
International Journal of Interactive Multimedia and Artificial Intelligence
December 2018, Vol. V, Number 3
ISSN: 1989-1660
unir
LA UNIVERSIDAD EN INTERNET

INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE
ISSN: 1989-1660 –VOL. 5, NUMBER 3

Dr. Alberto Magreñán, Real Spanish Mathematical Society, Spain
Dr. Monique Janneck, Lübeck University of Applied Sciences, Germany
Dr. Carina González, La Laguna University, Spain
Dr. David L. La Red Martínez, National University of North East, Argentina
Dr. Juan Francisco de Paz Santana, University of Salamanca, Spain
Dr. Héctor Fernández, INRIA, Rennes, France
Dr. Yago Saez, Carlos III University of Madrid, Spain
Dr. Guillermo E. Calderón Ruiz, Universidad Católica de Santa María, Peru
Dr. Giuseppe Fenza, University of Salerno, Italy
Dr. José Miguel Castillo, SOFTCAST Consulting, Spain
Dr. Moamin A Mahmoud, Universiti Tenaga Nasional, Malaysia
Dr. Madalena Riberio, Polytechnic Institute of Castelo Branco, Portugal
Dr. Juan Antonio Morente, University of Granada, Spain
Dr. Holman Diego Bolivar Barón, Catholic University of Colombia, Colombia
Dr. Sara Rodríguez González, University of Salamanca, Spain
Dr. Elpiniki I. Papageorgiou, Technological Educational Institute of Central Greece, Greece
Dr. Edward Rolando Nuñez Valdez, Open Software Foundation, Spain
Dr. Juha Röning, University of Oulu, Finland
Dr. Luis de la Fuente Valentín, Universidad Internacional de La Rioja - UNIR, Spain
Dr. Paulo Novais, University of Minho, Portugal
Dr. Giovanny Tarazona, Francisco José de Caldas District University, Colombia
Dr. Sergio Ríos Aguilar, Corporate University of Orange, Spain
Dr. Mohamed Bahaj, Settat, Faculty of Sciences & Technologies, Morocco
Dr. Javier Martínez Torres, Centro Universitario de la Defensa de Marín, Escuela Naval Militar, Spain
Dr. Abel Gomes, University of Beira Interior, Portugal
Dr. Edgar Henry Caballero Rúa, Inforfactory SRL, Bolivia

Editor's Note

The International Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI (ISSN 1989 - 1660) provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on AI tools or tools that use AI with interactive multimedia techniques. This was the first phrase that appeared into the website of the journal, whose launching had several motivations. First, IJIMAI was established on December 2008 in response to several agents, such as students, teachers, researchers, primarily in Spain and Colombia, who wanted to increase the impact of science in their environment. Second, IJIMAI was established to increase the number of scientific journals developed in Spain into the scope of Artificial Intelligence and Interactive Multimedia; there are very few journals about these topics in our country. Third, since the beginning we believed into an open access project, open for the whole stakeholders. Currently no money is needed to public a contribution in IJIMAI, and no money is needed to read all papers in IJIMAI as well; science should be open to achieve the maximum dissemination of knowledge. Finally, IJIMAI was established with the hope of being a long-term project; this 10th anniversary allows us to affirm that this goal is getting closer.

Since 2008 many changes have happened, the main ones are as follows:

- From 2008 to 2011 we only published one issue per year. These 4 years were very difficult. We had to work hard to find good quality papers.
- From 2012 to 2015 we increased our capacity of publication. We extended our issues from 1 to 4 per year. Also we achieved several indexations in different indexes and databases such as, DOAJ, INSPEC, DBLP, LATINDEX, among others. Many colleagues decided to work with us and they enrolled into the project. They helped us offering IJIMAI to several congresses such as IEEE-DCAI, IBERAMIA, WORLDCIST, etc.; they also helped us working as reviewers, editors and authors. IJIMAI was transformed into a robust project.
- From 2015 to 2018 IJIMAI was indexed at Web of Science through Emerging Science Citation Index. The visibility increased a lot and the number of received papers also increased notably. There are some data that reflect this status as you can read below.
  - Considering the years 2013-2018 the average number of articles per year is 47, while if we only consider the last two years, the average number is 53.
  - All research articles are peer-reviewed through a blind process in which identities of authors and reviewers are hidden.
  - The Editorial Board of IJIMAI is composed of important researchers in Computer Science. With the goal to provide a wide geographic coverage of the journal and its impact, the Editorial Board has members of 23 countries: Argentina, Australia, Bolivia, Canada, China, Colombia, Finland, France, Germany, Greece, India, Ireland, Italy, Japan, Malaysia, Morocco, Norway, Peru, Portugal, Spain, Slovakia, United Kingdom and USA.
  - I should highlight the high internationality of IJIMAI. There is a 66% of international institutions that publish papers in IJIMAI.
  - The current acceptance rate is 0.30.
  - Publication of special issues collaborating with authors of prestigious entities such as Hospital Universitario La Paz of Madrid (Spain) and technology-related companies such as Telefónica, BBVA or Banco Santander (Special Issue on 3D Medicine and Artificial Intelligence, Special Issue on Advances and Applications in the IoT and Cloud Computing, Special Issue on Big Data and AI, etc.).
  - Since the start of the journal in 2008 until December 2018, 365 articles have been published in the journal.
  - Taking into account last studies done, calculated with the same parameters than Clarivate Analytics, our impact factor is 1.05.
  - Finally, I would like to remark that IJIMAI has been indexed in the annual Ibero-American Journal Ranking, launched by the «Red Iberoamericana de Innovación y Conocimiento Científico» (REDIB) and Clarivate Analytics, appearing in position 27 of this ranking in which 748 journals are listed.

IJIMAI has to thank many colleagues who have helped the journal in many times. Thanks to all of them. However, there are proper names that have to be mentioned because they have been key in this project: Jesús Soto, Oscar Sanjuán, Carlos Montenegro, Juan Manuel Cueva, Enrique Herrera-Viedma, Francisco Mochón, Daniel Burgos, Ainhoa Puente, Elena Verdú and of course Miguel Arrufat.

Also, IJIMAI has to thank all authors for all the papers sent and all the papers published. Without authors there is no journal. As an expression of gratitude, to commemorate the 10th Anniversary, we have awarded the Prize for the most outstanding paper published in these 10 first years. Both quality and impact of the paper have been taken into account when awarding this Prize. The selection process has been as follows:

- A list of finalist papers was obtained from those receiving unless 0.5 citations/month since its publication according to Google Scholar data.
- Each Associate Editor of IJIMAI voted for his/her three favorite papers from those of the list of finalists.
- The winning paper has been the one receiving the highest number of votes.

The winners are Dr. Pekka Siirtola and Dr. Juha Röning from Finland with the paper *Recognizing human activities user-independently on smartphones based on accelerometer data*, published in June 2012. This is a very interesting research about recognition of everyday human activities in real time using a mobile phone which, according to Google Scholar data, has received 1.8 citations per month.

Given this special occasion, apart from original research articles, we have included three invited papers and testimonials written by key players in this issue. The result is a collection of high-quality papers that reflect the wide scope of the Journal, within the AI field. The research works presented in this issue cover different topics of interest as music creation, simulated sound propagation, factors influencing game users' retention, human computer interaction, human activity recognition, watermarking security, sentiment analysis, multi-issue negotiation, etc. The involved techniques are also diverse as those based on neural networks, particle swarm optimization, fuzzy logic, case-based reasoning, discrete wavelet transform, feature selection, evolutionary learning, multi-objective optimization, etc.

The first paper is an invited work, which is a bibliometric analysis of all the publications in IJIMAI during its first decade. The objective was to identify, through a science mapping, the most relevant structural aspects of these publications, providing a retrospective evaluation. This study has been undertaken by Baier-Fuentes et al. [1], resulting in interesting findings regarding co-citations, bibliographic coupling, co-authorship and the co-occurrence of keywords.

DOI: 10.9781/ijimai.2018.12.002

The second invited work is written by Dr. Pekka Siirtola and Dr. Juha Röning [2], authors of the awarded article above mentioned. In this new article they give an overview of progresses in the field of activity recognition and wearable sensor-based recognition, focusing on introducing their specific work, since the publication of their awarded paper in 2012. Currently, with an increasing market of wearable devices, research in this field presents promising.

Next two articles deal with sound related topics, although these are very different. Padilla and Conklin [3] focus on creative intelligence, proposing a music generator based on the corpus of masses of Palestrina. They combine statistical generation and pattern discovery using a template piece. This template solves the problem of coherence and imitation as fundamental novelty in counterpoint music generation. A first-order Markov model is used to generate two-voice counterpoint in florid style into the patterns selected. The paper concludes with a proposed relationship between musical quality and information content.

The other article about sound concerns simulated spatial sound propagation to provide the sense of realism and immersion in virtual environments. Lakka et al. [4] make an interesting review of the state-of-the-art techniques based on acoustic principles that apply, not only to real rooms, but also to 3D virtual environments such as those used in simulated training or computer gaming.

Regarding gaming, Rada et al. [5] present a novel study on the free-to-play video game industry: using a real database of the users of the "Red Eagle Origins" game, which is based on a famous Spanish TV Series, they measure the impact of TV series and social networks on the main game metrics: activation, retention and monetization.

Next article targets the Human Computer Interface (HCI) topic to improve the people ability to use a computer mouse. Mack [6] proposes an artificial neural network based filter to remove unwanted tremor-induced motion in computer mouse input. The filter is successful at removing a simulated Parkinson's tremor from computer mouse movements with minimal training data. Thus, this kind of solution can help those having diseases causing some tremor in hands.

Diseases diagnosis is the topic covered by the following articles. Arun et al. [7] present a novel approach for detection of depression using clinical data of people who have undergone a comprehensive assessment for cognitive function, mental health and cardiometabolic disorders. They use varied solutions as a Meta-Cognitive Neural Network (McNN) classifier with Projection-based learning (PBL), XGBoost for feature selection, and Particle Swarm Optimization (PSO) to select the best parameters of the McNN-PBL algorithm. The McNN-PBL classifier helps the clinician identify depressed patients resulting in improved treatment and prevention of progression of depression.

Next work also intends to help clinician to diagnose an illness, diabetes, but the proposed methods are very different, based on the case-based reasoning (CBR) concept. By using a fuzzy decision tree, Benamina et al. [8] integrate fuzzy logic and data mining to improve the case retrieval step in CBR. They achieve to reduce the complexity of calculating the degree of similarity that can exist between diabetic patients who require different monitoring plans.

Ending with the series of health related articles, Ayad and Khalil [9] present a new semi-blind image watermarking system for medical applications. The proposed scheme uses a combination of Singular Value Decomposition and Discrete Wavelet Transform to embed the watermark in a transparent manner and extract it with high fidelity. Simulation results shows its good imperceptibility and high robustness against several attacks.

Research of Gaona-García et al. [10] concerns the limitations associated with irrelevant search results in digital repositories. The authors present recommendations using information visualization strategies based on Simple Knowledge Organization Systems (SKOS) for the development of navigational search interfaces in digital repositories. The aim is to facilitate the access to learning objects on the basis of knowledge areas.

Nowadays, there is a great amount of digital information growing such as the professional content in repositories or the information produced by users with lots of personal opinions in, for example, social media. Therefore automatic processing of this information is the focus of many recent researches. One of this is the one presented in the article of Dehkharghan [11]. In the field of sentiment analysis, he describes a solution to extract the polarity of texts that uses a semi-automatic methodology to build phrase polarity lexicons. He bases on a premise that indicates that the polarity of the whole phrase cannot be estimated based on the polarity of its parts.

Also related to text mining, Revanasiddappa presents a new feature selection method based on Intuitionistic Fuzzy Entropy (IFE) for Text Categorization [12]. Unlike typical Fuzzy Entropy, which considers only membership degree, IFE considers also non-membership and hesitation degrees, improving handling of uncertainty.

Next work [13] is part of a collaborative decision support system applied to industrial diagnosis. Benkaddour et al. propose a global architecture of a recommender tool that provides diagnostic documents for industrial operators. Documents describing solutions and the information gathered from collaborative sessions and Web 2.0 tools are filtered by the recommendation system to improve the search results taking into account users' preferences. The aim is to find effective solutions to the breakdowns in a short time to improve productivity of companies.

Applicable also to the business world, the evolutionary learning agent presented by Ayachi et al. is able to estimate its opponent's deadline and reserve points in a bilateral multi-issue negotiation based on opponent's counter-offers [14]. They model the learning process as an optimization problem to learn its opponent's parameters and use a new concession strategy adjustment to improve the agent's outcome, which is shown to be very close to the best scenario.

Going back to the HCI topic, the paper of Raees and Ullah presents a novel eyes-based interaction technique for navigation [15]. Gestures and positioning of eyes are considered interaction instructions by the system and no extra device is needed other than an ordinary camera. The proposed system is applicable to navigation in 3D interactive virtual spaces as those of 3D gaming and simulation.

Ahmed et al. [16] target the probabilistic load flow problem considering the optimal location and size of Static Var Compensators (SVC) in radial distribution systems. The aim is to minimize the total power loss and voltage deviation. By using Pareto Envelope-based Selection Algorithm II (PESA-II), as multi-objective optimization method, with fuzzy logic decision maker, they achieve to reduce the number of iterations and computation time compared to the Monte Carlo method.

The article closing this issue deals with a very important topic, the target of humanity, which is happiness. Mochón [17] provides a survey of works about the impact of technology in happiness, with special consideration of social networks. Although technology has improved many aspects of our life, people is not happier than in previous generations. This and other contradictions about happiness are commented. The paper goes over positive effects of technology on happiness, as well as negative effects and possible actions to mitigate them, concluding with some reflections of the author.

As above mentioned, there are some colleagues, as the author of this last invited paper about happiness, who have been key in this process of building and consolidating IJIMAI. Below, the testimonials of some of them:

"For a long time I had been looking for a good journal. At a conference in 2012, after my lectures, Ruben asked me if I would like to publish in the (young) IJIMAI. After I had looked at the articles in detail, I agreed and was immediately positively surprised. After reviewing some manuscripts, proposing some ideas about the internet presence and my reports of positive experiences to colleagues, I was accepted into the team of "Editorial Board Members" and, sometime later, into the group of "Associate Editors". Meanwhile I have published regularly in the IJIMAI and I am always happy about the fast response times and the good support. What makes the IJIMAI special for me? There are three points: the constantly increasing quality of the contributions, the fast and always constructive feedback from Ruben and Elena and the fast response times of the reviewers. Elena and Ruben manage again and again to motivate us "as reviewers" to answer quickly and conscientiously. And this is one reason why authors like a journal. The other reason is surely that standard e-mails are avoided as much as possible and every author is taken seriously."

Jörg Thomaschewski
Hochschule Emden/Leer, Emden (Germany)

"For the past 6 years, I have dedicated my professional life to technological entrepreneurship, focusing on the latest technology trends in Silicon Valley. I was truly blessed as I was able to gain experience by working with the most distinguished experts of the field and learning about resilience, and how to stand up again when something does not go as expected. When I look back and think about the origins of IJIMAI, it is inevitable to find a remarkable parallelism between the creation of a startup and the conception of IJIMAI. I remember feeling the same illusion, investing the very same amount of effort, having clear goals, and experiencing a strong organic growth. I also think of this journal as a great family, that continuously nurtures with collaborations, a wider network, and exceptional contents. Therefore, I would like to conclude by dedicating a phrase of Émile Couéme to IJIMAI, as this sensational journal also has a bit of soul, represented in its optimism: "Every day and in every way, you are getting better and better." Congrats IJIMAI."

Óscar Sanjuán Martínez
CenturyLink (USA)

"IJIMAI is a journal with a focus on artificial intelligence but with a multidisciplinary character that I find very interesting. I started collaborating with IJIMAI as a referee three years ago. I was lucky when Rubén, the Editor-in-Chief of the journal, offered me to be more involved in the journal activities, being an Associate Editor. I am a researcher in the artificial intelligence field and I love the publishing world so the conjunction of these two subjects was really attractive for me. Currently I am the Managing Director and I really want to take this opportunity to thank every author who has submitted a manuscript to IJIMAI and, especially, to thank every reviewer because the reviewers' work does not look for a compensation and highly contributes to achieve quality in research publications. The enthusiasm and hard work of Rubén and colleagues from the Editorial Board, the involvement of UNIR, and the highly appreciated collaboration of authors and reviewers, are the key to the achieved milestones until now and more to come."

Elena Verdú
Universidad Internacional de La Rioja - UNIR (Spain)

"Artificial Intelligence (AI) is becoming the engine of the technological revolution that we are witnessing at this moment. AI together with the Internet of Things, Robotics and Big Data are increasingly present in our lives and draw a future of more development and progress that will contribute to improving the quality of our lives and the understanding of the world that surrounds us. There are many advances in AI and applications that are being made in the scientific community, and many international journals that are focusing their publications on AI. IJIMAI is one of the relevant Spanish journals that is serving as a showcase for some of the best AI contributions that are being developed in Computer Science. The Editor in Chief, Ruben Gonzalez, and his team with a serious job have made IJIMAI a quality scientific channel, with a serious referee process, excellent publication time and an attractive presentation. IJIMAI today attracts the attention of many researchers from Spain as well as from Latin American and Asian emerging countries that seek to know the latest advances in AI and disseminate their latest research in AI. I feel honored to collaborate with the journal from the editorial board. I think it is at a very good moment in its life as evidenced by the increase in publications and citations received, but without a doubt the best is yet to come: IJIMAI is going to become in the coming years one of the most important Spanish journals of AI for the dissemination of knowledge in AI in our international scientific community."

Enrique Herrera University of Granada (Spain)

"Artificial Intelligence (AI) is omnipresent in our daily life and has penetrated almost all research fields. IJIMAI is a pioneering journal that for 10 years ago spotted a hot area of interdisciplinary research that lies at the intersection between AI and interactive multimedia techniques. The journal has evolved over time and is delivering high quality publications thanks to the contributions of authors and editors. We have been indexed by high-quality online references for bibliographic information such as dblp and Web of Science – CLARIVATE and we are actively working on gaining other indexation recognitions."

Anis Yazidi
Oslo Metropolitan University (Norway)

"IJIMAI provides a rare combination of quality content with a fine editing work. Thanks to the thorough selection process and the cunning vision to the future, this publication plays a significant role to the scientific community. Indeed, IJIMAI becomes an irreplaceable resource to be up-to-date, to think of the next step, and to learn on the way."

Daniel Burgos
Universidad Internacional de La Rioja - UNIR (Spain)

"The collaboration with IJIMAI, preparing a special issue since 2014, has been one of the most enriching experiences in which I have participated during the last years due to its interdisciplinary and innovative character. The freedom I have enjoyed to select the topics has been an interesting challenge and has opened the opportunity for me to work on various aspects that are located on the frontier of research and to make contact with researchers from all over the world."

Francisco Mochón Morcillo
National Distance Education University (Spain)

"Think about where you were 10 years ago. Well, some people were thinking in how to create the International Journal of Interactive Multimedia and Artificial Intelligence – IJIMAI. This happened 10 years ago, starting was not easy but the first published issue of the journal was achieved in 2008, with several different concepts. One of them, which is very peculiar, is its cover. As everyone can see, covers correspond to own images that own editors have got because of their love for photography. I would like the reader of this note to visit every cover and choose its favourite one."

Carlos E. Montenegro
Francisco José de Caldas District University (Colombia)

"Congratulations to IJIMAI for 10th anniversary! It has been our great honor to be part of this journey. Our article "Recognizing human activities user-independently on smartphones based on accelerometer data" was published in IJIMAI in 2012. Thanks to IJIMAI and its open access approach, our article has gotten a lot of publicity and readers. In fact, it has been cited almost 150 times. While the IJIMAI's tremendous achievements over the past ten years are now a reason to celebrate, we hope that this is just a beginning. All the best for the next ten years!"

Pekka Siirtola and Juha Röning
University of Oulu (Finland)

To conclude this special Editor's Note, I want to thank you all for being a part of this project from all the Board of IJIMAI, specially our readers, who I hope enjoy this issue as much as we have enjoyed it during its creation.

Dr. Rubén González Crespo

[1] H. Baier-Fuentes, J. Cascón-Katchadourian, M.A. Martínez, E. Herrera-Viedma, José M. Merigó. A bibliometric overview of the International Journal of Interactive Multimedia and Artificial Intelligence. International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 9-16.

[2] P. Siirtola, J. Röning. Revisiting "Recognizing human activities user-independently on smartphones based on accelerometer data"– what has happened since 2012? International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 17-21.

[3] V. Padilla, D. Conklin. Generation of Two-Voice Imitative Counterpoint from Statistical Models, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 22-32.

[4] E. Lakka, A. G. Malamos, K. G. Pavlakis, J. A. Ware. Spatial Sound Rendering – A Survey, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 33-45.

[5] F. De Rada, A. Monchon, Y. Saez. TV Series and Social Media: Powerful Engagement Factors in Mobile Video Games, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 46-55.

[6] S. Mack. Removing Unclassified Hand Tremor Motion from Computer Mouse Input with Neural Networks, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 56-60.

[7] V. Arun, M. Krishna, Arunkumar B.V., Padma S.K., Shyam V. Exploratory Boosted Feature Selection and Neural Network Framework for Depression Classification, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 61-71.

[8] M. Benamina, B. Atmani, S. Benbelkacem. Diabetes Diagnosis by Case-Based Reasoning and Fuzzy Logic, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 72-80.

[9] H. Ayad, M. Khalil. QAM-DWT-SVD Based Watermarking Scheme for Medical Images, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 81-89.

[10] P. Gaona-García, C. Montenegro-Marín, E. Gaona-García, A. Gómez-Acosta, Y. Hassan-Montero. Issues of Visual Search Methods in Digital Repositories, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 90-97.

[11] R. Dehkharghani. Building Phrase Polarity Lexicons for Sentiment Analysis, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 98-105.

[12] M. B. Revanasiddappa, B. S. Harish. A New Feature Selection Method based on Intuitionistic Fuzzy Entropy to Categorize Text Documents, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 106-117.

[13] F. Z. Benkaddour, N. Taghezout, F. Z. Kaddour-Ahmed, I. Hammadi. An Adapted Approach for User Profiling in a Recommendation System: Application to Industrial Diagnosis, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 118-130.

[14] R. Ayachi, H. Bouhani, N. Ben Amor. An Evolutionary Approach for Learning Opponent's Deadline and Reserve Points in Multi-Issue Negotiation, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 131-140.

[15] M. Raees, S. Ullah. EVEN-VE: Eyes Visibility Based Egocentric Navigation for Virtual Environments, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 141-151.

[16] W. Ahmed, A. Selim, S. Kamel, J. Yu, F. Jurado. Probabilistic Load Flow Solution Considering Optimal Allocation of SVC in Radial Distribution System, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 152-161.

[17] F. Mochón. Happiness and technology: special consideration of digital technology and Internet, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), vol. 5, no. 3, pp. 162-168.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

A Bibliometric Overview of the International Journal of Interactive Multimedia and Artificial Intelligence

Hugo Baier-Fuentes[1], Jesús Cascón-Katchadourian[2], M.A Martínez[3], Enrique Herrera-Viedma[4], José M. Merigó[5,6] *

[1] Department of Business Administration, Universidad Católica de la Santísima Concepción. Av. Alonso de Ribera 2850, Concepción (Chile)
[2] Department of Information and Communication. Faculty of Communication and Documentation. Colegio Máximo de Cartuja, 18071, Granada (Spain)
[3] Department of Social Work and Social Services, University of Granada, Granada (Spain)
[4] Department of Computer Science and Artificial Intelligence, University of Granada, Av. Periodista Daniel Saucedo s/n, Granada, (Spain)
[5] School of Systems, Management and Leadership, Faculty of Engineering and Information Technology, University of Technology Sydney, 81 Broadway, Ultimo, 2007 NSW (Australia)
[6] Department of Management Control and Information Systems, School of Economics and Business, University of Chile, Av. Diagonal Paraguay 257, 8330015 Santiago (Chile)

Received 20 November 2018 | Accepted 28 November 2018 | Published 1 December 2018

Abstract

The International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI) published its first issue ten years ago. Currently, IJIMAI is indexed in the important database Emerging Sources Citation Index. This paper aims to identify, through a mapping of science, those most relevant aspects of the structure of publications made during the first 10 years of IJIMAI. Using VOSviewer software, the structural maps of the IJIMAI publications are analysed according to techniques such as bibliographic coupling, co-citations and co-occurrence of keywords. In addition, the evolution of the publications, citations and an analysis of the most cited papers of the journal are presented. The results show that IJIMAI has experienced a remarkable growth of both publications and citations in the last five years. We also observe that IJIMAI does not only capture the attention of the Spanish scientific community, but also of emerging countries such as India and Iran and emerging Latin American countries such as Colombia. With a such increasing behaviour, it is expected in the coming years that IJIMAI will position itself among the best journals with similar scientific scope.

Keywords

Bibliometrics, Web Of Science, VOS Viewer.

DOI: 10.9781/ijimai.2018.12.003

I. Introduction

As stated in its access platform, the International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI hereinafter), pretends to be an open access point to relevant research on advances in Artificial Intelligence tools or tools that use Artificial Intelligence with interactive multimedia techniques. IJIMAI published its first issue in December 2008, under the direction of its editors, Dr. Jesús Soto Carrión and Dr. Elisa García Gordo. This issue published research from diverse areas of knowledge, such as Medical Diagnosis, Semantic Metadata, Nature Conservancy and Intelligence perception. Until 2011, IJIMAI kept publishing an annual number. The period 2012 published 3 issues and finally, since 2013, it has kept publishing 4 issues per year. The year 2015 was specially important since the journal was indexed in the *Emerging Sources Citation Index* (ESCI hereinafter), considered an important database in the *Web of Science* (WoS hereinafter).

In 2018, IJIMAI celebrates its tenth anniversary, which stimulated our interest in carrying out a bibliometric analysis of all the publications made by the journal. According to Cobo et al. [1], Bibliometrics is a set of methods used to study the impact of a particular field, the impact of a set of researchers or a particular article. Among the main methods used in a Bibliometrics is performance analysis and science mapping [2]. The first one focuses on assessing the impact of scientific actors on a bibliographic database. Keep in mind that scientific actors involve countries, universities, departments, researchers. On the other hand, science mapping intended to show the structural and dynamic aspects of a field of research using techniques such as bibliographic coupling [3], the co-citations of documents [4], citation, co-authorship or co-ocurrence of keywords. In line with the above, we want to emphasize that the objective of this contribution is to identify, through a science mapping the most relevant structural aspects of the publications made in IJIMAI. For this, this study uses the VOSviewer Software [5].

* Corresponding author.
E-mail addresses: hbaier@ucsc.cl (Baier-Fuentes), jesuscascon@gmail.com (Cascón-Katchadourian), mundodesilencio@ugr.es (Martínez), viedma@decsai.ugr.es (Herrera-Viedma), jmerigo@fen.uchile.cl (Merigó).

It is important to note that advances in Computer Science and the Internet has intensified the popularity of these techniques among researchers [6], and therefore, at present it is very common to find a combination of performance analysis and / or science mapping applied for example, to specific fields of studies, such as *Management* [7], *Economics* [8], *International Entrepreneurship* [9] or *Knowledge Management* [10]. In recent years, these analyses have also become common in journals when they celebrate a significant anniversary. These analyses are interesting since they provide some general and historical results that allow the development of a retrospective evaluation [11]. Among some of the journals that have published these analyses we could/can name, for example, *Information Sciences* in the celebration of its fifty years [12], *European Journal of Marketing* in recognition of its forty years, *International Journal of Intelligent Systems* in its thirty years [13], *Knowledge-Based Systems* at the age of twenty-five years [14], *Journal of Product Innovation Management*, or more recently, *Journal of Knowledge Management*, in the celebration of his twenty years [15] [16]. Finally, keep in mind many other journals have already developed a bibliometric description to celebrate a special anniversary [17] [18].

This article is structured as mentioned below. Section II reviews the bibliometric methods, but in a special way, science mapping is introduced. Section III presents the results including several analysis of co-citations, bibliographic coupling, co-authorship and the co-occurrence of keywords. Finally, section IV summarizes the main findings and conclusions of the document.

II. Methods

This study performs a bibliometric analysis of the IJIMAI publications between 2008 and 2018. For this, data from the *Web of Science* belonging to *Clarivate Analytics* was collected. Note that the WoS contains a compilation of several citation databases, transforming WoS into the world's leading citation database that covers more than 18,000 high-impact journals. Note however, that there are several other citation databases in the world, such as Scopus, EconLit, Scielo, among others.

Bibliometrics is generally defined as the science that quantitatively studies bibliometric material [19]. Currently, it involves a variety of techniques and methodologies, including performance analysis and science mapping [20], the latter also known as bibliometric mapping. This methodology is an important research topic in the field of Bibliometrics [21], and focuses on finding representations of intellectual connections from a dynamically changing repository of scientific knowledge [22] . Our study seeks to analyse the structural aspects of scientific research published in the 10 years of IJIMAI, although we also disclose information about the most cited papers of the journal.

The advancement of information technologies has allowed this technique to become popular with force. Thanks to this, a wide range of software tools to analyse the bibliographic information has been developed. These include, for example, CiteSpace II [23], Bibexcel [24], VantagePoint [25], VOSViewer software [5], the latter being the one used in this study.

VOSViewer software builds the scientific maps based on some techniques based on bibliographic coupling, co-citations of documents, analysis of co-authorship or analysis of co-words. Note that bibliographic coupling [3] [26] [27] analyses the documents citing the same third document, while the analysis of co-citations [28] [4] focuses on the cited documents. The co-authorship analysis focuses on the authors and their affiliations to study social structure and collaborative networks [29] [30]. Finally, the co-words analysis focuses on the most important words or keywords of the documents, allowing a field of research to be structured conceptually [31] [32] [33].

Note also that the graphical visualization delivered by VOSviewer software is made through network structures, where the size of a circle shows the relevance of an element and the network connections allow identification of the most closely linked elements. The place of the circles and the colours are used to group the elements. Finally, keep in mind that VOSviewer software is an open-source bibliometric mapping tool, which can be easily downloaded from the site http://www.vosviewer.com/.

III. Results

A. Evolution of Publications, Citations and Most Cited Papers of IJIMAI

The first issue of IJIMAI was published in 2008. During its 10 years of circulation, it has published 325 documents, which have received appointments at least 400 times, according to data of WoS in autumn 2018. In WoS its h-index is 7, there are seven publications that have received 7 citations or more. Fig. 1 presents the evolution of publications since 2008. We should point out that IJIMAI is indexed in WoS from 2015, and its citation evolution cannot be well visualized in WoS.

During its first 4 years, IJIMAI published a number per year. The year 2012 published 3 annual issues. 2013 is the year in which it begins to publish 4 issues per year. In 2015 IJIMAI is indexed in the database Emerging Sources Citation Index (ESCI) provided by Clarivate Analytics, one of the citation databases that is part of the WoS. This allowed IJIMAI to overcome the barrier of the 40 studies in the same year, and of the 50 studies during the following two years. One explanation for this is that the journals indexed in ESCI database, and consequently in WoS database, are distributed and exposed to more and different researchers in various institutions around the world [34].

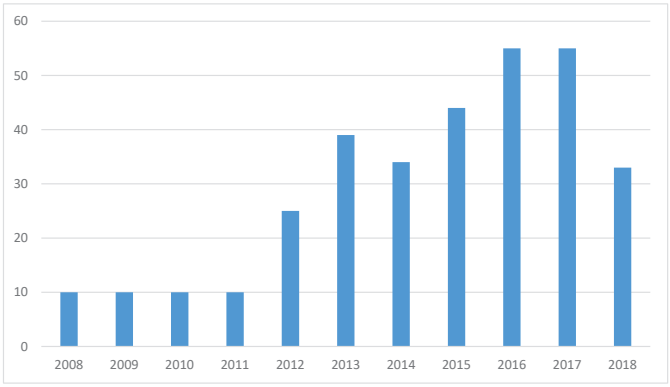


Fig. 1. Annual number of papers published in IJIMAI.

Additionally, we complete the analysis of WoS citations with information provided by Google Scholar. In Fig. 2, we show the citation report of IJIMAI according to Google Scholar. We observe that IJIMAI presents a remarkable h-index of 20 with 2046 citations received from 2013. Therefore, IJIMAI shows a remarkable growth of citations in the last five years as it is shown in the figure.

On the other hand, IJIMAI has published several documents that in total, have been cited more than 400 times by documents from different disciplines. In this sense, it is important to analyse the most cited articles of the journal. Table I presents the twenty-five most cited articles of IJIMAI. The indicators used in the table are defined in the footer.

Table I. The 25 most cited documents in IJIMAI

R	Title	Author/s	Year	TC	C/Y
1	Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data	Siirtola, P; Roning, J	2012	65	10.83
2	Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ).Example: Spanish Language Version	Rauschenberger, M; Schrepp, M; Perez Cota, M; Olschner, S; Thomaschewski, J	2013	31	6.20
3	Framework for Computation Offloading in Mobile Cloud Computing	Kovachev, D; Klamma, R	2012	22	3.67
4	Infected Fruit Part Detection using K-Means Clustering Segmentation Technique	Dubey, SR; Dixit, P; Singh, N; Gupta, JP	2013	19	3.80
5	A fuzzy c-means bi-sonar-based Metaheuristic Optimization Algorithm	Khan, K; Sahai, A	2012	15	2.50
6	A review about Smart Objects, Sensors, and Actuators	Gonzalez Garcia, C; Meana-Llorian, D; Pelayo G-Bustelo, BC; Cueva Lovelle, JM	2017	14	7.00
7	Multilayer Perceptron: Architecture Optimization and Training	Ramchoun, H; Amine, M; Idrissi, J; Ghanou, Y; Ettaouil, M	2016	14	4.67
8	Statistical Comparisons of the Top 10 Algorithms in Data Mining for Classification Task	Settouti, N; Bechar, MEA; Chikh, MA	2016	12	4.00
9	Big Data and Learning Analytics in Blended Learning Environments: Benefits and Concerns	Picciano, AG	2014	12	3.00
10	Big Data & eLearning: A Binomial to the Future of the Knowledge Society	Alonso, V; Arranz, O	2016	11	3.67
11	The Cambria Explosion of Popular 3D Printing	Chulilla Cano, JL	2011	11	1.57
12	Review of Current Student-Monitoring Techniques used in eLearning-Focused recommender Systems and Learning analytics. The Experience API & LIME model Case Study	Corbi, A; Burgos, D	2014	10	2.50
13	Local convergence for an improved Jarratt-type method in Banach space	Argyros, IK; Gonzalez, D	2015	10	2.50
14	GLOA: A New Job Scheduling Algorithm for Grid Computing	Pooranian, Z; Shojafar, M; Abawajy, JH; Singhal, M	2013	10	2.00
15	Text Analytics: the convergence of Big Data and Artificial Intelligence	Moreno, A; Redondo, T	2016	9	3.00
16	Social Voting Techniques: A Comparison of the Methods Used for Explicit Feedback in Recommendation Systems	Rolando Nunez-Valdez, E; Manuel Cueva-Lovelle, J; Sanjuan, O; Montenegro-Marin, CE; Infante Hernandez, G	2011	9	1.29
17	An Automated Negotiation-based Framework via Multi-Agent System for the Construction Domain	Mahmoud, MA; Ahmad, MS; Yusoff, MZM; Idrus, A	2015	8	2.00
18	Patterns of Software Development Process	Castro, SJB; Gonzalez-Crespo, R; Medina-Garcia, VH	2011	8	1.14
19	Combining Fuzzy AHP with GIS and Decision Rules for Industrial Site Selection	Taibi, A; Atmani, B	2017	7	3.50
20	Analysis of Gait Pattern to Recognize the Human Activities	Gupta, JP; Dixit, P; Singh, N; Aemwal, VB	2014	7	1.75
21	Assessing Road Traffic Expression	Silva, F; Analide, C; Novais, P	2014	7	1.75
22	Towards a standard-based domain-specific platform to solve machine learning-based problems	Garcia-Diaz, V; Pascual Espada, J; Pelayo G-Bustelo, BC; Cueva Lovelle, JM	2015	7	1.75
23	A Constraint-Based Model for Fast Post-Disaster Emergency Vehicle Routing	Amadini, R; Sefrioui, I; Mauro, J; Gabbrielli, M	2013	7	1.40
24	Dynamic, ecological, accessible and 3D Virtual Worlds-based Libraries using OpenSim and Sloodle along with mobile location and NFC for checking in	Gonzalez Crespo, R; Rios Aguilar, S; Ferro Escobar, R; Torres, N	2012	7	1.17
25	Ball convergence for Steffensen-type fourth-order methods	Argyros, Ioannis K.; George, Santhosh	2015	7	1.75

Abbreviations: R = Rank; TC = Total citations; C/Y Cites per year.

	Total	Desde 2013
Citas	2100	2046
Índice h	20	20
Índice i10	58	56

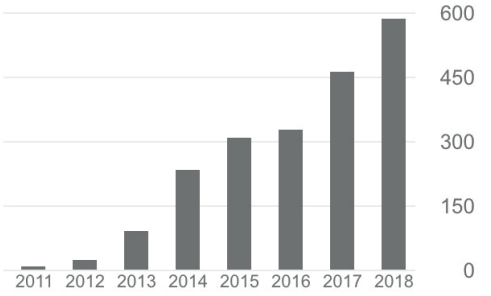


Fig. 2. Citation report of IJIMAI in Google Scholar.

B. Mapping Science of IJIMAI Publications

As previously mentioned, scientific mapping shows structural aspects of scientific research, becoming, therefore, an important methodology in the field of Bibliometrics [21]. Scientific mapping in practice, is a network structure often used to model the interaction between different scientific actors, such as authors, journals, keywords, references, among others [35]. In this study, therefore, a scientific mapping is presented that tries to analyze the most representative structures and connections between the different actors that have published during the 10 years of IJIMAI. Note that this analysis is presented in terms of techniques such as co-citations, bibliographic coupling and co-occurrence of keywords. For this, the VOSviewer software [5] is used, which has been developed specifically to build and visualize science mappings from files obtained from bibliographic databases, such as WoS, Scopus, PubMed and RIS [35].

Taking into account the above, the co-citations of the journals cited in IJIMAI are analysed. Note in this case, that co-citation occurs when two documents published in different journals receive a citation from the same third document from another journal. Fig. 3 that allows to perform this analysis, presents the information using a threshold of five citations and the hundred most representative co-citations links.

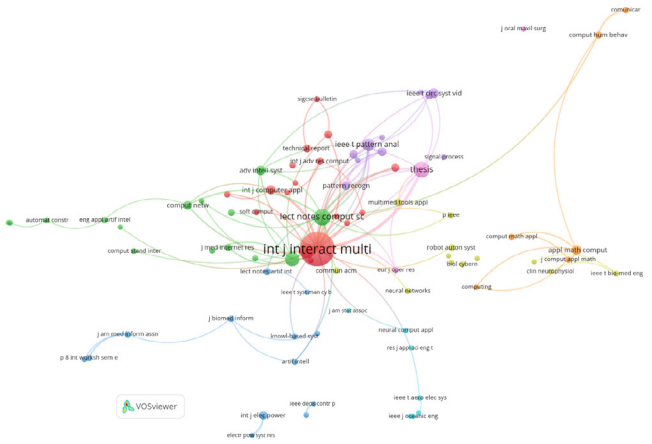


Fig. 3. Co-citation of journals in IJIMAI: minimum citation threshold of 5 and 100 links.

Note that the IJIMAI self-citations are easily observed occupying the centre of the map. This is not surprising, since it is very common in most journals, people developing science in a particular framework use close information sources. However, you can also see that computer journals are very influential. Among them, *Lecture Notes in Computer Science, Communications of the ACM, Lecture Notes in Artificial Intelligence, Soft Computing*. Also note the presence and influence of journals from other disciplines, but which are clearly linked to computer science. Such is the case of *Journal of Biomedical Informatics* or *Journal of the American Medical Informatics Association*.

Another interesting analysis to perform is the co-citation of authors. On the map presented in Fig. 4, it is noted to those authors with the highest number of citations and their connections with other authors. Fig. 4 is constructed with a threshold of five citations and the one hundred most representative links of IJIMAI.

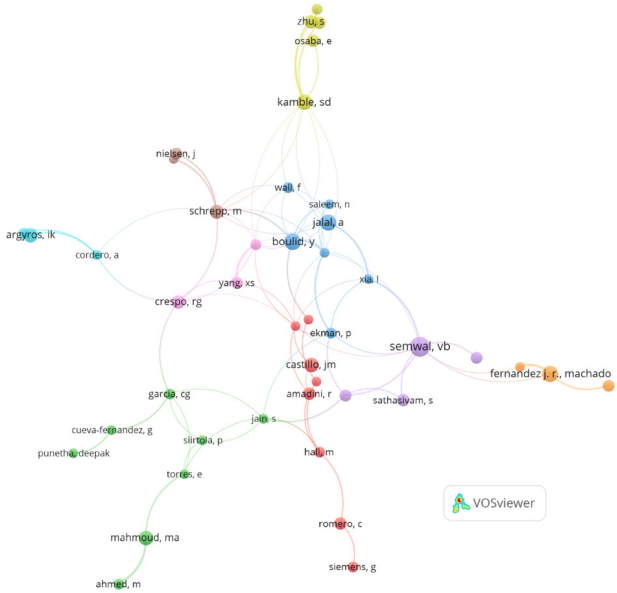


Fig. 4. Co-citation of authors cited in IJIMAI: minimum citation threshold of 5 and 100 links.

Note that there are several authors cited in the journal, but there is no clarity regarding a nucleus of authors that stand out in IJIMAI. Probably this is a consequence of the youth of the journal, and it is expected, therefore, that the most outstanding authors in the disciplines on which IJIMAI focuses will become a space in the same journal. Even so, some of the authors that stand out most in the map of co-citations of authors are, Semwal, V.B., Castillo, J.M., Ekman, P., Boulid, Y., among several others.

Another interesting subject to evaluate in this study is the bibliographic coupling of authors that publish in IJIMAI. Remember that the bibliographic coupling studies the authors of two documents that cite the same third document. Note that the connected documents appear on the map but not the third one, unless you also have a significant degree of bibliographic coupling through other documents [13]. In summary, the map presented in Fig. 5 shows the most productive authors among the entire set of IJIMAI documents.

This figure has the advantage of showing how the most productive authors of IJIMAI connect. Remember that the definition of the bibliographic coupling technique indicates that researchers who are close to each other tend to cite the same publications. Therefore, the interpretation of the figure allows us to highlight several clusters with authors that lead some lines of research in the journal. As it is logical, these authors tend to cite the same publications. For example, authors Thomaschewski, J., Taghezout, N. and Verdu, E. that notoriously form a cluster that develop a line of research.

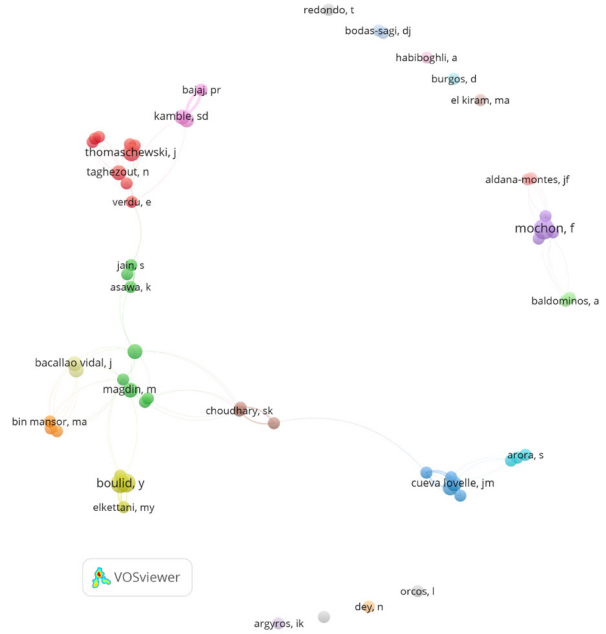


Fig. 5. Bibliographic coupling of authors publishing in IJIMAI: minimum publication threshold of 2 and 100 links.

As well as the previous analysis, it is also interesting to observe how the institutions that publish in the journal are coupled bibliographically. For this, Fig. 6 displays the data considering a minimum threshold of two documents and one hundred connections.

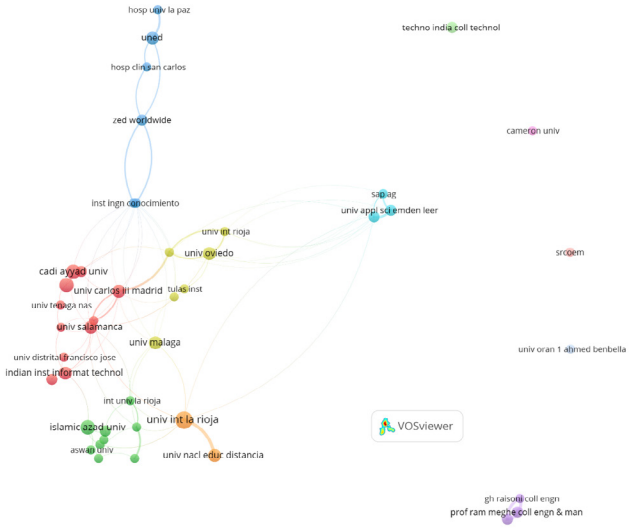


Fig. 6. Bibliographic coupling of institutions publishing in IJIMAI: minimum publication threshold of 2 and 100 links.

Fig. 6 shows the leading institutions in the publications in IJIMAI and their connections. As in the previous analysis, the visualization of this figure represents the trends in research carried out by the institutions that publish in the journal, since they tend to cite the same references. It is observed in general, that Spanish universities have a strong presence in the journal. However, unlike other analysis, in which the institutions of the same country are usually coupled bibliographically (see for example, [36] [37] [12] [13] , this visualization does not allow to observe a marked clustering by institutions of the same country. One explanation for this is that IJIMAI is still in an expansion stage and it is likely that institutions and researchers from the same country still do not know the journal.

Another interesting analysis to carry out is the bibliographic coupling at the country level. Remember that the visualization of this technique shows the countries that publish the most in the journal and that tend to cite the same documents. Fig. 7 shows the bibliographic coupling of the countries that publish in IJIMAI with a minimum threshold of one study and one hundred more representative links.

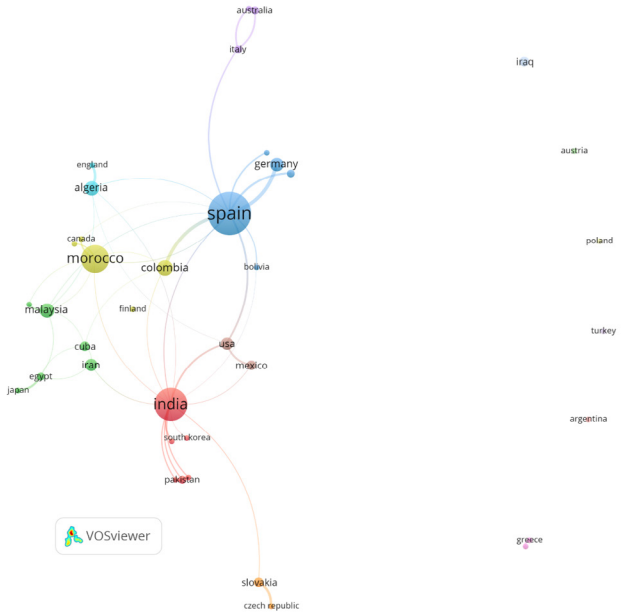


Fig. 7. Bibliographic coupling of countries publishing in IJIMAI: minimum publication threshold of 1 and 100 links.

Note that the most productive countries in the journal are Spain, India and Morocco. The case of this last country, is especially striking. Morocco has been implementing several policies aimed at the development of ICTs in the country, that is probably motivating research in scientific fields related to IJIMAI. On the other hand, and as in the previous visualization, there is no clear bibliographic coupling between neighboring countries. We believe, however, that as IJIMAI grows and positions itself among related journals, the nearby countries will tend to be bibliographically coupled with more notoriety.

Finally, we perform an analysis of co-occurrences of keywords [31]. The keywords can be extracted from the title and the summary of a publication, or also they can be obtained from the list of keywords provided by the author of a publication. According to van Eck and Waltman [38], the number of co-occurrences of two keywords is the number of publications in which both keywords appear together in the title, summary or list of keywords. In such a way, Fig. 8 presents the co-occurrence of keywords in the IJIMAI publications, with a minimum threshold of two co-occurrences and one hundred more representative links.

The visualization of Fig. 8 shows that a good diversity of terms is used in the publications of the journal. Although the terms most frequently used in publications are *Big Data*, *Machine Learning*, *Artificial Intelligence* and *Classification*. In general, a first interpretation of the figure is that there is a clear orientation of IJIMAI towards areas more related to *Artificial Intelligence* and *Data Mining*. In fact, the clusters located on the left side of the figure are related to the aforementioned areas. In addition, given the closeness of the terms, it can be assumed that these terms co-occur quite frequently in the IJIMAI publications. Other less frequent clusters are related to disciplines such as *Semantic Web* and *Interactive Multimedia*.

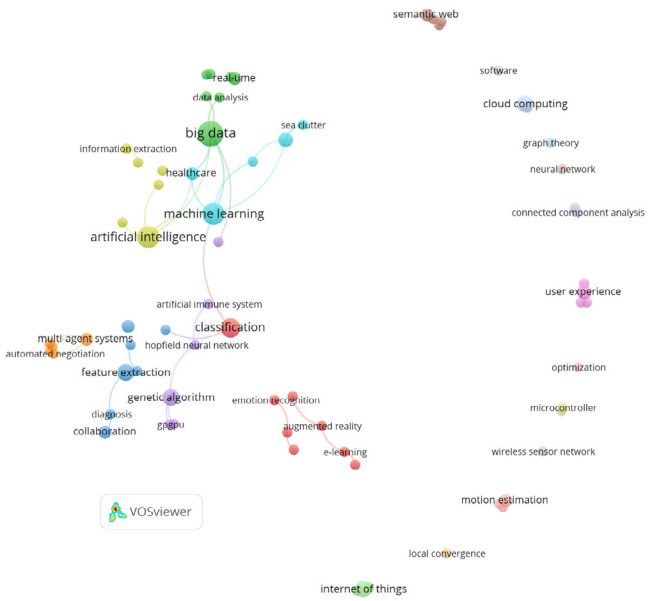


Fig. 8. Co-occurrence of keywords (Keyword Plus) in IJIMAI: minimum occurrence threshold of 2 and 100 links.

IV. Conclusions

IJIMAI is a young international magazine related to the Artificial Intelligence and tools that use Artificial Intelligence with interactive multimedia techniques. It celebrates its tenth anniversary in 2018. This fact has motivated the development of this work, which presents a bibliometric description, showing the main tendencies of the journal during its first ten years. This study uses the WoS database mainly, and specifically, the ESCI database, one of the databases of references that are part of the WoS database.

The bibliometric analysis developed was based mainly on science mapping or bibliometric mapping, considered one of the main bibliometric techniques [21]. In order to map the publications of IJIMAI, the VOSviewer software is used, that allows to show the structures of the publications of different scientific actors that have published during the 10 years of the journal. The techniques used in this science mapping are the bibliographic coupling and co-citation of authors, institutions and countries. Additionally, a co-occurrence analysis of keywords is incorporated and the evolution of the publications and the most cited articles of IJIMAI are also shown.

The results show that the co-cited journals are related to Computer Science and Information Systems in general. Spain, on the other hand, is the country with the largest presence in IJIMAI, a situation that is also common, since IJIMAI is an international journal from the University of La Rioja, located in Spain. This is also explained by a strong presence of Spanish universities publishing in the journal. Other countries that stand out in the journal, are India and Morocco, and to a lesser extent, emerging countries such as Colombia, Bolivia, Iran among several others. The science mapping analysis ends with an analysis of the most frequent and co-occurring keywords among them. In this sense, it can be pointed out that the main areas in which the IJIMAI publications are focused are *Big Data*, *Machine Learning*, *Artificial Intelligence* and *Classification*.

This work has been done to contribute with a general vision of the state and structure of the investigations published in IJIMAI. Like any document, this work presents some limitations that should be considered. In the first place, it must be taken into account that in the future, the presented data will change because the bibliometric data are dynamic and evolve. Obviously, IJIMAI must continue to grow and position itself among similar journals and the various disciplines that it declares in its scientific scope. This will allow in the future, to develop a more complete bibliometric analysis that involves other complementary methodologies to the scientific mapping, as for example, the analysis of bibliometric performance [20]. By doing this, the document can provide more information that the reader can assess according to their interests and priorities. Secondly, this analysis performs a science mapping analysis taking into account the authors who publish in IJIMAI. However, keep in mind that many of the authors who published in the first issues of the journal, they could change their membership through these ten years, therefore, some of the results shown in this scientific mapping may not represent the current affiliation of some authors. Finally, through this scientific mapping of the publications of IJIMAI, we want to give them our congratulations and encourage them to continue contributing and supporting the development of the various disciplines of interest to the journal.

Acknowledgements

This project was funded by the FEDER financial support from the Project TIN2016-75850-P provided by the Spanish Ministry of Science, Innovation and Universities. The last author acknowledges support from the Fondecyt Regular program (project number 1160286) of the Chilean Government through Conicyt.

References

[1] Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F., 2011. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. Journal of Informetrics 5, 146–166.

[2] Van Raan, A., 2005. Handbook of Quantitative Science and Technology Research. Chapter Measuring Science. Springer Netherlands.

[3] Kessler, M.M., 1963. Bibliographic coupling between scientific papers. American Documentation 14, 10–25.

[4] Small, H., 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science 24, 265–269.

[5] Van Eck, N.J., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84, 523–538.

[6] Ding, Y., Rousseau, R., Wolfram, D., 2014. Measuring scholarly impact : methods and practice, Firs Editi. ed. Springer, Cham, Heidelberg, New York, Dordrecht, London.

[7] Podsakoff, P.M., MacKenzie, S.B., Podsakoff, N.P., Bachrach, D.G., 2008. Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. Journal of Management 34, 641–720.

[8] Bonilla, C., Merigó, J.M., Torres-Abad, C. 2015. Economics in Latin America: A bibliometric analysis. Scientometrics 105, 1239–1252.

[9] Baier-Fuentes, H., Merigó, J.M., Amorós, J.E., Gaviria-Marin, M., 2018. International entrepreneursip, an overview from bibliometric analysis. International Entrepreneurship and Management Journal, https://doi.org/10.1007/s11365-017-0487-y.

[10] Gaviria-Marin, M., Merigó, J.M., Baier-Fuentes, H., 2018. Knowledge management: a global examination based on bibliometric analysis, Technological Forecasting and Social Change, https://doi.org/10.1016/j.techfore.2018.07.006.

[11] Schwert, G.W., 1993. The Journal of Financial Economics A retrospective evaluation (1974-91)*. Journal of Financial Economics 33, 369–424.

[12] Merigó, J.M., Pedrycz, W., Weber, R., de la Sotta, C., 2018. Fifty years of Information Sciences: A bibliometric overview. Information Sciences 432, 245–268.

[13] Merigó, J.M., Blanco-Mesa, F., Gil-Lafuente, A.M., Yager, R.R., 2017. Thirty Years of the International Journal of Intelligent Systems: A Bibliometric Review. International Journal of Intelligent Systems 32, 526–554.

[14] Cobo, M.J., Martínez, M.A., Gutiérrez-Salcedo, M., Fujita, H., Herrera-Viedma, E., 2015. 25 years at Knowledge-Based Systems: A bibliometric analysis. Knowledge-Based Systems 80, 3–13.
[15] Biemans, W., Griffin, A., Moenaert, R., 2007. Twenty Years of the Journal of Product Innovation Management: History, Participants, and Knowledge Stock and Flows. The Journal of Product Innovation Management 24, 193–213.
[16] Gaviria-Marín, M., Merigó, J.M., Popa, S. 2018. Twenty years of the Journal of Knowledge Management: A bibliometric analysis, Journal of Knowledge Management 22, 1655-1687.
[17] Cancino, C., Merigó, J.M., Coronado, F., Dessouky, Y., Dessouky, M. 2017. Forty years of Computers & Industrial Engineering: A bibliometric analysis. Computers & Industrial Engineering 113, 614–629.
[18] Merigó, J.M., Cobo, M.J., Laengle, S., Rivas, D., Herrera-Viedma, E., 2018. Twenty years of Soft Computing: a bibliometric overview. Soft Computing In Press. doi:10.1007/s00500-018-3168-z.
[19] Pritchard, A., 1969. Statistical Bibliography or Bibliometrics? Journal of Documentation 25, 348–349.
[20] Small, H., 1999. Visualizing science by citation mapping. Journal of the American Society for Information Science 50, 799–813.
[21] Morris, S.A., Van der Veer Martens, B., 2009. Mapping research specialties. Annual Review of Information Science and Technology 42, 213–295.
[22] Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F., 2011. Science mapping software tools: Review, analysis, and cooperative study among tools. Journal of the American Society for Information Science and Technology 62, 1382–1402.
[23] Chen, C., 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology 57, 359–377.
[24] Persson, O., Danell, R., Wiborg Schneider, J., 2009. How to use Bibexcel for various types of bibliometric analysis, in: Åström, F., Danell, R., Larsen, B., Schneider, J. (Eds.), Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at His 60th Birthday. International Society for Scientometrics and Infometrics, Leuven, pp. 9–24.
[25] Porter, A.L., Cunningham, S.W., 2005. Tech mining: exploiting new technologies for competitive advantage. John Wiley & Sons Inc., Hoboken, NJ.
[26] Tur-Porcar, A., Mas-Tur, A., Merigó, J.M., Roig-Tierno, N., Watt, J. 2018. A bibliometric history of the Journal of Psychology between 1936 and 2015. Journal of Psychology 152, 199-225.
[27] Valenzuela, L., Merigó, J.M., Johnston, W., Nicolás, C., Jaramillo, F. 2017. Thirty years of the Journal of Business & Industrial Marketing: A bibliometric analysis, Journal of Business & Industrial Marketing 32, 1-18.
[28] Blanco-Mesa, F., Merigó, J.M., Gil-Lafuente, A.M. 2017. Fuzzy decision making: A bibliometric-based review. Journal of Intelligent & Fuzzy Systems 32, 2033–2050.
[29] Martínez-López, F.J., Merigó, J.M., Valenzuela, L., Nicolás, C. 2018. Fifty years of the European Journal of Marketing: A bibliometric analysis. European Journal of Marketing 52(1-2), 439–468.
[30] Peters, H.P.F., van Raan, A.F.., 1991. Structuring scientific activities by co-author analysis: An exercise on a university faculty level. Scientometrics 20, 235–255.
[31] Callon, M., Courtial, J.P., Turner, W.A., Bauin, S., 1983. From translations to problematic networks: An introduction to co-word analysis. Social Science Information 22, 191–235.
[32] Mulet-Forteza, C., Martorell-Cunill, O., Merigó, J.M., Genovart-Balaguer, J., Mauleon-Mendez, E. 2018. Journal of Travel & Tourism Marketing: A bibliometric analysis, Journal of Travel & Tourism Marketing 35, 1201-1221.
[33] Wang, W., Laengle, S., Merigó, J.M., Yu, D., Herrera-Viedma, E., Cobo, M.J., Bouchon-Meunier, B. 2018. A bibliometric analysis of the first twenty-five years of the International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 26, 169–193.
[34] Huang, Y., Zhu, D., Lv, Q., Porter, A.L., Robinson, D.K.R., Wang, X., 2017. Early insights on the Emerging Sources Citation Index (ESCI): an overlay map-based bibliometric study. Scientometrics 111, 2041–2057.
[35] Gutiérrez-Salcedo, M., Angeles Martínez, M., Moral-Munoz, J., Herrera-Viedma, E., Cobo, M., 2018. Some bibliometric procedures for analyzing and evaluating research fields. Applied Intelligence 48, 1275–1287.
[36] Cancino, C.A., Merigó, J.M., Coronado, F.C., 2017. A bibliometric analysis of leading universities in innovation research. Journal of Innovation & Knowledge 2, 106–124.
[37] Laengle, S., Merigó, J.M., Miranda, J., Słowiński, R., Bomze, I., Borgonovo, E., Dyson, R.G., Oliveira, J.F., Teunter, R., 2017. Forty years of the European Journal of Operational Research: A bibliometric overview. European Journal of Operational Research 262, 803–816.
[38] Van Eck, N.J., Waltman, L., 2014. Visualizing Bibliometric Networks, in: Ding, Y., Rousseau, R., Wolfram, D. (Eds.), Measuring Scholarly Impact: Methods and Practice. Springer, pp. 285–320.

Hugo Baier-Fuentes

Hugo Baier-Fuentes is a PhD in Business by the University of Barcelona (Spain). Currently he is an Assistant Professor and Academic Director MBA at the Faculty of Economics and Business Administration at the Universidad Católica de la Santísima Concepción, Chile. He also holds a Bachelor in Industrial Engineering from the Adventist University of Chile. He has published in several international journals including Technological Forecasting and Social Change, European Journal of International Management and International Entrepreneurship and Management Journal. His research includes the study of bibliometrics, international entrepreneurship and the international strategy of Latin American small firms. His works have been presented in the leading conferences of Spain, including Spanish Academy of Management (ACEDE), and the Academy of International Business Conference.

Jesús-Daniel Cascón-Katchadourian

Jesús-Daniel Cascón-Katchadourian is PhD in Documentation Studies and Master in Documentation and Scientific Communication. He has a degree in History and another in Documentation from the University of Granada. He received the first national award by the Ministry of Education, Culture and Sport. His interest areas are related with cartography, georeferencing and geolocation of historical photography, bibliometric studies and information retrieval. He currently enjoys a postdoctoral contract at the University of Granada.

María-Ángeles Martínez-Sánchez

María-Ángeles Martínez-Sánchez is a PhD in Social Sciences, Master in Information and Scientific Communication, and graduate in Social Work from the University of Granada. She is an Associate professor in the Department of Social Work and Social Services. Her most recent research revolves around science analysis, bibliometric studies, digital libraries, and social services. She is author of 10 JCR publications and different international conference contributions.

Enrique Herrera-Viedma

Enrique Herrera-Viedma is Professor in Computer Science and A.I in University of Granada and currently the new Vice-President for Research and Knowledge Transfer and Vice-President for Publications in IEEE SMC Society. His current research interests include group decision making, consensus models, linguistic preference modeling, aggregation of information, information retrieval, bibliometrics, digital libraries, web quality evaluation, recommender systems, social networks, and social media. In these topics he has published more than 250 papers in JCR journals and coordinated more than 20 research projects. Dr. Herrera-Viedma is member of the government board of the IEEE SMC Society and an Associate Editor of several international journals such as the IEEE Trans. On Syst. Man, and Cyb.: Systems, IEEE Trans. On Fuzzy Systems, Knowledge Based Systems, Soft Computing, Fuzzy Optimization and Decision Making, Applied Soft Computing, Journal of Intelligent and Fuzzy Systems, and Information Sciences. Since 2014 he has been selected as a Highly Cited Author in both categories Computer Science and Engineering by Clarivate Analytics.

José M. Merigó

José M. Merigó is currently a Full Professor at the School of Information, Systems and Modelling at the Faculty of Engineering and Information Technology at the University of Technology Sydney and a Part-Time Full Professor at the Department of Management Control and Information Systems at the School of Economics and Business at the University of Chile. His research is focused on computational intelligence, decision theory, aggregation operators, bibliometrics and uncertainty. He has been a Senior Research Fellow at the Manchester Business School of the University of Manchester (UK) and Senior Assistant Professor at the University of Barcelona at the Department of Business Administration. He has published >500 papers including 20 books and >200 publications indexed in Web of Science Core Collection. He is an editorial board member of several international journals including Computers & Industrial Engineering, Applied Soft Computing, Kybernetes, Technological and Economic Development of Economy, International Journal of Fuzzy Systems and Journal of Intelligent & Fuzzy Systems. Since 2015 he has been selected as a Highly Cited Author in Computer Science by Clarivate Analytics.

Revisiting "Recognizing Human Activities User-Independently on Smartphones Based on Accelerometer Data" – What Has Happened Since 2012?

Pekka Siirtola*, and Juha Röning

Biomimetics and Intelligent Systems Group, University of Oulu (Finland)

Received 19 November 2018 | Accepted 26 November 2018 | Published 1 December 2018

unir
LA UNIVERSIDAD EN INTERNET

Abstract

Our article "Recognizing human activities user-independently on smartphones based on accelerometer data" was published in the International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI) in 2012. In 2018, it was selected as the most outstanding article published in the 10 years of IJIMAI life. To celebrate the 10th anniversary of IJIMAI, in this article we will introduce what has happened in the field of human activity recognition and wearable sensor-based recognition since 2012, and especially, this article concentrates on introducing our work since 2012.

Keywords

Accelerometers, Wearable Sensors, Human Activity Recognition, Machine Learning.

DOI: 10.9781/ijimai.2018.12.001

I. Introduction

HUMAN activity recognition is a field of science where classification methods are applied to inertial sensor data to recognize human activities. Some early preliminary activity recognition studies had already been done in the 1990's (such as [1, 2]) but [3] can be considered as the first proper inertial sensor -based activity recognition article. It concentrated on recognizing daily activities using accelerometers, and how clothes could be made aware of context.

Our human activity recognition article "Recognizing human activities user-independently on smartphones based on accelerometer data" [4] was published in IJIMAI in 2012. For the study, a daily activity data set were collected from eight healthy subjects. The trousers' front pocket was fixed as the phone placement, but the subject was allowed to determine whether the phone was placed in the left or right pocket. The participants performed five different activities: walking, running, cycling, driving a car, and idling, that is, sitting/standing. The total amount of the data collected was about four hours. These activities were selected for the study because normal everyday life consists mainly of these five activities.

Study used a window length of 7.5 seconds, and altogether 42 features were extracted from windows. These included for instance standard deviation, mean, minimum, maximum, five different percentiles (10, 25, 50, 75, and 90), and a sum and square sum of observations above/below certain percentile (5, 10, 25, 75, 90, and 95). The classification was obtained using a two stage procedure. In the first classification stage, a model was trained to decide if the studied subject was active (walking, running or cycling) or inactive (driving a car or idling). In the second stage, the exact activity label was obtained, and therefore, one model was trained to classify an active activity as walking, running or cycling, and the other to classify an inactive activity as idling or driving.

* Corresponding author.
E-mail address: pekka.siirtola@oulu.fi

The models were trained offline using the collected daily activity data set. In addition, these models were implemented to smartphones (Symbian^3- and Android-phones) and also used in online tests. To compare different classifiers, the classification was performed using two different classification methods, kNN (k nearest neighbours) and QDA (quadratic discriminant analysis). The most descriptive features for each model were selected using a sequential forward selection (SFS). QDA classifiers for offline and online recognition were trained using the whole training data set, similar to kNN classifier for the offline recognition. However, because of the limited computational power of the smartphone, the activity recognition on the device using kNN was performed using only a limited number of randomly chosen instances from training data.

The offline recognition results show that the method enables accurate results. Each activity is recognized with high accuracy. The average classification accuracy using QDA was 95.4% and using kNN, 94.5%.

For the online experiments, the application for real-time classification was tested by seven persons carrying Nokia N8 smartphone in their trousers' front pocket. In addition, classification on Android device was tested by five subjects, again, carrying the phone on their trousers' front pocket. Based on the experiences gathered using Nokia phones, only QDA classifier was implemented to Android-phone. The online recognition rates were almost identical to offline results. On a device running Symbian^3 operating system the average classification accuracy using QDA was 96.2% and using kNN, 94.1%, and on Android phone the recognition accuracy using QDA was 94.5%.

The article showed that user-independent activity recognition works reliably and operating system independently. In fact, it was one of the first articles showing that human activities can be recognized reliably in real-time in real-life conditions using smartphone hardware and smartphone sensors. The article got immediately a positive reception from the research community, and so far, according to Google Scholar, it has been cited 146 times making it the most cited article ever published in IJIMAI (numbers checked 14.11.2018). In

addition, it was selected as the most outstanding article published in the 10 years of IJIMAI life. In addition, it was a key part of Dr. P. Siirtola's Doctoral thesis "Recognizing human activities based on wearable inertial measurements: methods and applications" [5] which was published in 2014. This article introduces how the field of human activity recognition and recognition based on wearable sensor data has changed, and especially, how the research of the authors of [1] has progressed since 2012.

II. Ready-To-Use Activity Recognition

A sequel to [4] was presented in "Ready-to-Use Activity Recognition for Smartphones" [6]. It extended the work by introducing body-position independent human activity recognition method. This means that while in [4] the phone position was fixed as trousers' pocket, in [6] there was more options as the phone position. In fact, a data set containing data from five body positions (trousers' front pocket, jacket's pocket, at backpack, at brachium, and at the ear) were gathered for this study. The participants performed five different activities: walking, running, cycling, driving a car, and sitting/standing. However, there is no data from each activity from each body position. For instance, subjects were not allowed to cycle while holding a phone at the ear because of safety issues. Moreover, data were collected when a phone was laying on the table. Therefore, six activities were recognized. The total amount of the data collected was about fifteen hours.

A 3D accelerometer is a sensor consisting of three accelerometers that are approximately perpendicular to each other. However, as a side results, it was noted in the article that in reality, in each phone these sensors are aligned a bit differently due to manufacturing differences. Because of this difference, accelerometer values from different phones are at a different level, and this leads to differences in measured values and, eventually, misclassifications. On the other hand, it was noted that despite the differences in sensor values, the shapes of the signals are approximately the same but absolute values differ by some constant. This difference is normally fixed by using automated or user-driven calibration. However, because the difference in the signal level is the only major difference caused by different calibration, in our study, differences were eliminated simply by subtracting the mean of the window's values from each value of the window. This way, calibration differences can be eliminated already in the feature extraction phase and automated or user-driven calibration was not needed.

For the model training, 19 features were extracted from magnitude signal and from the signals combining two out of three acceleration channels after calibration differences were eliminated, so together 76 features were extracted and models for offline and online experiments were trained using these. QDA was decided to be used as a classifier, and in order to achieve the highest possible recognition rates, the most descriptive features for each model were selected using SFS method. Moreover, to obtain reliable user-independent results, the training was performed using the leave-one-out method, so that each person's data in turn was used for testing and the rest of the data were employed for model training.

Offline experiments showed that using the presented method, it was possible to recognize activities with a high accuracy user-independently. In addition, all the activities were detected almost perfectly from each body-position. Moreover, hardware variations between phones was not an issue: training data was collected using five different smartphones and activities from each phone were detected reliably.

Again, online experiments were performed using phones running two different operating systems, Symbian^3 and Android. Online recognition accuracies were also high, however, not as high as offline recognition results. The difference between recognition rates of online and offline scenarios was most likely caused by real-life situations from which there was no training data. Moreover, it was not only noted that recognition can be done body position –independently without caring hardware differences, it was also noted that mobile application was light. In fact, Symbian^3 -version of the application used only around 15% of CPU capacity of the Nokia N8, and thus, the application did not eat too much battery.

III. Towards Personalized Activity Recognition Models

The main weakness of articles [4] and [6] was that they were based on user-independent models. When dealing with data collected from humans, the challenge is that people are different: they are unique for instance in terms of physical characteristics, health state or gender. All of these affect to the inertial data that are measured. In fact, it is shown that user-independent models do not work accurately for instance if they are trained with healthy study subjects and tested with subjects who have difficulties to move [7]. One more challenge is real-life, real-time conditions. It has been shown that when models that work in laboratory conditions are used in real-life conditions outside the laboratory, the results can be far from excellent [8]. In such cases, the recognition model is not general enough, and therefore, it cannot react to the changing and unseen conditions. It is especially relevant to overcome problems arising from real world conditions when the aim is to build a model that is used outside laboratory. In fact, non-stationary environments are considered one of the modern machine learning's greatest challenges [9]. Therefore, to be able to use models in real life problems, model used in the recognition process should be non-stationary instead of stationary. Moreover, because of problems related to different types of people, the focus of research should be on personal and personalized prediction models instead of user-independent models. However, the challenge of personal and personalized models is that they require personal training data. This normally would require an extensive, separate data collection session for each user.

A. From User-Independent to Personal Human Activity Recognition Models Using Smartphone Sensors

One method to obtain personalized recognition models without user-interruption was presented in [10]. The study presented a method for smartphones to obtain light weight personalized human activity recognition models unobtrusively by using the sensors of a smartphone. The proposed method consisted of four phases:

1. In the first phase, sensor fusion-based recognition model is trained and used to recognize activities from the streaming data. To maximize the recognition rate of this model, it is trained using a large number of features and these features can be based on more than one type of sensors of a smartphone (for instance accelerometers, GPS values, gyroscopes, and magnetometers).
2. When classifying streaming data using sensor fusion-based user-independent model, it can be assumed that recognition process is reliable leading to reliable classification results. Therefore, by combining these recognition results, and using them as labels, and the data related to them, a personal training data set can be gathered without user-interruption.
3. When personal data from each of the recognized activities is available, a new user-dependent recognition model can be trained. In order to make this personal recognition model light, only a small number of features extracted from a one sensor are used to train the model.
4. Streaming data can then be classified using a computationally efficient, single sensor-based user-dependent model.

The data used in the study was the same as the one used in [6], but in this case both accelerometer and magnetometer data was used.

Therefore, the user-independent model used in phase 1 was trained using features extracted from both accelerometer and magnetometer signals, and personal user-dependent model trained in phase 3 was trained using accelerometer reading only, making it more energy efficient than user-independent model.

Experiments were done with LDA (linear discriminant analysis), and QDA classifiers, and the experiments showed that the presented method improved classification accuracy when compared to traditional user-independent model. In fact, the recognition accuracy improved in nine tested cases out of ten, on average the improvement varied from 3 to 4 %-units.

B. Personalizing Human Activity Recognition Models Using Incremental Learning

Using the approach presented in [10], recognition accuracy can be improved but the problem of the approach is that personalization is based on model re-training. Therefore, in order to build personal recognition models, all the streaming data needs to be kept stored. This is problematic as it requires a lot memory, and model re-training requires a lot calculation capacity.

Incremental learning refers to recognition methods that can learn from online information and adapt to new environments. The advantage of incremental learning is that this adaptation can be done without model re-training and user-interruption. Instead, the idea is that models can be updated, instead of re-training, automatically based on streaming data [11]. Therefore, in the case of human activity recognition, incremental learning can not only be used to adapt to new environments, but also to data of new unseen person to build personalized recognition models.

Our method to personalize human activity recognition models using incremental learning was presented in [12]. In fact, our study does not only show that personalization based on incremental learning improves the recognition rates compared to results of user-independent model, it also compares three base classifiers: LDA, QDA and CART (classification and regression tree).

The experiments were made using publicly open data set [13]. This data contains data from seven physical activities (walking, sitting, standing, jogging, biking, walking upstairs and downstairs). Data set contains measurements from 10 study subjects. However, only nine persons data were used in the experiments, as apparently one of the study subjects had placed sensor in different orientation than others making the data totally different to other subjects' data. In fact, because of this problem, and other problems that can be found from publicly open datasets [14], in 2018 we introduced OpenHAR [15] which is a MATLAB toolbox to provide easy access to 10 publicly open human activity datasets.

In [12], incremental learning was based on Learn++ algorithm [16]. Learn++ is an ensemble method where the idea is to process incoming streaming data not as single values but instead as chunks. For each chunk, a new group of weak base models are trained and combined to a group of previously trained base models through weighted majority voting as ensemble model [17]. The following idea was used in our study to personalize models: in the first place, user-independent models were trained and added to ensemble. When streaming data from a new unseen person were obtained, ensemble model was used to label this new data. This data and predicted labels can then be used to train personal recognition models which can be added to ensemble. This means that instead of re-training the whole recognition model, personalized models can be obtained by updating the existing model by adding new base models to ensemble. Therefore, once the new base models were added to ensemble, the data used to train these base models were no longer needed, and it can be erased from the memory. This makes the approach very efficient computationally. Moreover, every time a new base model was added to ensemble, the model becomes more personal, and simultaneously this continuous learning also enables a solution to adapt to new environments and unseen situations. However, problem with the data chunks used to personalize the recognition process was that they were small. Thus, they did not contain much variation leading easily to over-fitted models. To avoid over-fitting, noise injection method we presented in [18] was applied to training data sets to increase the size of training data and increase its variation.

According to the experiments shown in [12], in most cases personalization reduces the average error rate: when new base models were added to Learn++, error rates decreased. In fact, the improvement was significant: QDA and CART improved results in 7 cases out of 9 and LDA with all study subjects. With CART the average error rate dropped from 18.0% to 15.7% (13.1% improvement), with LDA from 14.1% to 9.5% (33.1% improvement), and with QDA from 11.1% to 9.1% (17.9% improvement). Comparison was made to user-independent model. Therefore, while the average error rate using QDA was the smallest, the biggest benefit from personalization can be achieved when LDA was used as a base classifier.

IV. Using Human Activity Recognition Methods In Health Applications

Back in 2012, most of the activity recognition studies were based on inertial sensor data, as devices did not normally include any other sensors. However, nowadays the situation is different. For instance, wrist-worn Empatica E4 device includes not only inertial sensor, but also thermometer, electrodermal activity sensor, which is used to measure galvanic skin response, and photoplethysmography sensor, which can measure blood volume pulse, heart rate, and heart rate variability. This has opened new research possibilities, and nowadays, similar methods that were used to train human activity recognition models can be used to train models for health and medical application.

In [19] we used data collected using Empatica E4 to early detection of migraine attacks. Data was collected from seven volunteer study subjects. Five of them were women and two were men, and the age of the study subjects varied from 30 to 60 years. They had different types of migraines, for instance five of them had aura symptoms, while two did not have. Moreover, most of them did not use preventive medication. However, all of them used medication during the migraine attacks. All of the study subjects had migraine attacks quite often. In fact, this was a criterion for joining the study as it helped the data gathering process. Frequent attacks enabled a shorter data collection period, which ensured that data consisted of several migraine attacks for every study subject. Data gathering session was long; study subjects wore Empatica E4 on their non-dominant hand approximately 27 days. Altogether, data set included 200 days of data.

When data was pre-processed, it was noted that the quality of signals was not good during the daytime due to physical activity which caused disturbances to signals. Due to this problem, only sleep time data was used in the study. Moreover, sleep time data were divided into two classes: (1) nights before a day without a migraine attack and (2) nights before a day with a migraine attack, and therefore, class (2) contains information and measured values about the pre-ictal stage of a migraine attack. The idea behind this approach is to inform the user after he/she wakes up in the morning if he/she will have a migraine attack that day, and therefore, user can take predictive medication if needed.

Similar to human activity recognition studies, features were extracted from sleep time data by considering one night as one window. Altogether, 110 features were extracted from sleep data, these included for instance standard deviation, mean, max, min, different percentiles from each signal, as well as correlations between different signals. However, the problem was that though the data set was extensive, 27 nights of data per

study subject, considering one night as one window compressed the data set so much that reliable models cannot be built based on it. Moreover, the data set was imbalanced, set includes only a few nights of data from class (2), and most of the data were from class (1).

In order to increase the number of observations, it was decided to not base the recognition on the features extracted from each night. Instead, we used the differences between nights as features. Differences were calculated so that (1) nights before a migraine attack were compared to nights before a day without a migraine, and (2) nights before a day without a migraine were compared with each other. This approach increased the number of data significantly: before this procedure, data set had 200 samples (data from 200 nights), while after applying this approach, the number of observations was 2265. In addition, to avoid over-fitting the number of observations were increased using our noise injection presented in [18].

Experiments were again made using LDA and QDA classifiers. The results showed that migraine attacks cannot be detected beforehand using user-independent model. However, the results using personal models were encouraging: balanced accuracy for detecting attacks one night prior was 70% using LDA, and as high as 84% using QDA. While the average detection rate using QDA was high, the results also show that balance accuracy varies greatly between study subjects (from 60% to 95%), which shows how complicated the problem actually is. In fact, the future work is to determine reasons for this variation. For instance, it is possible that some migraine types are more difficult to predict than others are, or it is possible that our data set was not comprehensive enough to build reliable models for different study subjects.

V. Conclusion

A lot has happened in the field of human activity recognition using wearable sensors since 2012 when our article "Recognizing human activities user-independently on smartphones based on accelerometer data" was published in IJIMAI. Our article was one of the first articles where human activity recognition was done in real-time on smartphone using smartphone's own sensors. Therefore, it had its own small role in shaping the field as it is now.

While the field has developed, also our research has progressed and diversified: we have moved from stationary recognition models to models that enable continuous life-long learning, from user-dependent models to personalized models, and from movement recognition to more comprehensively measuring humans, which enables understanding what is happening inside human body.

The future of human activity recognition, and especially, recognition based on data from wearable sensors, looks interesting. Wearable devices and their market develop rapidly, and new sensors are introduced to devices which enables new types of applications. Moreover, market development enables bigger user tests for researchers, and more business opportunities for application developers.

References

[1] J. Farringdon, A. Moore, N. Tilbury, J. Church & P. Biemond: Wearable sensor badge and sensor jacket for context awareness. Proc. Wearable Computers, 1999. Digest of Papers. The Third International Symposium on, 107–113.

[2] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz & M. Depairon (1999) Physical activity monitoring based on accelerometry: validation and comparison with video observation. Medical & Biological Engineering & Computing 37(3): 304–30.

[3] K. Van Laerhoven & O. Cakmakci (2000) What shall we teach to our pants? Wearable Computers, 2000. The Fourth International Symposium on, 77–83.

[4] P. Siirtola & J. Röning (2012) Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data. International Journal of Interactive Multimedia and Artificial Intelligence 1(5): 38-45.

[5] P. Siirtola. Recognizing Human Activities Based on Wearable Inertial Measurements-Methods and Applications. Doctoral dissertation, Department of Computer Science and Engineering, University of Oulu, (Acta Univ Oul C 524), 2015.

[6] P. Siirtola & J. Röning (2013). Ready-to-use activity recognition for smartphones. In Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on (pp. 59-64). IEEE.

[7] M. Albert, S. Toledo, M. Shapiro, and K. Kording (2012) Using mobile phones for activity recognition in parkinsons patients, Frontiers in neurology, vol. 3, 2012.

[8] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola & I. Korhonen (2006). Activity classification using realistic data from wearable sensors. IEEE Transactions on information technology in biomedicine 10(1): 119-128.

[9] M. Sugiyama & M. Kawanabe (2012). Machine learning in non-stationary environments: Introduction to covariate shift adaptation. MIT Press.

[10] P. Siirtola, H. Koskimäki & J. Röning: From User-independent to Personal Human Activity Recognition Models Exploiting the Sensors of a Smartphone, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2016., Bruges, Belgium 27-29 April 2016, 471--476.

[11] A. Gepperth & B. Hammer (2016). Incremental learning algorithms and applications. In European Symposium on Artificial Neural Networks (ESANN).

[12] P. Siirtola, H. Koskimäki & J. Röning: Personalizing human activity recognition models using incremental learning. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2016., Bruges, Belgium 25-27 April 2018, 627--632.

[13] M. Shoaib, S. Bosch, O.D. Incel, H. Scholten. and P.J.M Havinga. (2014) Fusion of Smartphone Motion Sensors for Physical Activity Recognition. Sensors, 14, 10146-10176.

[14] P. Siirtola, H. Koskimäki & J. Röning: Experiences with Publicly Open Human Activity Data Sets - Studying the Generalizability of the Recognition Models, Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, 291--299, 2018.

[15] P. Siirtola, H. Koskimäki & J. Röning: (2018, October). OpenHAR: A Matlab Toolbox for Easy Access to Publicly Open Human Activity Data Sets. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (pp. 1396-1403). ACM.

[16] R. Polikar, L. Upda, S. S Upda, & V. Honavar. Learn++: An incremental learning algorithm for supervised neural networks. IEEE transactions on systems, man, and cybernetics, part C (applications and reviews), 31(4):497–508, 2001.

[17] V. Losing, B. Hammer, & H. Wersing. Choosing the best algorithm for an incremental on-line learning task. In European Symposium on Artificial Neural Networks (ESANN), April 27-29 2016.

[18] P. Siirtola, H. Koskimäki & J. Röning: (2016). Personal models for eHealth-improving user-dependent human activity recognition models using noise injection. In Computational Intelligence (SSCI), 2016 IEEE Symposium Series on (pp. 1-7). IEEE.

[19] P. Siirtola, H. Koskimäki, H. Mönttinen & J. Röning: (2018). Using Sleep Time Data from Wearable Sensors for Early Detection of Migraine Attacks. Sensors (Basel, Switzerland), 18(5).

Pekka Siirtola

Pekka Siirtola received his M.Sc. degree in Mathematics in February 2006 and Dr. Tech degree in April 2015, both from University of Oulu, Oulu, Finland. Currently, he is working as a post-doc researcher at the Biomimetics and Intelligent Systems Group at the University of Oulu. His research interests include data mining, and machine learning, and especially, how data from wearable sensors can be used in health and wellness applications.

Juha Röning

Juha Röning is Professor of Embedded System at the University of Oulu. He serves also as Visiting Professor of Tianjin University of Technology, P. R. China. He is principal investigator of the Biomimetics and Intelligent Systems Group (BISG). In 1985 he received Asla/ Fullbright scholarship. From 1985 to 1986 he was a visiting research scientist in the Center for Robotic Research at the University of Cincinnati. From 1986 to 1989 he held a Young Researcher Position in the Finnish Academy. Prof. Röning was nominated as Fellow of SPIE in 2000. He has two patents and has published more than 300 papers in the areas of computer vision, robotics, intelligent signal analysis, and software security. He is currently serving as a Board of Director for euRobotics aisbl. He is also a steering board member of ARTMIS-IA.

Generation of Two-Voice Imitative Counterpoint from Statistical Models

Victor Padilla[1]*, Darrell Conklin[2]

[1] UNIR, Universidad Internacional de La Rioja, Logroño (Spain)
[2] Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, San Sebastian, Spain and IKERBASQUE: Basque Foundation for Science, Bilbao (Spain)

Received 24 July 2018 | Accepted 18 September 2018 | Published 6 October 2018

Abstract

Generating new music based on rules of counterpoint has been deeply studied in music informatics. In this article, we try to go further, exploring a method for generating new music based on the style of Palestrina, based on combining statistical generation and pattern discovery. A template piece is used for pattern discovery, and the patterns are selected and organized according to a probabilistic distribution, using horizontal viewpoints to describe melodic properties of events. Once the template is covered with patterns, two-voice counterpoint in a florid style is generated into those patterns using a first-order Markov model. The template method solves the problem of coherence and imitation never addressed before in previous research in counterpoint music generation. For constructing the Markov model, vertical slices of pitch and rhythm are compiled over a large corpus of dyads from Palestrina masses. The template enforces different restrictions that filter the possible paths through the generation process. A double backtracking algorithm is implemented to handle cases where no solutions are found at some point within a generation path. Results are evaluated by both information content and listener evaluation, and the paper concludes with a proposed relationship between musical quality and information content. Part of this research has been presented at SMC 2016 in Hamburg, Germany.

Keywords

Artificial Intelligence, Music Informatics, Music Generation, Sequential Pattern Mining, Statistical Models Of Music.

DOI: 10.9781/ijimai.2018.10.002

I. Introduction

In music history, when a style ends and is replaced by a new way of thinking, theorists and musicologists try to explain the "obsolete" musical ideas. It is at this moment when the old style is classified, and the rules of the old style are systematized for future generations of musicians. According to Jeppesen [1], the history of music theory and style are far from being identical, and it is important to take into account the constantly recurring mistakes of the theorists with regard to the description of musical style. The principal inaccuracies of theorists according to Jeppesen [1] are:

1. An inclination that is common to these writers to theorize on their own account.
2. The moment of inertia which causes theorists to transfer rules from older textbooks to new without proper critical revision.
3. Inability of theorists, when describing the practices of past times, to discriminate between these and the elements of style typical of their own contemporaries, (which was the case with Fux).
4. Pedagogic considerations, which often tend to a simplification or relaxation of the set of rules belonging to the style, but often also to a stricter rendering of these rules "for the sake of exercise".

Perhaps the most famous example of this theoretical approach is the "Gradus Ad Parnassum" of Johann Joseph Fux (1725). Fux presents a pedagogical method that breaks the learning task into well-defined graduated stages, from note against note through to florid counterpoint. This continues to be a standard pedagogic counterpoint text. However, for modern music informatics, generating music based on such stylistic rules is not a good approach: in fact, music from Renaissance to Romanticism can be written following basically the same rules. For example, stylistic differences between Bach and Palestrina counterpoint cannot be completely defined by basic generic rules (for example, prohibition of parallel fifths or octaves, voice leading at suspensions, etc.), and implementing specific exceptions to them can be a very complex task. The rule-based system achieved will become brittle and imprecise. Musical style must be learned from examples in order to model as closely as possible a musical style.

The corpus of pieces we are working with comprises 717 movements from Palestrina masses, comprising almost 350,000 vertical dyads (slices), providing a massive amount data for training statistical models of counterpoint. Even though these data are naturally within the musical style, composing a piece of music in the Renaissance style is not so simple as performing random walks through a statistical model. Counterpoint is full of imitations, canons, motifs and augmentations, and such devices cannot be captured by a first-order Markov model trained on limited data [2]. For solving these limitations and to provide coherence to the generated pieces, we take a piece from the corpus, referred to as a *template*, and discover its repeated patterns based on different viewpoints. The discovered patterns are used to cover the template piece and are adhered to in new music generated by the statistical model.

There are several arguments for using the masses of Palestrina as a test collection for our system. They are a model for a standard Renaissance style in counterpoint. Many universities and conservatories teach

* Corresponding author.
E-mail address: victor.padilla@unir.net

this style as basic training for new students in composition. Another important aspect is the homogeneity of the corpus of pieces. There are no significant differences in style between the first and the last mass, and the number of pieces is big enough to build a probabilistic model.

At the end of the article, we present some methods for the evaluation of the results. To assess an automatic system of musical composition is always a difficult task and in this paper we use two diverse alternatives: a listener evaluation in a live performance setting, and an objective information content method. For the listener evaluation, even when asking an expert group in a "Turing test" setting, the results cannot be very reliable, as they depend on many variables such as the performance, involvement of the experts, and their degree of knowledge of the musical style. An objective way to measure the relative quality of generated pieces is by using the information content of the generated piece, which is the negative log probability according to the statistical model.

This article is organized as follows: Section II reviews different works for generation and analysis of music. Section III describes the corpus of Palestrina masses. Section IV explores the concept of probability with respect to zero- and first-order Markov models of the Palestrina corpus. For obtaining patterns from a template, slicing the scores into dyads, and generating from a statistical model, we are using the concept of horizontal and vertical viewpoints developed and refined by Conklin [3]-[6]. Section V describes how to apply restrictions imposed by the template during music generation. Section VI presents and analyses the results obtained, and Section VII positions our work within previous research specifically on counterpoint analysis and generation.

II. Prior Work

Artificial Intelligence is a mature field having a broad variety of applications, ranging from driverless cars, natural language and speech processing, and computer players for board games, but is hard to define in formal terms for of Computational Creativity [7]. Often the scientist and the artist speak different languages and have different goals in mind, but this provides an interesting breeding ground for research with a lot of challenges to solve. Since the 1950s different techniques from artificial intelligence have been used for algorithmic composition and music generation. In this section we comment on some of the milestones of music generation, divided into two groups: rule-based approaches and machine learning approaches. In many cases, the lines between some methods are blurred and employ a mixture of both types of model creation.

A. Ruled-based

The generation of new music based on rules has a long tradition, from the works Hiller & Isaacson [8] using the ILLIAC computer at the University of Illinois. This early work was designed as a series of experiments on music composition, and rules for counterpoint were used in the generation of the first and second movements of the Illiac Suite. Rothgeb [9] encoded rules of eighteenth-century harmony for specifying adequate chords given bass notes. Many works on logic programming can be formulated as constraint satisfaction problems (CSPs). In this area, Ebcioglu [10] implemented rules for counterpoint translating constraints of fifth-species counterpoint to Boolean functions. In 1988 he subsequently developed 350 rules for the harmonic and melodic generation of Bach chorales [11]. Following Ebcioglu's work, several lines of research have been developed for harmony or counterpoint. Tsang & Aitken [12] harmonize four-part chorales, and Ovans & Davison [13] create a CSP system for the first-species counterpoint. Ramirez & Peralta [14] build a constraint logic programming system, for harmonizing a melody. Phon-Amnuaisuk [15] implemented a constraint system for harmonizing chorales in the style of J. S. Bach, adding control over the harmonization process in a flexible way. In recent years, Herremans & Sorensen [16] work with different counterpoint species using a variable neighborhood search algorithm. Komosinski & Szachewicz [17] address the difficulty of evaluating penalty (or reward) values for each broken (or satisfied) rule. The use of an additive function counting broken rules is known to have several drawbacks. First of all, it assumes that one can somehow determine the importance of each rule Another important drawback is that breaking one very important rule is equivalent to breaking several less important rules. Therefore, simple additive rule weighting function is found to be weak, and they propose to use a dominance relation. The implementation of rules by fuzzy logic to generate two-voice first species counterpoint is analysed by Yilmaz and Telatar [18].

Expanding the idea of "rule", a formal grammar may be viewed as a set of rules that expand high-level symbols into more detailed sequences of elements, in the same way that a language is constructed in a hierarchical structure of linguistic constituents. Some efforts to codify rules by hand and extend them to a grammar for automatic composition were made by Roads [19], Holtzman [20] and Jones [21]. In the 1990s some systems appear that hybridize rule-based systems with evolutionary algorithms using a fitness function from a rule set. McIntyre [22], Horner & Ayers [23] and Phon-Amnuaisuk et al. [24] worked in this direction in the context of four-part harmonization. For species counterpoint, Polito et al. [25] extracted rules and used them to define a fitness function followed by agents that cooperate to produce the composition. Gwee [26] worked with species counterpoint and a fitness function based on fuzzy rules.

B. Machine Learning

Machine learning is the subfield of computer science that studies the ability of computers to learn without being explicitly programmed. Many different machine learning techniques exist and for music generation learning is mainly unsupervised (without negative examples). In this section, we will survey some research on music generation using grammatical inference, Markov chains, and Artificial Neural Networks.

1) Grammatical Inference

As noted earlier, a grammar can be defined as a set of rules formally describing a language. The problem with a grammatical approach to composition is the difficulty in defining the rules manually. To tackle this issue, Cruz-Alcázar & Vidal-Ruiz [27] implemented several methods of grammatical inference inducing stochastic regular grammars to parse the compositions and make new pieces. Gilbert & Conklin [28] present a method to find tree structures in musical scores using a *probabilistic context-free grammar* for melodic reduction. The method is applied to parse phrases from Bach chorale melodies, and the statistical model is also used to evaluate the information content of the pieces. Following the ideas of Gilbert & Conklin, Groves [29] explores the generation of melodies from a probabilistic analytical model of melodies. Quick & Hudak [30] present a new class of generative grammars called *probabilistic temporal graph grammars* to handle temporal aspects of music in a way that retains a coherent metrical structure.

2) Markov Chains

Markov Markov chains are stochastic processes transitioning in discrete time steps through a finite set of states. In music composition, the transition matrices may be trained from a corpus of pre-existing compositions. In an order-n Markov chain, the next state depends on the last n states. In a hidden Markov model (HMM), the states are hidden and the goal is to infer an optimal state sequence for an observed sequence. This approach has been studied by Farbood &

Schoner [31] who train a second-order HMM to generate Palestrina-style first-species counterpoint to a specified cantus firmus line. The method uses Markov chains which capture the rules of counterpoint using probabilistic tables for harmony, melody, parallel motion, and cadences. Herremans et al. [31] use a first-order Markov model from a corpus of first species counterpoint and compare the ability of variable neighborhood search, iterative random walk and Gibbs sampling, to generate a hidden counterpoint line. Results are evaluated by information content (average negative log probability of the fragment using the dyad transitions of the transition matrix).

Working with HMMs and Bach chorales, Allan & Williams [33] create a system to compose four voice textures given a soprano part. The chord sequence is generated using the Viterbi algorithm, and for the passing notes, a second HMM is employed. Whorley & Conklin [34] present a multiple viewpoint system of four-part harmony to evaluate and improve an iterative random walk technique. This is evaluated using information content and also with a small set of rules of harmony.

3) Artificial Neural Networks

Artificial Neural Networks (ANNs) are biologically inspired models made of interconnected sets of nodes in several layers. Some of them are the input connections while others have output signals with several interconnected layers between.

Focusing on polyphony and counterpoint, Hild et al. [35] develop a model designed to solve a four-part chorale harmonization in Bach's style. The system was called HARMONET and had a three-layered architecture. An evolution of HARMONET was MELONET [36] and improved by Hörnel & Degenhardt [37]. Since then, many hybrid systems have been developed, for example NETNEG [38] which used an ANN trained with sixteenth-century classical music compositions. That generates melodic segments and polyphony was generated by a rule-based system of agents. Verbeurgt et al. [39] join Markov chains for constructing motifs and a trained ANN to assign the absolute pitches. Adiloglu & Alpaslan [40] used the back-propagation algorithm to generate two-voice first species counterpoint pieces. Some researchers have recently applied methods from deep learning to the chorale harmonization problem [41][42].

C. Generation Using Patterns

As suggested by Conklin [2], the construction of computational methods for musical style imitation has been far more difficult than initially imagined. Listening to music and perceiving its structure is an easy task for specially trained musicians but building computational models to mimic these processes is a hard problem. The use of patterns during generation can help to ensure coherence and intra-opus repetition in generated pieces. A method for the detection of melodic phrases in the masses of Palestrina is described by Knopke & Jurgensen [43], based on the use of suffix arrays to find repeated patterns. Sidorov et al.[44] present an approach to music analysis, in which an inferred grammar is explains the structure of a musical work.

Data mining is a process of extracting small pieces of valuable information from large data. A special situation is when data is in the form of sequences, and several sequential pattern mining methods have been developed in the last decade [45]. Using the idea of viewpoints [2]-[5], music can be converted into a string (or parallel strings) of features and analysed using sequential pattern mining methods. In our research, for analysing patterns, we are using the gap-BIDE algorithm [46] with zero gaps between sequences. This will be further explained in Section IV.B.

Some researchers have developed methods for combining patterns and constraints with Markov models. Pachet, Roy and Barbieri [47] try to solve unary and adjacent binary constraints with Markov models using arc-consistency techniques and re-normalization. Collins et al. [48] describe and evaluate a computational model of stylistic composition using discovered patterns to constrain a Markov model of vertical slices. Conklin [5], focused on trance music, explores a new approach to generating high probability and coherent chord loops from a statistical model trained on a chord sequence corpus. David Cope's Experiments in Musical Intelligence [49] is a system for algorithmic composition heavily based on the conservation of patterns.

III. Palestrina's Masses

The style of Palestrina can be seen as a combination of melodic lines in a polyphonic environment, characterized by the tension between harmonic and melodic elements. The line is the starting point of Palestrina's style, and the harmony does not have an independent sphere of interest as in Bach counterpoint. Jeppesen [1] analyses Palestrina's style using different aspects such as rhythm, modes, lyrics, melody, harmony and dissonance. His point of view is very clear about the secondary role played by harmony: "…The exactions arising out of harmonic aspects are really only intended to ensure the sonority of the individual harmonic moments. ´Harmonic´ does not signify here any independent sphere of interest; chords had not yet reached a stage when they had their own vigorous life, as in Bach's works". All that is required in vertical chords is clearness and sonority. The imitation is the base for constructing the polyphony. Accidentals are limited to F, G and C-sharp and B and E flat, found in plainsongs and related to Gregorian modes. De la Motte [51] mention that, "Palestrina just polished and refined a language developed by Josquin 70 years before". The accidentals of Palestrina are a natural evolution of the polyphony of Josquin des Prés, where E flat was always associated with B-flat and C, F and G sharp embellishments of the notes D, G and A.

The corpus of pieces we are working with consists of 101 masses composed by Palestrina (see Table I). These masses were published between 1554 and 1601, after his death in 1594. The date of composition of the different pieces is very difficult to determine, and each mass consists of various movements: Kyrie, Gloria, Credo, Sanctus, Benedictus, Agnus Dei. Each movement is divided into sections based on the text. The masses and the movements vary in the number of voices from three to six. For example, Benedictus in many masses is written in three voices and Kyrie in five or six. Table I describes is the corpus of pieces we have, using the data of music21 [52], a Python-based toolkit for computer-aided musicology developed by MIT.

Taking into account just two voices, the number of vertical slices available is almost 350,000 which provides enough information for constructing a reasonably accurate first-order Markov model, as is explained in the next section.

Table I. Corpus of Pieces of Palestrina from music21

Mass part	Pieces
Agnus	186
Benedictus	99
Credo	98
Gloria	101
Kyrie	129
Sanctus	104
Total:	**717**

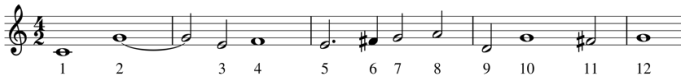
IV. Viewpoints for Pattern Discovery

For the generation of polyphony, both horizontal (melodic) and vertical (harmonic) aspects must be modelled and we use the concept of *viewpoints*, developed and refined by Conklin [2]-[5], from the horizontal and vertical perspective. In our work, the generation of imitative counterpoint is developed at two levels, corresponding to

the short- and long-term models of viewpoints. Long term or stylistic aspects are modelled using vertical slices and short term or intra-opus aspects using patterns discovered in a template piece. A template piece of music is transformed into a higher level description derived from the basic surface representation, by converting the sequence of basic events into sequences of derived viewpoint elements. A linked viewpoint is a combination of two or more viewpoints that models their interaction simultaneously. Following these steps, pattern discovery is performed on the transformed representation.

A. Horizontal Viewpoints

Each voice of the chosen template piece is cut into phrases which are assumed to be separated by rests. This division of phrases is possible in Palestrina vocal music, where music is thought taking into account the phrases of the text, and rests are written between separate ideas, never as part of a musical idea. Once the score is divided, each phrase of Palestrina music is treated as a sequence of linked viewpoint values. To better understand the concept of viewpoint, we take a melody of Palestrina. The sequence of notes is converted to a sequence of features derived from the musical surface (Fig. 1), for example, absolute pitch (pitch), name of note (spell), melodic contour, duration contour, interval (diaintc), or an abstract interval class (scalestep), as will be explained below. A pattern is a sequence of features $(v_{1,...},v_i)$ where each v_i is a feature (e.g. scale step linked with duration contour).



	1	2	3	4	5	6	7	8	9	10	11	12
pitch	60	67	64	65	64	66	67	69	62	67	66	67
spell	C	G	E	F	E	F#	G	A	D	G	F#	G
contour (pitch)		+	+	+	-	+	+	+	-	+	-	+
contour (dur)		+	-	+	-	-	+	=	=	+	-	+
diaintc		J5	m3	m2	m2	M2	m2	M2	J5	J4	m2	m2
scalestep		J45	Mm3	Mm2	Mm2	Mm2	Mm2	Mm2	J45	J45	Mm2	Mm2

Fig. 1. Different viewpoints applied to a melody of Palestrina. Agnus from Beata Marie Virginis. altus, bars 5 to 9.

The scalestep viewpoint groups successive intervals and is flexible enough to find patterns in Renaissance style. The values of that viewpoint are:

- Unison and Octave (J18)
- Minor second and Major second (Mm2)
- Minor third and Major third (Mm3)
- Perfect fourth and Perfect fifth (J45)
- Minor sixth – Major sixth (Mm6)
- Minor seventh - Major seventh (Mm7)

The repetitions of patterns in Palestrina are not merely exact transpositions of intervals. For example, a minor second can be converted to a major second, as is shown in Fig. 2. Using the syntax above, and taking into account just the scalestep viewpoint, the pattern indicated in Fig. 2 would be represented as:

$$\text{J45, Mm3, Mm2, Mm2, Mm2, Mm2, Mm2} \quad (1)$$

This pattern has seven components and represents the boxed segment in Fig. 2 (the soprano, altus and bassus of the Agnus from Beata Marie Virginis).

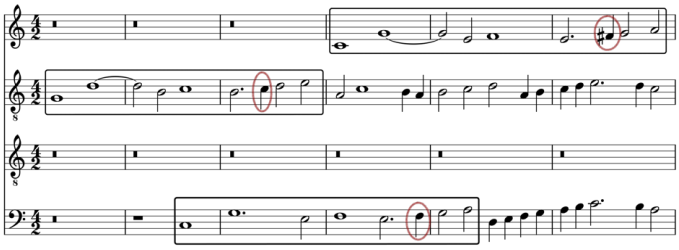


Fig. 2. Agnus from Beata Marie Virginis, bars 1 to 6. Palestrina. Pattern with different intervals.

B. Pattern Discovery

Data mining is the computational process of discovering interesting patterns in large data sets. This interdisciplinary subfield of computer science is growing, and the number of algorithms and researchers in the field highlights its importance. Sequential pattern mining has become an essential data mining task, with broad applications, including market and customer analysis, web log analysis or pattern discovery in protein sequences A survey on sequential pattern mining and the approach of the different algorithms, addressing efficiency and scalability, has been summed up by Khan & Jain [43].

Algorithms for sequential pattern mining include SPADE, Sequential PAttern Discovery using Equivalence classes [50], PrefixSpan, Prefix-projected Sequential pattern mining [51] , GSP, Generalized Sequential Pattern algorithm [52] CloSpan, Closed Sequential pattern mining [53], BIDE, BI-Directional Extension [54] or SPAM, Sequential Pattern Mining using A Bitmap Representation [55]. In our experiments we are using gap-BIDE [44], an extension of the BIDE algorithm for mining closed sequential patterns with possible gap constraints. Currently, we are working at zero gap level without taking into account gaps in the sequences.

To apply sequential pattern mining to Palestrina masses, each piece in the corpus is converted to a viewpoint sequence, with phrase boundaries indicated by rests, as explained in Section IV.A. The linked viewpoint for discovering patterns is:

$$\text{scalestep} \otimes \text{contour(dur)} \quad (2)$$

This particular linking of viewpoints allows the discovery of flexible patterns, e.g. augmented and diminished patterns. The viewpoint can also capture inversions, though it should be noted that equivalence according to the viewpoint is a necessary but not a sufficient condition for inversion. The vocal lines in Palestrina are very flexible and the imitations are sometimes just hinted at. Fig. 3 contains examples of patterns found in the Agnus II from the mass

The pattern 91 is an inverted scale made by six short equal notes plus a long note. Pattern 3 is more complex, typical of Palestrina music, where neither the durations are exact nor the melodic contour, but an attentive listener can easily identify the similarity of the two instances of this pattern. Regarding duration and melodic contour, pattern 1 is clearest but possibly a human musicologist would have taken the previous note of this pattern dismissing the different duration and interval of the first note in the different presentations.

C. Ranking Patterns

A huge number of patterns can be typically found in a template piece: some trivial, and some method for ranking them is necessary [6]. In this paper we establish a ranking of patterns based on a binomial distribution that computes the probability of obtaining an observed number of occurrences in a given number of sequence positions within the template piece.

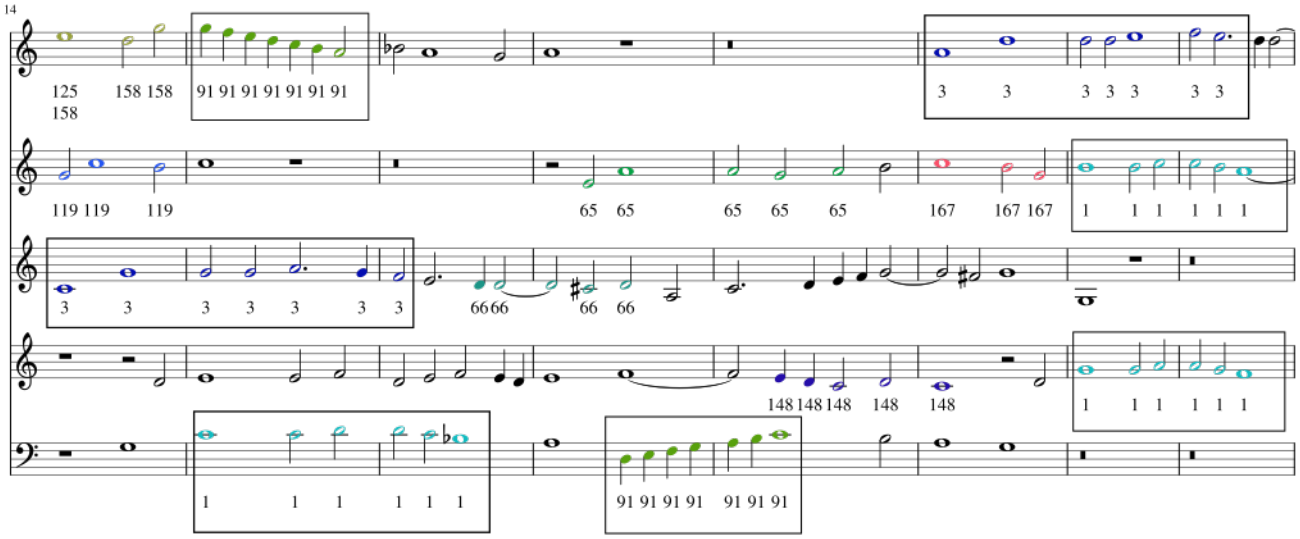
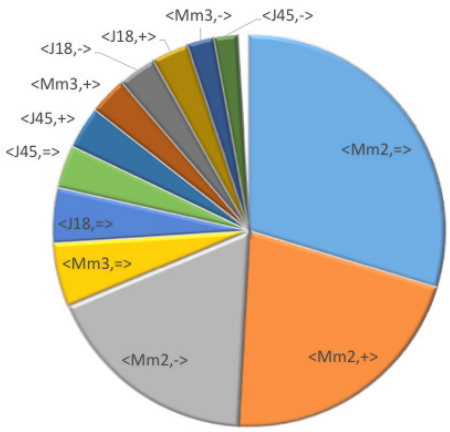


Fig. 3. Agnus II from Ascendo ad Patrem, bars 14 to 21. Palestrina. Patterns detail of patterns using scalestep ⊗ contour(dur).



Patterns
<J45,+>, <Mm3,->, <Mm2,+>, <Mm2,->, <Mm2,->, <Mm2,+>, <Mm2,=>
<J45,+>, <Mm3,->, <Mm2,+>
<J45,+>, <Mm3,->, <Mm2,+>, <Mm2,+>, <Mm2,->, <Mm2,+>

Fig. 4. A list of features and their counts (upper pie chart). Example of pattern encoding (lower table) from Beata Marie Virginis of Palestrina (see Fig. 1).

To rank patterns, it is necessary to know the background distribution of pattern components. Fig. 4 shows a distribution of the values of the viewpoint defined in (1). For example, <Mm2,=> indicates a scale step of a minor or major second with an equal duration, or <J45,+> indicates a scale step of perfect fourth or fifth where the second note has a higher duration. Clearly, the most probable interval is the second (69.4%) divided into same duration as the previous note (30.0%), higher duration than the previous note (21.4%) and lower duration than the previous note (18.0%). The rest of the melodic intervals have a much lower probability. This illustrates that, for example, patterns comprising predominantly <Mm2,=> features are not surprising and will not be significant unless they occur very frequently in the piece. The binomial pattern ranking, as described below, handles these effects in the piece.

The background probability of a pattern gives the probability of finding it in a random segment with the same zero-order distribution as the corpus. The background probability (of a pattern) using a zero-order model of the corpus is:

$$b_p = \prod_{(i=1)}^{l} c\,(v_i)/N \tag{3}$$

where:

- $c(v_i)$ is the total count of feature v_i,
- N is the total number of places in the corpus where the viewpoint is defined.

Using the background probability of a pattern, its interest I can be defined using the binomial distribution which gives the probability of finding exactly k occurrences of the pattern in a sequence of length t, where the background probability is b. Then the negative log probability of finding at least the observed number of occurrences of the pattern.

$$I(p) = -\ln B(k;t;b) \tag{4}$$

where:

- l_i gives the cumulative probability (right tail) of the binomial distribution,
- t approximates the maximum number of positions that can be possibly matched by the pattern,
- k is the number of times the pattern appears in the template piece.

with t calculated as follows:

$$t = \sum_{i=1}^{p} l_i - c + 1 \tag{5}$$

where:

- p is the number of phrases,
- l_i is the length phrase i,
- c is the length of the pattern.

D. Building the Template

Fig. 5 is one example of different patterns found in one fragment of Agnus from Beata Marie Virginis of Palestrina, ordered by their interest value Eq. (3). The number followed by a colon (:) indicates the interest for each pattern. The same melody can be covered by patterns in many different ways. To cover a template, the ranking according to their interest value is used. If a lower pattern in the ranking overlaps with one higher, the lower pattern is not considered. This simple greedy algorithm is repeated until reach the last pattern, trying to complete the template. In this way, the template will be covered by non-overlapping interesting patterns. The template obtained will be used for creating the new piece as is explained in the next section.

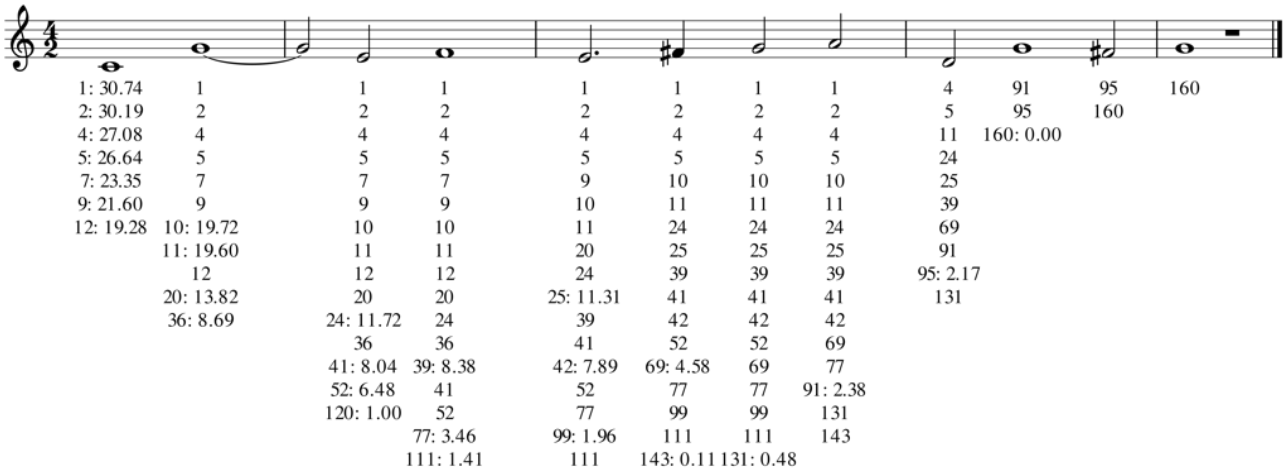


Fig. 5. Agnus from Beata Marie Virginis,of Palestrina. Altus, bars 5 to 9. Different patterns ordered by their interest (Equation 4). The first number gives the rank of the pattern, and the second the value for the first position of a pattern.

E. Vertical Viewpoints. Markov Model

For constructing the Markov model, two voices are selected and cut into slices (see Fig. 6). In this first approach, we have taken the highest and lowest voice for a better result, removing the inner voice. Usually, the music that follows harmonic constraints entrusts to the lower part (bass) an important role in the harmonic context, while the higher part (soprano) is more appropriate for defining melodies.

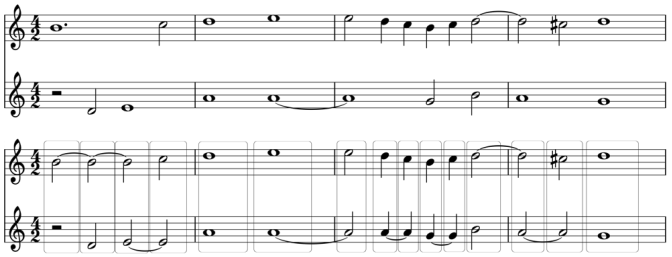


Fig. 6. An example of two-voice slicing.

The slicing process is the same as the full expansion method explained by Conklin [4], dividing when a new event appears in one voice. In our method, we do not retain ties between notes. In Renaissance vocal music, whether a note is repeated or tied sometimes depends on the text and furthermore durations are conserved from the template.

Taking into account pitch and duration, the number of slices in the corpus is 347,748. The zero-order Markov model is calculated counting the number of repeated slices and dividing by the total. The number of different slices is 1582 distributed as is shown in Fig. 7.

The vertical axis is the number of repetitions (logarithmic scale) and the horizontal the rank in the slice ordered by repetitions. Counting the number of next unique slices (first-order Markov model), also ordered by the number of repetitions (zero-order model), the results are shown in Fig. 8, where the number of different paths ranges from 0 and 183.

The piece now can be treated as a sequence of regular simultaneities where it is possible to apply different constraints that filter the possible paths. For example, based on the melody of Palestrina at the top of the Fig. 9, we illustrate the system with different restriction levels for creating a new upper voice. The upper voice is generated applying a random walk among the possible vertical slices using a first-order model. It is a short phrase, and it was easy to find solutions through forward generation with just one template and different viewpoint constraints in the lower voice. Ranking from strongest to weakest, and using linked viewpoints, they are labelled as pitch ⊗ duration, scalestep ⊗ duration and duration.

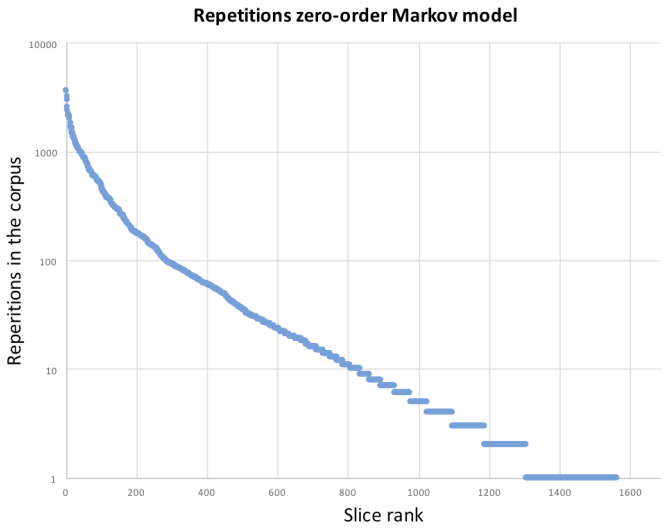


Fig. 7. Zero order distribution of repetitions.

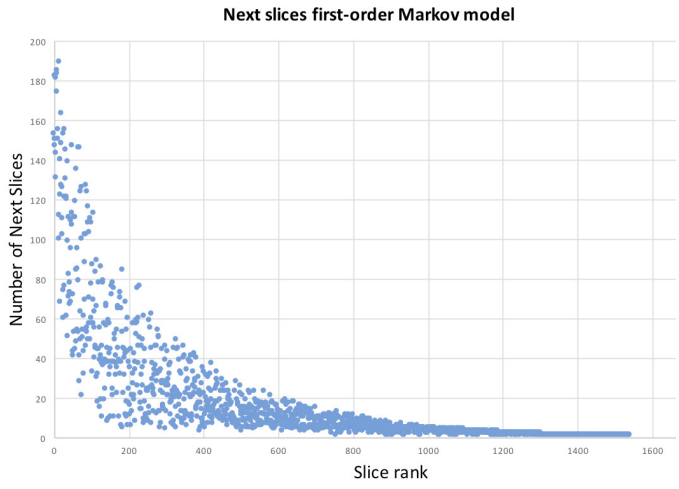


Fig. 8. Distribution of unique next slices, first-order Markov model.

V. Applying the Model to the Template

This section explains a method for generating new music based on a template which has been annotated with patterns, as described

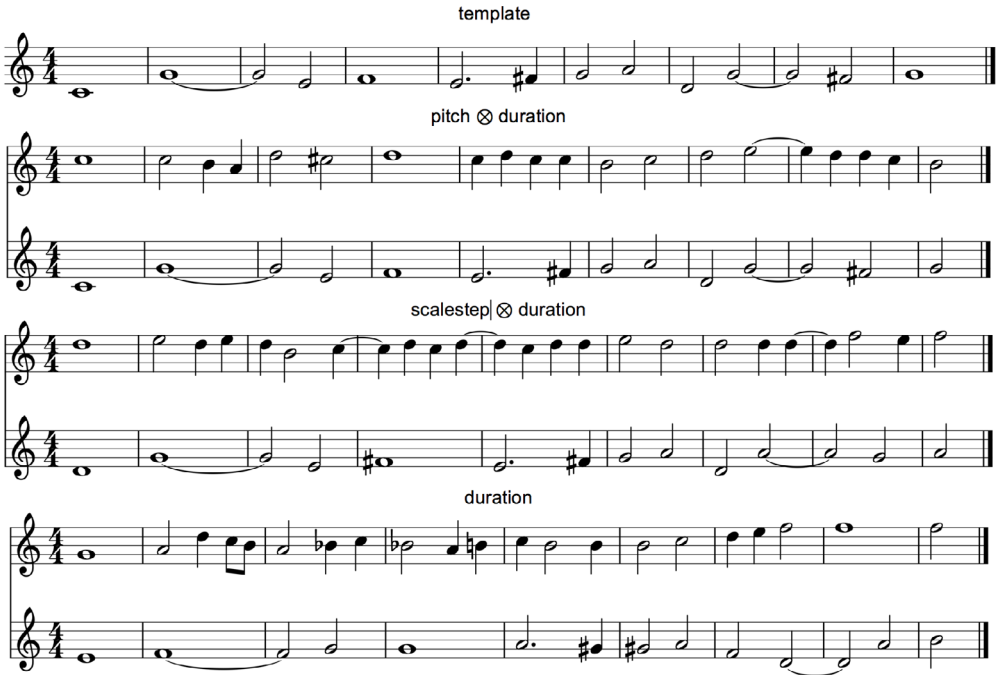


Fig. 9. Generation of upper voice based on different constraints in lower voice.

in Section IV. The idea is to fill the template with the slices and probabilistic paths obtained by the first-order model. The steps are as described below.

A. Forward Generation

For generating new music, one piece from the corpus is chosen, and patterns are discovered in the piece using the viewpoint scalestep ⊗ contour(duration) as mentioned earlier. Once the final template is constructed, for constraining notes within areas covered by patterns, the viewpoint

$$\text{scalestep} \otimes \text{contour(dur)} \otimes \text{contour(pitch)} \quad (6)$$

is used. Note that this represents a slightly more restrictive linked viewpoint than that used for pattern discovery (1), in that the regions are also required to conserve pitch contour. The generated music, therefore, conserves the abstract qualities of scale step, duration contour, and pitch contour. Further, in this article, the exact rhythm from the Palestrina template is used, therefore, the conservation of duration is assured. For describing the method, we take Benedictus from the mass Descendit Angelus Domini as a template and proceed with the next steps:

- Remove internal voices retaining the highest and lowest.
- Divide the template into regions organized by the patterns. If the region is a pattern, the viewpoint shown in Eq. (6) is used for horizontal restrictions. If the region is not a pattern, just the duration viewpoint remains;
- Filter the vertical slices by the different constraints. If at one point it is not possible to find a next slice, a backtracking algorithm is performed (see Section V. B).

There is a probability associated with each piece, using the statistical model. Different results will be obtained choosing pieces with different overall probabilities, as will be commented in Section VI.

B. Backtracking Algorithm

Due to the severe restrictions forced by the template, it is possible to encounter some points where all slices to continue the piece have zero probability at the slices generated. This problem was due to the bottleneck arising from the availability of very few continuations for some slices of the corpus. To solve this problem, a double backtracking algorithm has been implemented at two different levels, pattern and template. At the pattern level, the system goes one, or several steps back if no possible solutions are obtained for some slice. If the backtracking at pattern level reaches the first slice, the system goes back one (or several steps back) from the patterns of the template. This method is faster and permits a scattered group of solutions uniformly distributed.

VI. Results and Evaluation

The method described in Section V have been used to generate new pieces based on the Benedictus from the mass *Descendit Angelus Domini* as the template. This Benedictus is composed of just three voices (as most of the Benedictus in Palestrina masses are). The main purpose of taking a three voices piece is that we have to remove just one staff, and the counterpoint, imitations, and harmony are less affected than in a four or five voices piece.

The evaluation of a system for generating music is always a complex task. If the generation is limited to a very narrow and particular type of composition (first-species counterpoint, i.e.), the evaluation can be defined regarding "how many rules have been broken". Some examples in this sense can be analysed in [34][18][40][31]. In the case of Palestrina, the evaluation of broken rules such as parallel fifths or octaves of two consecutive slices is not possible because slice transitions are taken directly from the corpus. It is therefore not possible to find parallel fifths or octaves of two consecutive slices unless they specifically are in the Palestrina corpus. Some unusual melodic movements in Palestrina style possibly appear, mainly related to the use of accidentals, but is very difficult to measure violations of accidental use automatically.

A. Computational Evaluation

Some research has found a close relation between *information content* and the "quality" of the results in a statistical model [3][34] The information content can be defined for a sequence of slices $e_{1,\ldots}, e_n$ as follows:

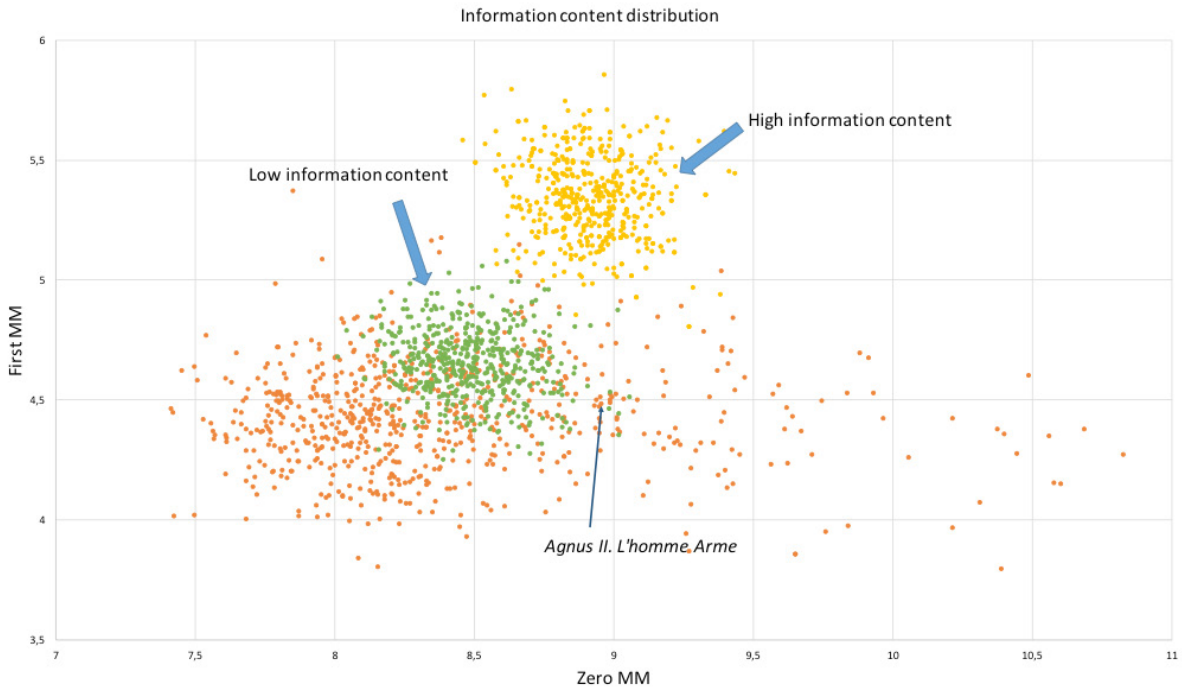


Fig. 10. Information content distribution using zero and first-order Markov Model. In orange the original corpus of pieces of Palestrina. In green and yellow the pieces generated based on the Agnus II from the mass L'homme Armé with low and high information content.

$$-\frac{1}{n}\sum_{i=2}^{n} \log_2 P(e_i \mid e_{i-1}) \quad (7)$$

$P(e_i|e_{i-1})$ where is the probability of eventi e_i in the first-order Markov model.

Taking *Agnus II* from the mass *L'homme Armé* as a template, we have generated 1000 pieces distributed as follows:

- 500 pieces, choosing the next slice of the first-order Markov model from the upper third of the next slice distribution. This option will produce a group of solutions of low information content.
- 500 pieces, choosing the next slice of the first-order Markov model from the lower half of the next slice distribution. This option will produce a group of solutions of high information content.

The distribution of the original corpus of pieces and the pieces generated using the zero- and first-order Markov model is shown in orange in Fig. 10. In this figure, the pieces generated, based on the *Agnus II* from the mass *L'homme Armé*, are in green and yellow. The green cluster indicates the pieces generated with the low information content and the yellow one the pieces with high information content. The information content of the original template is 4.50 as indicated in Fig. 10.

Taking just the values of the information content of the first-order Markov model from the original template, the distribution is shown in Fig. 10, where the 1000 pieces generated are divided into two groups of entropy based on the probability of next slices. Choosing the next slice from the upper third, ranking them from highest to lowest probability (blue bars, left part), they are centred around 4.65, that is slightly higher than the information content of the original template, 4.50. Taking next slices from the second half of the ranking, the information content is clearly higher, centred around 5.4. This second group explores less probable links between slices.

B. Listener Evaluation

For listener evaluation, we have taken three pieces: the original Palestrina and one of each group of Fig. 11 (low information content, high information content). The information content of these pieces is 4.65 and 5.54 respectively. They were interpreted by a professional choir at "The Vortex Jazz Club" in London, which is a small concert hall with a maximum capacity of 100 people. The audience was asked to identify which of the three pieces was the original and asked for their degree of confidence in the decision, on a five-point scale. 55 questionnaires were returned. The results showed that 49.1% identified correctly the template and 50.9% were deceived, as is seen in Fig. 12. On the questionnaire were some a questions about musical knowledge like "*Do you play an instrument?, How long have you played or sung?, What types of music do you listen?*". Analysing the responses, 33/55 (60%) have played an instrument or sung for more than five years, and 40/55 (72.7%) listen to classical music, which indicates at least an average musical knowledge and a trained ear.

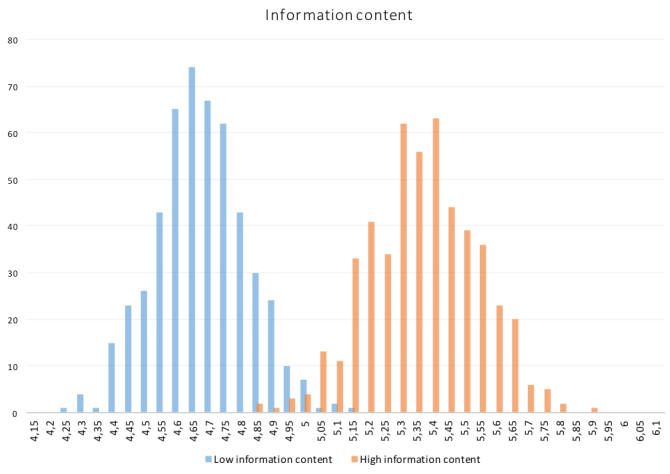


Fig. 11. Distribution of information content first-order Markov model of 1000 pieces generated divided into two groups of entropy based on the probability of next slices.

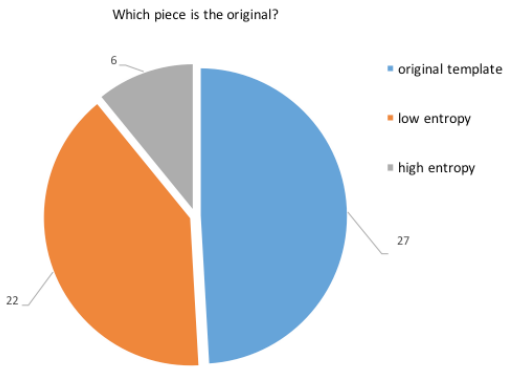


Fig. 12. Results of the question: which piece is the original?

The results obtained indicates a clear difference between the pieces generated with high and low entropy. Low entropy pieces are closer to the original style and even can deceive around the 50% of an audience with musical knowledge. It is possible that the melody and jumps in the vocal line of the high entropy piece are identified as "out of style" even without breaking counterpoint rules. Fig. 13 shows the results of the question, "How confident are you on a scale 1 to 5?".

Fig. 13. Results of the question: How confident are you on a scale 1 to 5?

VII. Discussion

In this section, we take a closer look at the main milestones and the different approach to the counterpoint generation for comparing with our research. Starting from the past century, Gjerdingen [59] creates a computer program for counterpoint species writing rules by hand. Gjerdingen recognises that the counterpoint rules remind an Old Testament patriarch calling out the commandments, "…*Thou shalt have no tritones; Thou shalt not leap dissonance; Thou shalt not commit parallel fifths; and so on.*" Those rules are implemented using functions that control the melodic line and vertical intervals. The evaluation is made taking a *cantus firmus* and comparing examples of the book of Jeppesen & Glen [60]. Their main goal is to create small fragments of counterpoint, given a *cantus firmus*, following rules as closer as the examples of the book, but not to the real Renaissance style.

One step forward can be considered the work of Farbood & Schoner [31]. In this case, the rules are not implemented by hand. Each rule is implemented as a probability table where illegal transitions are described by probability zero. The transition probabilities for generating a counterpoint line are obtained by multiplying the individual values from each table, assuming the rules are independent. A database of species counterpoint (12 pieces composed by human and 44 generated by the computer) is used as training data for the Markov model. The research is limited to first species counterpoint, and the main objective is to infer rules from a corpus of pieces, not to generate florid imitative counterpoint.

Adiloglu & Alpasl [40] also work on first species, but in their case using neural networks combined with back-tracking algorithm. The input layer represents the notes coming from the cantus firmus and the output layer the new counterpoint generated. The result is evaluated in two steps, by the counterpoint rules (parallel fifths, and octaves) and by asking a group of musical experts. In this case, the rules are not inferred from the corpus as in the Farbood & Schoner [31] work. Adiloglu & Alpaslan [40] ask a group of experts about the quality of the music generated, a musicologist, a composer and a choir director. According to their opinion, the melodies generated are generally correct but there are some cases in which the rules were not obeyed. The music experts also commented that the melodies produced do not always sound interesting. It is important to underline that first species, note against note, is the very first exercise for the student in counterpoint. In our case, we are comparing real masterpieces against pieces automatically generated.

Herremans & Sörensen [16] develop a variable neighbourhood search (VNS) algorithm that can generate musical fragments of arbitrary length consisting of a first species counterpoint melody given a cantus firmus. The VNS is a local search algorithm that starts from a randomly generated melody and improves it by changing one or two notes at a time. When no improving fragments can be found in any of the neighbourhoods, a local optimum is reached. In order to get out of this local optimum, a perturbation strategy is used. The algorithm reverts back to the best found fragment and changes a predefined percentage of the notes to a random allowed pitch.

The previous research work with species counterpoint (mainly first species), and it is not taken into account the complexity of imitations and structure of the counterpoint. Knopke & Jürgersen [43] try to identify common melodic phrases in the masses of Palestrina mapping multi-character music symbols into single-character tokens to build a suffix array structure. This research is focused just on analyses and not in generation. They claim that this system identifies all transpositions, inversions, retrogrades and retrograde inversions of unknown melodic segments. One limitation is that they do not use abstract viewpoints, just pitch and rhythm, and slight changes in imitations are disregarded.

VIII. Conclusions and Future Work

This paper presents a method for generating new music based on the corpus of masses of Palestrina. To sum up, comparing our research with previous work, our research is made working with real pieces of Palestrina, regarding the complexity of the counterpoint. The species are simple exercises invented for training, but they are not proper compositions. An exercise in species is a short piece of music without any kind of coherence or imitation, just one *cantus firmus* (usually in whole notes) and a second melody in the same duration (first species) or shorter (rest of species). For the template, we use different combinations of linked viewpoints, much more flexible than simple pitch and rhythm. Small mutations in imitations are very common, and the patterns detected should be robust to these changes. Regarding the work of Knopke & Jürgersen [43], for the identification is required the exact matching of the patterns. For example, non-exact intervals (a fourth by a fifth or a third minor by major) or mutations in a melody, very common in counterpoint music, are not considered. Filling a template identifying imitation is a way of retaining some grade of coherence into the piece. The counterpoint is based on the motifs that are repeated in different voices. One possibility explored in this article is to have a template to fill. This reinforce a strong, but possible

constraint that limits the number of possible paths. The employment of templates extracted from a piece is just one possibility for working with coherence, but this concept is nor even considered in the previous works.

The method, for practical purposes, is limited to two voices taking a template from the corpus without overlapping patterns. The gap-BIDE algorithm and the binomial distribution explained in Section IV.B works correctly detecting patterns. From an intuitive point of view, the ranking of the patterns discovered, in most of the cases, is related to the importance of the pattern in the piece. The greedy covering algorithm is quite simple and will be revised in a future version. Though this aspect is not the main goal of the project, a deeper research finding strengths and weaknesses of the method for finding patterns, template extracted and the covering algorithm should be done.

Regarding the Markov model, a first-order imodel s a good approach for ensuring correctly linked slices with rhythm and pitch constraints, preventing "weak" successive chords having grammatical errors such as parallel fifths, and parallel octaves, without implementing these devices using specific rules. This model does not organize harmonic regions, and "non-idiomatic" melodic movements can appear, mainly associated with accidentals. In this sense, a second-order model implementation could be an improvement for generating better melodies, but the training data would decrease exponentially. The main goal of this work is that the template complements some weaker aspects of the first-order Markov model and provides some kind of melodic coherence. In other systems, for example, David Cope's EMI [49], the coherence is achieved analysing bigger slices of the pieces, somehow inspired by the idea of *Musikalisches Würfelspiel* of some classical composers. In our case, the slices are reduced to the minimum rhythmic value and the possible structural information obtained, sparse. The template, therefore, provides the necessary scaffolding for the melodic ideas.

Section V.B commented on the double backtracking algorithm performed if no solution is found. The processing time is very high to find solutions using random walks when the group of optimum linked slices is very small, and in some cases, there may not be a solution due to the hard requirements of the patterns selected. The backtracking algorithm is faster than a simple random walk and provides a group of solutions homogeneously distributed. Another possibility that could be implemented in a future version is a depth-first search to explore all the different paths, which might lead to more heterogeneity in the results.

This model is made and tested for two voices due to the sufficient population of dyads in the corpus, but it is possible to extend to three or more voices using different viewpoints such as vertical intervals and duration. The zero-order Markov model will grow significantly, and the slices with higher probabilities will possibly decrease, augmenting the dead-end solutions, but hopefully, the corpus is large enough to find paths and create new and interesting pieces.

Acknowledgment

This research was supported by the project Lrn2Cre8 which is funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859.

References

[1] K. Jeppesen, The Style of Palestrina and the Dissonance, Dover Publications, INC ed., 2012.

[2] D. Conklin, "Music generation from statistical models," in *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, Aberystwyth, Wales, 2003.

[3] D. Conklin and I. H. Witten, "Multiple viewpoint systems for music prediction," *Journal of New Music Research,* vol. 1, no. 24, pp. 51-73, 1995.

[4] D. Conklin, "Representation and discovery of vertical patterns in music," in *Music and Artificial Intelligence*, C. Anagnostopoulou, M. Ferrand and A. Smaill, Eds., Springer-Verlag Lecture Notes in Artificial Intelligence 2445, 2002, pp. 32-42.

[5] D. Conklin, "Chord sequence generation with semiotic patterns," *Journal of Mathematics and Music,* vol. 10(2), pp. 92-106, 2016.

[6] D. Conklin and M. Bergeron, "Discovery of contrapuntal patterns," in *ISMIR 2010:11th International Society for Music Information Retrieval Conference*, Utrecht, 2010.

[7] J. D. Fernandez and F. Vico, "AI methods in algorithmic composition: A comprehensive survey," *Journal of Artificial Intelligence Research,* vol. 48, pp. 513-582, 2013.

[8] L. Hiller and L. Isaacson, "Musical composition with a high-speed digital computer," 1958.

[9] J. Rothgeb, Harmonizing the unFig.d bass: A computational Study., Thesis, Ph.D, Yale University ed., New Haven, Connecticut, 1968.

[10] K. Ebcioğlu, "Computer Counterpoint," in *International Computer Music Conference*, New York, 1980.

[11] K. Ebcioğlu, "An Expert System for Harmonizing Four-part Chorales," *Computer Music Journal,* vol. 12, no. 3, pp. 43-51, 1988.

[12] C. P. Tsang and M. Aitken, "Harmonizing music as a discipline of constraint logic programming," in *Proceedings of the International Computer Music Conference ICMC*, Montreal, Canada, 1991.

[13] R. Ovans and R. Davison, "An interactive Constraint-Based expert assistant for music composition," in *Proceedings of the Canadian Conference on Artificial Intelligence*, Ottawa, Canada, 1992.

[14] R. Ramirez and J. Peralta, "A constraint-based melody harmonizer," in *ECAI'98 Workshop on Constraints and Artistic Applications*, Brighton, 1998.

[15] S. Phon-Amnuaisuk, A. Tuson and G. Wiggins, "Evolving musical harmonisation," in *Proceedings of the International Conference on Artificial Neural Nets and Genetic Algorithms*, Portorož, Slovenia, 1999.

[16] D. Herremans and K. Sörensen, "Composing fifth species counterpoint music with a variable neighborhood search algorithm," *Expert Systems with Applications,* vol. 40, no. 16, pp. 6427-6437, 2013.

[17] M. Komosinski and P. Szachewicz, "Automatic species counterpoint composition by means of the dominance relation," *Journal of Mathematics and Music,* vol. 9, no. 1, pp. 75-94, 2015.

[18] A. Yilmaz and Z. Telatar, "Note-against-note two-voice counterpoint by means of fuzzy logic," *Knowledge-Based Systems,* vol. 23, no. 3, pp. 256-266, 2010.

[19] C. Roads, "Grammars as representations for music," *Computer Music Journal,* vol. 3(1), pp. 48-55, 1979.

[20] S. R. Holtzman, "Using generative grammars for music composition," *Computer Music Journal,* vol. 5(1), pp. 51-64, 1981.

[21] K. Jones, "A space grammar for the stochastic generation of Multi-Dimensional structures," in *International Computer Music Conference*, San Francisco, USA, 1980.

[22] R. A. McIntyre , "Bach in a box: the evolution of four part baroque harmony using the genetic algorithm," in *IEEE Conference on Evolutionary Computation*, 1994.

[23] A. Horner and L. Ayers, "Harmonization of musical progressions with genetic algorithms," in *Proceedings of the International Computer Music Conference*, 1995.

[24] S. Phon-Amnuaisuk, "Control language for harmonisation process," in *International Conference on Music and Artificial Intelligence*, Edinburgh, Scotland, UK, 2002.

[25] J. Polito, J. M. Daida and T. F. Bersano Begey, "Musica ex machina: Composing 16th-Century counterpoint with genetic programming and symbiosis," in *Proceedings of the International Conference on Evolutionary Programming*, 1997.

[26] N. Gwee, Complexity and Heuristics in Rule-Based Algorithmic Music Composition, Ph.D. thesis, Louisiana State University, 2002.

[27] P. Cruz-Alcázar and E. Vidal-Ruiz, "Learning regular grammars to model musical style: Comparing different coding schemes," in *International Colloquium on Grammatical Inference*, 1998.

[28] E. Gilbert and D. Conklin, "A probabilistic context-free grammar for

melodic reduction," in *In international Workshop on Artificial Intelligence and Music, at IJCAI-07, The Twentieth International Joint Conference on Artificial Intelligence*, Hyderabad, India, 2007.

[29] R. Groves, "Towards the Generation of Melodic Structure," in *4th International Workshop on Musical Metacreation (MUME 2016). Held at the Seventh International Conference on Computational Creativity, ICCC 2016*, París, 2016.

[30] D. Quick and P. Hudak, "Grammar-based automated music composition in Haskell," in *In Proceedings of the first ACM SIGPLAN workshop on Functional art, music, modeling & design*, Boston, Massachusetts, USA, 2013.

[31] M. Farbood, and B. Schoner, "Analysis and synthesis of Palestrina-style counterpoint using Markov chains," in *Proceedings of the International Computer Music Conference*, Cuba, 2001.

[32] D. Herremans, K. Sörensen and D. Conklin, "Sampling the extrema from statistical models of music with variable neighbourhood search," in *Proceedings of the Sound and Music Computing Conference*, Athens, Greece, 2014.

[33] M. Allan and C. Williams, "Harmonising chorales by probabilistic inferences," *Advances in Neural Information Processing Systems,* vol. 17, pp. 25-32, 2004.

[34] R. P. Whorley and D. Conklin, "Music Generation from Statistical Models of Harmony," *Journal of New Music Research,* vol. 45, no. 2, pp. 160-183, 2016.

[35] H. Hild, J. Feulner and W. Menzel, "HARMONET: A neural net for harmonizing chorales in the style of JS Bach," *Advances in Neural Information Processing,* p. 267–274, 1992.

[36] J. Feulner and D. Hörnel, "MELONET: neural networks that learn harmony-based melodic variations," in *Proceedings of the International Computer Music Conference*, San Francisco, 1994

[37] D. Hörnel and P. Degenhardt, "A neural organist improvising baroque-style melodic variations," in *Proceedings of the International Computer Music Conference*, Thessaloniki, Hellas, 1997.

[38] D. Hörnel and P. Degenhardt, "A neural organist improvising baroque-style melodic variations," in *Proceedings of the International Computer Music Conference*, Thessaloniki, Hellas, 1997.

[39] K. Verbeurgt, M. Fayer and M. Dinolfo, "A hybrid Neural-Markov approach for learning to compose music by example," in *Proceedings of the Canadian Conference on Advances in Artificial Intelligence*, 2004.

[40] K. Adiloglu and F. N. Alpaslan, "A machine learning approach to two-voice counterpoint composition," *Knowledge-Based Systems,* vol. 20(3), pp. 300-309, 2007.

[41] T. Feynman T, L and M. Gotham, M. Johnson and J. Shotto, "Automatic Stylistic Composition of Bach Chorales with Deep LSTM," in *ISMIR*, 2017.

[42] D. Precup and Y.W. Teh, "DeepBach: a Steerable Model for Bach Chorales Generation," *Proceedings of the 34th International Conference on Machine Learning*, 2017. pp. 1362-1371, 2007.

[43] I. Knopke and F. Jürgensen, "A system for identifying common melodic phrases in the masses of Palestrina," *Journal of New Music Research,* vol. 38, no. 2, pp. 171-181, 2009.

[44] K. A. Sidorov, A. Jones and A. D. Marshall, "Music Analysis as a Smallest Grammar Problem," in *In Proceedings of the 15th International Society for Music Information Retrieval (ISMIR)*, Canada: International Society for Music Information Retrieval, 2014.

[45] I. Khan and A. Jain, "A Comprehensive Survey on Sequential Pattern Mining," *International Journal of Engineering Research and Technology,* vol. 1, no. 4, June 2012.

[46] J. Wang and J. Han, " BIDE: Efficient mining of frequent closed sequences," in *Proceedings. 20th International Conference on Data Engineering*, Boston, Massachusetts, USA , 2004.

[47] F. Pachet, P. Roy and G. Barbieri, "Finite-Length Markov Processes with Constraints," in *In Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, Barcelona, Spain, 2011.

[48] T. Collins, R. Laney, A. Willis and P. H. Garthwaite, "Developing and Evaluating Computational Models of Musical Style," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing,* vol. 30 (1), pp. 16-43, 2016.

[49] D. Cope, "Experiments in musical intelligence (EMI): Non-linear linguistic-based composition," *Journal of New Music Research,* Vols. 18(1-2), pp. 117-139, 1989.

[50] D. Cope, Virtual music: computer synthesis of musical style, MIT Press, Cambridge, 2001.

[51] D. De la Motte, Contrapunto, S.A. Idea Books ed., Barcelona, 1981, p. 420.

[52] M. Cuthbert and C. Ariza, "music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data", *in Proceeding of International Society for Music Information Retrieval.* 2010.

[53] M. Zaki, "SPADE, An efficient algorithm for mining frequent sequences," *Machine learning,* Vols. 42(1-2), pp. 31-60, 2001.

[54] J. Han, J. Pei, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal and M. Hsu, "Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth," in *Proceedings of the 17th International Conference on Data Engineering*, Washington, DC, USA, 2001.

[55] R. Srikant and R. Agrawal, "Mining sequential patterns: Generalizations and performance improvements," in *International Conference on Extending Database Technology*, 1996.

[56] X. Yan, J. Han and R. Afshar, "CloSpan: Mining: Closed Sequential Patterns in Large Datasets," in *Proceedings of the 2003 SIAM International Conference on Data Mining*, San Francisco, California, USA, 2003.

[57] C. Li and J. Wang, "Efficiently Mining Closed Subsequences with Gap Constraints," in *SIAM International Conference on Data Mining*, Atlanta, Georgia, USA, 2008.

[58] J. Ayres, J. Gehrke, T. Yiu and J. Flannick, "Sequential Pattern Mining using A Bitmap Representation," in *Proceedings of the International Conference on Knowledge*, Edmonton, Canada, 2002.

[59] R. Gjerdingen, "Concrete musical knowledge and a computer program for species counterpoint," in *Explorations in Music, the Arts, and Ideas: Essays in Honor of Leonard B. Meyer*, Pendragon Press, 1988, pp. 199-220.

[60] K. Jeppesen and H. Glen, Counterpoint: the Polyphonic Vocal Style of the Sixteenth Century, Englewood Cliffs, N. J., Prentice-Hall, 1960, p. 330.

[61] V. Padilla and D. Conklin, "Statistical generation of two voice florid counterpoint," in *Proceedings of the International Conference on Sound and Music Computing*, Hamburg, Germany (2016).

Víctor Padilla Martín-Caro

Dr. Víctor Padilla Martín-Caro was born in Madrid, Spain. Currently, he is the Vice-Dean of Social Sciences and Humanities at UNIR, Universidad Internacional de La Rioja. Previously, he worked as a musical research at Lancaster Institute for the Contemporary Art at Lancaster University, UK. His research interests range from musical composition, optical music recognition (OMR), to music information research (MIR) for pattern discovery.

Darrell Conklin

Prof. Darrell Conklin is an Ikerbasque Research Professor at the Department of Computer Science and Artificial Intelligence at the University of the Basque Country, San Sebastian, Spain and the group leader of the Music Informatics Group. His research areas include music generation, statistical modelling, and pattern discovery in music.

Spatial Sound Rendering – A Survey

E. Lakka [1], A. G. Malamos [2], K. G. Pavlakis [2], J. A. Ware[1]*

[1] Faculty of Computing, Engineering and Science, University of South Wales, Treforest (United Kingdom)
[2] Media, Networks and Communications Lab Dept. of Informatics Engineering Technological Educational Institute of Crete Heraklion, GR 71004 (Greece)

Received 23 January 2018 | Accepted 3 May 2018 | Published 1 June 2018

unir
LA UNIVERSIDAD EN INTERNET

Abstract

Simulating propagation of sound and audio rendering can improve the sense of realism and the immersion both in complex acoustic environments and dynamic virtual scenes. In studies of sound auralization, the focus has always been on room acoustics modeling, but most of the same methods are also applicable in the construction of virtual environments such as those developed to facilitate computer gaming, cognitive research, and simulated training scenarios. This paper is a review of state-of-the-art techniques that are based on acoustic principles that apply not only to real rooms but also in 3D virtual environments. The paper also highlights the need to expand the field of immersive sound in a web based browsing environment, because, despite the interest and many benefits, few developments seem to have taken place within this context. Moreover, the paper includes a list of the most effective algorithms used for modelling spatial sound propagation and reports their advantages and disadvantages. Finally, the paper emphasizes in the evaluation of these proposed works.

Keywords

Real-time Spatial Audio, Acoustics, 3D Sound, Auralization, Simulation of Sound Propagation, Real-time Systems, Virtual Reality, Immersive Environments.

DOI: 10.9781/ijimai.2018.06.001

I. Introduction

THE last few years a growing interest in virtual environment technologies have been witnessed and their inclusion have been seen in a wide variety of different applications, including distance learning (e.g. serious gaming), the entertainment industry (e.g. online games, live events), architectural design, the production of art, various training scenarios, along with scientific and engineering research [1] [2]. As a result, work documented in the relevant recent literature has focused on improving the realism and the sense of the immersivity within a Three Dimensional (3D) Virtual Environment (VE), taking into account not only the realistic visual rendering but also the spatial sound propagation. The fact that sound rendering constitutes an integral part of the production of realistic VE, is not a coincidence, because it can offer additional details and visceral sense to a 3D immersive world.

Specifically, spatial auditory allows the user of VE to recognise the location of a sound source(s) [3], deduce information of the environment around the sound source(s) and, in general, to conceive the immersive environment in the same way as the listener recognizes the sound in the real world. Additionally, from the physical/algorithmical point of view, sound propagation techniques are used to simulate the sound waves as they travel from each source to the listener by taking into account the interactions with various objects in the scene [4]. In other words, spatial sound rendering in a VE goes far beyond traditional stereo and surround sound techniques, through the estimation of physical attributes, which are involved in sound propagation. Thus, characteristics such as surface reflection, diffusion, reverberation, and wave phenomena (interference, diffraction) can be included for the formation of spatial impressions of a virtual 3D scene (more details in Section 1.1).

To summarise, for several years great effort has been devoted to achieve high quality visual rendering for the development of interactive virtual worlds [5].Moreover, considerable attention has been paid to engage multiple senses in 3D interactive applications, for the reason that it constitutes a vital factor in order to improve the immersion and realism for the user experience [6]. In other words, the 3D VE can be described as a mosaic of technologies that includes visual and auditory rendering in order to simulate the real physical world. Accordingly, the question which arises is whether researchers take into account the attribute of spatial sound propagation for the design of 3D VE and what method is used to accomplish the desired result. Additionally, another issue is whether there are similar efforts in web 3D applications, in order to accomplish more sophisticated web virtual environments, with the contribution of the immersive audio.

The objective of this paper is to review empirical research studies and thus facilitate an understanding of the methods which are used in the (web) 3D virtual applications and in room acoustics, in order to render the spatial sound.

The remainder of the paper is organized into the follow sections: Section I.A describes the sound propagation phenomena; Section II analyses the most effective algorithms for the spatial sound propagation and they have been divided into categories; our conclusions are drawn in the final III section.

A. Sound Propagation Phenomena

This section gives a brief overview of the physical models of sound propagation and other acoustic effects. For the purpose of simulating sound in virtual environments the comprehension of these issues is fundamental. Hence, an explanation of physical phenomena is

* Corresponding author.
E-mail addresses: efilakka@gmail.com (E.Lakka), amalamos@ie.teicrete.gr (A. G. Malamos), pavlakis@staff.teicrete.gr (K. G. Pavlakis), andrew.ware@southwales.ac.uk (J. A. Ware).

described, in order that the methods and algorithms for realistic sound rendering can be understood.

1) Reflection

During the propagation of a sound wave in an enclosed space, the wave hits objects or room boundaries and its free propagation is disturbed. Moreover, during this process at least a portion of the incident wave will be thrown back, a phenomenon known as reflection. If the wavelength of the sound wave is small enough in respect to the dimensions of the reflecting object and large compared with possible irregularities of the reflecting surface, a specular reflection occurs. This phenomenon is illustrated in the Fig. 1a, in which the angle of reflection is equal to the angle of incidence. In contrast, if the sound wavelength is comparable with the corrugation dimensions of an irregular reflection surface, the incident sound wave will be scattered in all directions. In this case, the phenomenon is called diffuse reflection (Fig. 1b).

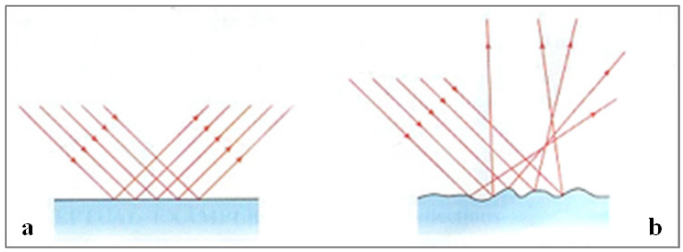


Fig. 1. Reflection – Sound Physical Phenomenon. (a) Specular reflection phenomenon; (b) Diffuse reflection phenomenon.

2) Diffraction

Another easily experienced characteristic of a sound wave is the diffraction which occurs when, for example, listening to without seeing another person from behind a door. Diffraction is the spread of waves around corners (Fig. 2b), behind obstacles or around the edges of an opening (Fig. 2a). The amount of diffraction increases with wavelength, meaning that sound waves with lower frequencies, and thus with greater wavelengths than obstacles or openings dimensions, will be spread over larger regions behind the openings or around the obstacles [7].

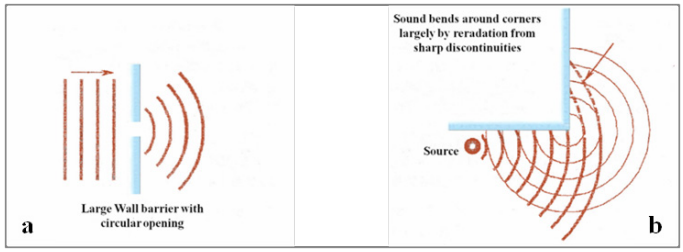


Fig. 2. Diffraction - Sound Physical Phenomenon. (a) Behind obstacles or around the edges of an opening; (b) Around corners.

3) Refraction

Refraction is the change in the propagation direction of waves when they cross obliquely the boundary between two mediums where their speed is different. This phenomenon should be considered for a realistic sound simulation. The path of a refracted wave can be found using Fermat's principle, which states that sound waves take the path with the least travel time (Fig. 3). For transmission of a plane sound wave from air into another medium, the refraction index in (1) is used, for calculating the geometric conditions [8].

$$n = c'/c = \sin\theta'/\sin\theta, \quad (1)$$

where c' and c the sound speed in the two media, θ the angle of incidence and θ' the angle of refraction.

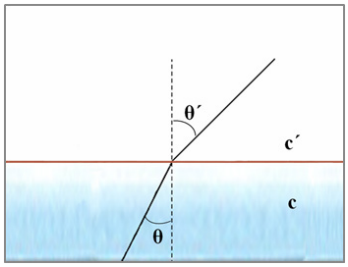


Fig. 3. Refraction - Sound Physical Phenomenon.

II. Background

The spatial sound rendering has been utilized in many approaches to simulate a realistic aural environment. As a result, a significant number of algorithms have been proposed to develop an innovative solution for this issue [9]. In our research, we used the below classification of these methods in order to survey the sound propagation technique:

Spatial Sound

- Sound propagation/Audio rendering
  - Acoustic Wave Equation Methods
    - Finite Element Method
    - Boundary element method
    - Finite-difference time-domain
    - Digital Waveguide Mesh
  - Geometric Methods
    - Enumerating Propagation Paths
      - Image Sources
      - Ray Tracing
      - Beam Tracing
    - Radiosity
  - Hybrid Methods
- Web Spatial Sound

This paper focuses on the most effective auralization techniques and reports both the positives and negatives of each of them. With this in mind, an evaluation of each proposed algorithms/methods is presented. Finally, the fact that despite the significant amount of work that has been carried out with regard to the potential for sound synthesis and sound propagation, there is still a lack of studies regarding the critical issue of spatial sound in web environments is highlighted.

A. Sound Propagation/Audio Rendering

1) Acoustic Wave Equation Methods

The most precise propagation algorithms that are intended to simulate several sound effects are based on numerically solving the acoustic wave equation [10]. Specifically, the acoustic wave equation, which describes the physics of sound propagation, is presented by the mathematical equation (2):

$$\theta^2 p/\theta t^2 - c^2 \nabla^2 p = f(x,t) \quad (2)$$

where p(x,t) is the acoustic pressure, c is the speed of sound and f(x,t) a source term describing one or more external sound sources.

Apart from this, the sound propagation has been expressed by the frequency-domain equalization of Helmholtz (3). As it is discussed

below, there are several methods which use it, in order to estimate the spatial sound in a virtual environment [11].

$$\nabla^2 P+\omega^2/c^2 P = 0 \text{ in } A^+ \quad (3)$$

where P (x, ω) is the (complex-valued) pressure field, ω is the angular frequency, and A+ is the acoustic domain.

Taking into account all the above, the ultimate goal of this section is to give an overview of the most well-known numerical techniques, by defining them; listing their advantages and disadvantages; providing examples of studies in which they have been applied. More details on the theoretical basis of the most popular state-of-the-art numerical structural-acoustic methods was presented in "Review of numerical solutions for low-frequency structural-acoustic problems", by Atalla and Bernhard [12]

a) Finite Element Method (FEM)

Finite element methods (FEM) [13] (volumetric techniques) have as major aim to solve numerically the wave (Helmholtz) equation in boundary conditions. In others words, FEMs try to solve the wave equation through the division of the space into a finite number of small elements (voxels) [14], [15].

Much research on sound propagation using FEM has been done. One of the first studies was developed in 1979 by Dennis W. Quinn [16]. In his work, the FEM was used to compute the sound propagation in non-uniform ducts which contain flow. To verify this approach, it was compared with other solutions or limiting cases and the sample calculations gave satisfactory results, in the case of two dimensional flows within no uniform ducts.

Moreover, a variety of approaches are demonstrated in the review paper examining the use of FEM in acoustic modeling, by Thompson [17]. The review was concluded with the prediction of the continued research on solutions to the challenging problem of developing efficient techniques for acoustic simulation, using finite-element methods, on the ground that it gives satisfactory results when compared to its simplicity.

Additionally, the group of Chou [18] used the FEM to assess the sound field distribution based on the indoor space and chamber volume, taking into account the effects of shape, absorption property, and room boundary, on sound delivery, in order to determine the improvement of interior sound. They reached the conclusion that their proposal can be applicable to the predesign analysis of interior architecture in order to improve the interior noise and reduce the construction costs.

To sum up for this method, it is one of the initial techniques that have been applied for the sound propagation. Additionally, the technique has also been extended in subsequent studies and has been used in hybrid approaches – namely in combination with other methods, in order to overcome some difficulties, as outlined in the literature review. Examples of hybrid methods will be presented in detailed in the corresponding paragraph.

Finally, FEM is not appropriate for open scenes, for the reason that these environments are characterised by sparse scattering geometry and uniform wave propagation speed. For this case, another acoustic wave equation method has been proposed as more suitable technique [19]. It has the similar philosophy as FEM and will be analysed in the next section.

b) Boundary Element Method (BEM)

The boundary element method (BEM) is a traditional technique which can provide numerical solutions and be applied to a range of engineering and scientific issues.

Particularly, it can be described by the simplicity for the reason that it demands only a mesh of the boundary of the domain. As a result, this technique is more effortless to be used than other classical finite element methods. Hence, the BEM is an integral part in the field of linear acoustics and it has reached a level of maturity in these areas because it can be overcome many of difficulties with low cost and offers adequate solution.

In comparison with the better-known FEM, the BEM differs in the element structure [20]. Specifically, the algorithm can be divided in the follow steps: (1) the Helmholtz equation is transformed into the boundary integral equation; (2) pressure and velocity are solved on the boundary, as a result of that the pressure is calculated at any point in the domain [21].

Overall, it constitutes one of the most accurate propagation algorithms for simulating various acoustic effects, evidenced by the fact that it has attracted much attention from research teams. Indeed it is no coincidence that the first audio representation approaches have been proposed using the BEM. Specifically, the first systematic study on the sound propagation in underwater was carried out by Chen and Schweikert [22], using a boundary element fluid model in combination with a boundary element structural model. The proposed method had the advantage of the readily used in practice, because the theoretical analysis was used in order to a mathematical model be constructed which can be an evaluated tool available for the next studies. Furthermore, Cunefare et al. [23] developed a boundary element method in order to solve the exterior acoustic radiation problem which was produced for the phenomenon of wavenumbers. The strong point of their work was the effectiveness of their method and the overcoming of the major drawbacks using boundary integral methods in acoustic problems (uniqueness of solution, singular integral kernels). Afterwards in 1996, the group of Zhenlin [24] provided a study using BEM to predicting the acoustic performance of expansion chamber mufflers with mean flow and compared it with the corresponding methodologies in the literature with satisfactory results. Moreover, Katz [25] proposed a solution to calculate a portion of the head-related transfer function (HRTF) of an individual based on precise geometrical data based on the BEM. They had found a cutting-edge solution to alter the geometry of the individual through the model in ways which were not possible with real objects. Lastly, Bapat et al. [26] presented a technique in which BEM was adapted in order to model 3-D half-space acoustic wave problems. They concluded with the presentation of results in which they highlighted the decrease of the requirements in the CPU time and memory storage of their algorithm in contrast with previous methods.

c) Finite-Difference Time-Domain (FDTD)

FDTD has become a commonly used algorithm in room acoustic modeling, for the reason that it is characterized by computational power that is becoming more readily available. It has as its major principle the fact that derivatives in the wave equation are replaced by corresponding finite differences. Furthermore, FDTD has higher precedence, because it produces better suited impulse responses to auralization than FEM and BEM, which typically calculate frequency domain responses.

Due to the above benefits of FDTD, several researchers have proposed interesting approaches for sound propagation using this technique. One of the first examples was presented by Blumricha and Heimannb [27], in which they investigated a simulation of sound propagation in an inhomogeneous atmosphere, avoiding many of the necessary approximations. The contribution of their study was the determination of single atmospheric influences to the whole effect on sound waves. An additional work has been described in the same year (2002) by Salomons group [28]. An interesting approach was analysed for sound propagation in systems with inhomogeneous moving media and finite-impedance surfaces using the FDTD model. Their results have been verified and have been in accordance with the solutions of the

Helmholtz equation. Furthermore, Heutschi et al. [29] based on FDTD in order to capture the typical ground impedances in the low-frequency range for outdoor soils adding small computational cost. Extensive results carried out showed that this method had good agreement in comparison with analytical solutions. Lastly, in [30] an adapted FDTD model was presented to be applied to a terrain-following coordinate system taking to account the orography of the ground surface. The results demonstrated in this study match state of the art methods.

However, the complexity of numerical methods increases linearly function according to the surface area of the primitives or the volume of the acoustic space, and as at least a cubic function of the maximum simulated frequency. Recently, many wave-based pre-computation techniques have been proposed for interactive applications. The reason being is to take advantage of numerical methods but not use it in real time, which increases the computational cost.

d) Digital Waveguide Mesh (DWM)

DWM approaches are numerical simulation techniques which use discrete waveguide elements; each of this element carries waves along its length in single dimension. One of the first systematic analysis was described by Duyne and Smith [31] in which an adequate technique was developed extending the DWM to model the propagation of the wave in a membrane. Additionally, Savioja group [32] presented a detailed analysis of DWM and proposed some methods in order to overcome its major drawback which is the direction dependent dispersion. Also, Murphy [33] developed an innovative environment which offered for research into the application of DWM-based models for virtual acoustic spaces.

However, DWM methods suffer from directional dispersion of sound, that is, sound does not travel with the same speed in different directions on the spatial grid [34]. Due to the fact that DWM also had significant drawbacks, the following research should use more sophisticated techniques to overcome the limitations in numerical methods and produce a realistic sound simulation.

2) Geometric Methods

The main benefit of the acoustic wave equation methods is to yield accurate results for the physical propagation of sound, by solving numerically the wave equation. However, they also exhibit a number of drawbacks. Specifically, their complexity depends both on the simulation frequency and on the surface areas of the objects or the volume of the acoustic space [35]. So, these methods are sensitive to the complexity of the virtual environment, which means that they are mainly limited to realistic scenes. To put it differently, these techniques are computationally too expensive for dealing with the whole audible frequency range and they are inadequate for high frequency auralization. As a result, these techniques are not appropriate for interactive sound propagation in complex environments. [19], [36-42].

For this reason, geometric methods were used in order to overcome some of the above disadvantages of wave methods, so that the 3D sound can be embedded into more sophisticated and complicated applications.

Firstly, from a theoretical point of view, sound and light are both waves; as a result, they share many common properties and similar techniques can be used in order to be rendered in a virtual scene. With this in mind, many algorithms which have been used for light rendering, they were adapted to solve the sound rendering in a 3D scene.

However, the extensive analysis of these two phenomena identifies some significant differences between them. Firstly, from the physical point of view, sound waves cover a much broader range of different wavelengths (the wavelengths of audible sound fall between 0.02 and 17 meters and their correspondent frequencies are 20kHz to 20Hz), more than three orders of magnitude larger than visible light. Moreover, as a result of the long sound wavelengths in comparison with the dimensions of most common objects in a room, the sound modeling requires less detail of room geometry. In general, the most common behavior of sound waves reflection in large objects (such as walls) is the specular. On the other hand, significant diffraction is occurred in the around edges of objects. Small objects have meaningful effect on the sound propagation only in the case that frequencies are over 4 kHz, so they can usually be excluded from auralization algorithms, especially in the presence of other sources with considerable reflection and diffraction phenomena. Secondly, sound travels through air roughly 106 times slower than light. As a result, the acoustic signal is perceived as a combination of direct and reflected sound, because its speed causes obviously different arrival times for sound propagating in different paths. Lastly, since sound is a coherent wave phenomenon, the computation of the reflected and scattered sound waves should incorporate the phase (complex amplitude) of the incident and reflected waves. In contrast, the incoherent light demands only the sum of the power [43].

Despite the above differences between sound and light, the algorithms of sound propagation borrow the most of the light rendering algorithms and techniques.

a) Enumerating Propagation Paths

Image Sources

The main principle of the image method is to enumerate specular reflection paths, by taking into account virtual sources. The virtual sources are constructed like mirroring location of the initial source, with respect to all polygonal surfaces of the environment. In other words, a sound source is reflected against all surfaces in a model; as a result, a set of image sources is produced. These are again reflected against all the surfaces by an iterative process, until a termination condition is satisfied. Response length or reflection order could be examples of these conditions in order to finish the algorithm and return the result [44].

Additionally, a hierarchical image-source tree can be used in order to depict the result of the image-source computation. In this structure, the root is the sound source and each branch represents an image-source. Consequently, an impulse response for acoustic scene can be calculated as the sum of all image-sources, which are included in the environment and they are the branches of that tree. Also, from the computational point of view, this method is used for the computation of the first reflections to avoid the significant increase of cost which depends directly on the number of image sources. [45].

A growing body of literature has examined and analyzed the image-source algorithm in the field of sound propagation. The interest in this area is demonstrated by the increasing number of reviews and the systematic study on this topic by other researchers [3], [9], [7], [44].

In an early study, Gibbs and Jones [46] used the image source method in order to measure the variation of sound pressure using a constructed rectangular model which possess a significant number of absorption configurations. Later, Santon [47] presented an approach for the estimation of speech intelligibility in rooms, using the image-source technique. His proposal takes into account the directional distribution of the echoes and is based on the concept of received energy being partitioned into useful and disturbing energies. This method was evaluated and was characterised as a clear improvement over the other techniques for predicting speech intelligibilities. Another study by Allen and Berkley [48] developed the impulse response between two points in a small rectangular room. This work was characterized by simplicity and was implemented for a room with rigid walls with only specular reflections and no diffractions. In 1984, Borish [49] improved the image-source method described by Santon [47], in order to extend for arbitrary polyhedra with any

number of sides. Later, Heewon Lee and Byung-Ho Lee [50] presented an algorithm for the simulation of sound ray paths in an arbitrary polyhedral room, which is based on the image model algorithm. After evaluations, they concluded that their method improved the efficiency of the image model technique by compensating for the drawbacks of corresponding methods. Furthermore, Vorlander [51] proposed a new method which combined both image-source model and the ray tracing. In 1992, Renate Heinz [52] developed an innovative approach in which they extended the image source method with a diffuse background signal to its result. In this study, only the specular parts of the reflections were considered and analysed. The following decade, Lehmann and Johansson [53] introduced a new method which provided an approximation of the acoustic energy decay (energy–time curve) in room impulse responses generated using the image-source technique. This study gave the advantage to researchers to undertake a preliminary analysis of a simulated reverberant scene without the need for time-consuming image method simulations. One year later, McGovern [54] addressed two major disadvantages of the image source method (redundant or unnecessary mathematical operations). In the first case (redundant mathematical operations), the use of look-up tables was proposed and in the second case (unnecessary mathematical operations), he used a sorting method. The evaluation implied that this method resulted in a substantially reduced computation time, and thus had potential applications for real-time auralization in interactive scenes. In addition, in 2013, André's group [55] analyzed the issue of auralisation which was based on geometric acoustic models. Specifically, the direct sound and reflections from each source were computed dynamically by the image-source method with the combination with HRTFs. Subsequently, trying to improve image methods for real time approaches, Charalampous and Michael [56] implemented an image source method variant in which they compared three different tree traversal approaches, depth-first, breadth-first and best-first.

As reported in the above studies, the fundamental benefit of these methods is their robustness. Specifically, they offer the security that all specular paths up to a given order or reverberation time will be found. On the other hand, the computational complexity of these methods grows exponentially, despite the fact that they compute only specular reflection. A solution to this problem has been achieved through the introduction of a ray tracing algorithm, which is analyzed in the next section.

Ray Tracing

The concept of this method is to compute propagation paths of sound that arrive to a receiver by generating rays emanating from the source and following them through the environment until an adequate number of rays has been found that reach a receiver position. [9], [7]. To put it another way, during the first phase the sound rays are produced in all directions. These rays are reflected at surfaces and the aim of this step is to find which of them hit any listener because this means that the specific ray is audible. It is obvious that the sound propagation has a number of similarities with the problem of solving global illumination by ray tracing, which is used in light rendering, and a similar method can be applied in the sound field.

According to the way sound sources emit rays and depending on how these rays interact with objects of the environment, different versions of the ray tracing algorithm have been developed. This means that these methods take into account the directions of rays (fixed/random) or the absorption of the incoming wave by the surfaces. Specifically, the size of the absorption of any surface should be identified from a coefficient which depends on the wavelength of the incoming sound. Finally, there are many ray tracing methods, which can also handle diffuse reflections and take into account the percentage of diffusion or scattering, but in these cases the computation time significantly increases [57].

There is a considerable amount of literature on the issue of sound propagation using the ray tracing technique, a fact that demonstrates the significance of the specific algorithm.

The first systematic study on the distribution of early reflected sound over the audience areas in concert halls was carried out in 1968 by Krokstad using a ray tracing technique [58]. Afterwards, Kulowski in 1982 [59] indicated a method of determining a quantitative measure of the ray tracing technique error. The importance of this study lies in the fact that it offered the option to enumerate sound decay curves of comparative credibilities, which is especially useful when curves are being modeled in different observation regions or even in different rooms. Two years later, the same author presented an algorithm which was based on the ray tracing method. This proposal gave the opportunity to model the acoustical field in rooms using small computers, because the calculation time was decreased sufficiently, a fact which has been confirmed using several examples [60]. Furthermore, in 1993, Lehnert [61] not only analyzed the two main kinds of inherent systematic errors of this method (errors due to a detection problem and errors due to limited spatial resolution), but also proposed an algorithm for this purpose which was validated perceptually. In the same year (1993), van Maercke and Martin [62] designed an approach to artificial reverberation in acoustics, without taking to account the diffuse in the environment of simulation. It contained a ray tracing algorithm for the calculation of echograms and implemented a beam method to predict short time impulse responses and criteria maps. Additionally, Li, Taherzadeh and Attenborough [63] used and extended the previous ray tracing scheme, in order to predict the sound field near a flat impedance ground in a refracting atmosphere that includes the effect of vector wind and turbulence explicitly. Before that, Mueller and Ullmann [64] described an approach to enhance sound by high quality 3D audio information through acoustic ray tracing. The main objective of this work was to compute a fixed sound source for a constant listener, with the prospect of an extension of the method in which 3D audio for moving listeners could be generated in interactive environments. Furthermore, Alpkocak and Sis [65] presented an approach to calculate the impulse response of a room using the ray tracing algorithm. This is based on assumptions that the environment was linear time-invariant system and the impulse response was calculated by sending Dirac impulses into the system as input and then the output gave the response. Moreover, ray tracing was attractive for further study. For example, the group of Dreher [66] tried to study the 3D ray tracing algorithm in the environmental noise context. The major goal of this approach was to reduce the computation time using different acceleration structures. Other work from the group of Okada [67] was presented in which novel ray tracing method was developed in order to solve sound diffraction problems using the ray tracing due to calculate sample values of the integrand. Besides, the evidence from this study intimated that this method offered applicability, after its evaluation with a prototype system of interactive ray tracing.

Furthermore, in the last few years, considerable attention has been paid to sound propagation for more sophisticated and realistic implementations. Specifically, Taylor et al. [68] introduced an attractive method for tuning geometric acoustic simulations based on ray tracing. They demonstrated that their technique achieved a significant performance improvement over prior geometric acoustic methods for the same number of contributions. As a result, the system had the ability to render acoustic spaces composed of thousands of triangles interactively. Also, Mo et al. [69] developed an efficient algorithm which is based on ray tracing, in order to simulate sound propagation in large outdoor scenes. The peculiarity of these environments was the variety of objects and the complexity of objects' boundaries. After that, the same group presented a paper in which the ray tracing was

used in order to improve the efficiency of outdoor sound propagation, without significant limitations of the scene, by the utilizing of analytic ray curves as tracing primitives [70]. In the same way, in one of the most recent publications [71], Schissler and Manocha computed the propagation paths from each source to the listener in large, dynamic scenes using ray tracing. Through this approach, they succeeded in accelerating the computation of impulse responses for interactive sound rendering in a 3D virtual environment.

As mentioned in the literature review, one of the essential advantages of this method is the simplicity. Furthermore, the speed and the efficiency on GPU hardware are additional strong points of the ray tracing algorithm. Finally, it offered the possibility to consider both higher order reflections, without significant computation cost increase and diffuse reflections. On the other hand, ray tracing is a stochastic method and one of its drawbacks are sampling artifacts as well as possibly lost important sound paths due to the limited sampling steps. There is no one guarantee that all significant paths will be considered. In addition, diffraction is theoretically possible but cannot be solved efficiently. This appears because the more distant a ray has traveled, the more sampling artifacts will occur. As a result, many rays are computed that will never reach the listener [7], [72].

The evidence from two algorithms analysis (image source-ray tracing) for sound propagation intimates that they have both advantages and disadvantages. Therefore it is worthwhile developing a combination of both in order to obtain fine temporal resolution in sampling rate quality, taking to account the phenomenon of scattering and faster audibility check of image sources [8]. This combination is called a "hybrid method" with key advantage of this approach being that a number of weak points in one algorithm can be ameliorated using strengths of the other. More details on this are given below.

Beam Tracing

This methodology was firstly developed in computer graphics in order to utilize the spatial coherence in generating realistic images and after that it was adapted for the sound rendering. In the sound propagation case, it classifies the propagation paths from a sound source, using the recursive method to trace the pyramidal beams (i.e., sets of rays) through the acoustic scene Particularly, for each beam, polygons in the scene are considered for intersection with the beam in front-to-back visibility order. During the algorithm firstly, polygons are detected, secondly, the original beam is clipped to delete the shadow region, thirdly, a transmission beam is constructed to match the shadow region, fourthly, a reflection beam is produced by mirroring the transmission beam over the polygon's plane, and finally conceivably other beams are created in order to model other types of scattering (see Fig. 4) [9], [72].

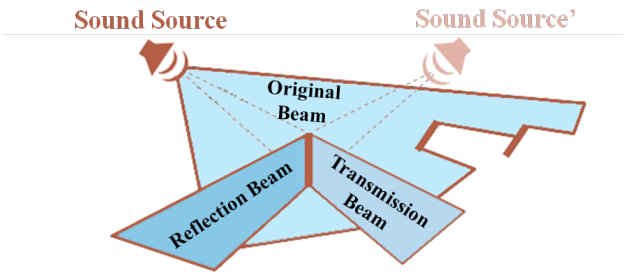


Fig. 4. Beam Tracing Method.

To review, there is a considerable amount of literature on this algorithm and especially if it is combined with another method of sound propagation. Preliminary work was carried out in the early 1990s, by Lewers [73] in which he proposed a beam tracing model in order to predict the behavior of sound in a room. Then, Funkhouser's group [74] developed new beam tracing algorithms that greatly accelerate computation of reverberation paths in a distributed virtual environment by taking advantage of the fact that sounds can only be generated or heard at the positions of listeners, which were represented by avatars. In this work, they succeeded in developing a faster beam tracing sound propagation technique which could support real-time computing.

Additionally, Funkhouser et al. [75] describe a beam tracing method based on precomputed spatial subdivision and "beam tree" data structures that enables real-time simulation of sound for static sound sources in interactive virtual environments. The offered advantages of this technique, in contradiction with earlier literature, were the scale, the accuracy, and the interactivity.

Moreover, others have analyzed a new theory which extended geometrical acoustics with diffraction phenomena [76]. The strength of their contribution lies in the fact that it is a new beam tracing method for enumerating sequences of diffracting edges efficiently and without aliasing in densely occluded polyhedral environments; it resulted in a practical approximation to the simulated sound field in which diffraction was considered only in shadow regions; additionally, it resulted in a real-time auralization system demonstrating that diffraction dramatically improved the quality of spatialized sound in virtual environments.

Further important research was proposed by Funkhouser's group [77]. Specifically, in this work, a beam tracing method was developed in order to enable interactive updates of propagation paths from a static source to a moving listener in large indoor area. Firstly, the major advantage of this approach is the ability to support auralization in large building environments. Secondly, it simulates sound propagation due to edge diffraction. Third, it finds all propagation paths up to a given termination criterion without exhaustive search or risk of under-sampling and finally it updates propagation paths at interactive rates.

Furthermore, Ajaj, Savioja and Jacquemin [78] developed an innovative environment for the purpose of the sound propagation in an interactive virtual scene. In their development, they based their approach on the beam tracing algorithm in order to implement the one part of this application for a real-time acoustic simulation. Also, the group of Laine [79] proposed an advanced solution which improved the previous method of beam tracing, in order to accomplish an optimized algorithm for finding and efficiently updating specular reflection paths for a moving listener. They demonstrate that the proposed algorithm performs well both with complex, lightly occluded room models and with moving sound source at interactive rates with moderate model complexity.

Furthermore, Antonacci, Sarti and Tubaro, in 2008, proposed a novel technique that could enable the fast tracing of a large amount of acoustic beams through the iterative lookup of a special data structure that could represent the global visibility between reflectors. After two years, this group analyzed an extension of the previous work in which diffraction and diffusion were simulated in the model to succeed more realistic results, without considerably increasing of the computational efficiency. This expanded approach demonstrated that not just the construction of the beam-tree but also the whole path-tracing process can be performed entirely on the visibility maps [80], [81].

As others have highlighted, beam tracing is currently considered to be the fastest commonly used geometric room acoustics modeling technique [72], [79], [82], [83]. This algorithm has advantages over both image source and ray tracing. Comparing with the image source method, the beam tracing advances in the fact that fewer virtual sources must be considered for the sound propagation algorithms, which is an important factor of effectiveness in geometrical complex scenes. Since each beam represents the region of space for which a corresponding

virtual source (at the apex of the beam) is visible, higher-order virtual sources must be considered only for reflections of polygons intersecting the beam [9]. On the other hand, in contrast to ray tracing, beam tracing works with object-precision polyhedral volumes that support well-defined intersections with diffracting edges. As a result, beam tracing is not affected by aliasing phenomena [76], [84], [85]. Additionally, it advances in the geometric coherence, because the algorithm calculates for each beam an infinite number of potential ray paths from the source to the listener. As a result, beam tracing is independent of the sampling artifacts, such as the ray tracing. [77].

Finally, Charalampous and Michael in their review paper [72], underline that beam tracing transcends as a deterministic method, in comparison to ray tracing. At this point, it should be noted that deterministic method will produce the same results when run multiple times. For example, a deterministic algorithm for detecting sound reflections, in a specific model, will detect the exact same reflection paths up to a given order of termination each time executed.

In contradiction with the above analysis of beam tracing benefits, this method is characterized from drawbacks too. For example, the geometric operations required to trace beams through a 3D model, such as intersection and clipping, are relatively complex for the reason that there is the possibility for each beam to be reflected and/or obstructed by several surfaces [77], [9]. Likewise, the group of Funkhouser [74] at the Bell Laboratories point out another disadvantage of the beam tracing, in that they notice the fact that the particular algorithm is difficult in environments with curved surfaces and non-linear refracting objects. They propose, as suitable solutions in these cases, the conservative beam tracing methods combined with validation of constructed paths.

b) Radiosity

Some preliminary work for the radiosity method was carried out in the 1950s, in the field of thermodynamics. Although, the basic equations of this technique have been presented in an optics paper by the Yamauti [86] and have been attractive for computer graphics since 1980s.Without delay, this technique was suggested and developed in acoustics, for the reason that both light and sound share many similar properties. On the other hand, there are some principal differences between radiosity in acoustics and radiosity in computer graphics, such as time-dependence. Particularly, sound is in contradiction with light, because it travels so slowly through the air. This fact causes a significant time delay which cannot be disregarded by any model of sound propagation. As mentioned below, it is one of the limiting aspects of acoustical radiosity because of the high computational cost [87].

There is a considerable amount of literature on acoustic radiosity. Some preliminary work was carried out several years ago. In 1993, the group of Shi [88] proposed a modified radiosity algorithm in order to implement both visual and room auditory rendering. After a few years, Tsingos and Gascuel [89], [90] presented new approaches for the simulation of room acoustics based on hierarchical radiosity. These are novel approaches because complex phenomena, such as sound global specular and diffuse reflections, were taken into account for the first time, in order to make their methods promising for virtual acoustics applications. Afterwards, at the beginning of the next decade, Nosal and co-workers [91], [92] developed a radiosity algorithm for rectangular rooms while they focused on addressing the problem of the applicability of the method in the light of the inherent assumption of diffuse reflection. Furthermore, in 2004, a mini review of radiosity method in simulating sound fields with diffusely reflecting boundaries was presented by Kang [93]. Additionally, Nosal et al. extended their work ([91], [92]) documented in another paper in which they investigated how acoustical radiosity performs in predicting real room sound fields (a squash court, a classroom, and an office). Moreover, Siltanen et al. [94] proposed a new extended acoustic radiosity method to solve the room acoustic rendering equation in order to handle both diffuse and non-diffuse reflections. Finally, different from all the previous literature, Muehleisen [95] suggested the radiosity for the prediction of sound pressure levels in six sided rectangular rooms.

As reported previously (in Section 2.1.2.2), the sound radiosity is limited by the fact that it is time dependent, in comparison with the corresponding method in light. Nevertheless, this feature can be an advantage on the ground that this computational cost is incurred only in the initial sound rendering. Specifically, the estimation for the sound rendering in an area will need to be done once at the beginning for a given source. After that, the remaining computational costs are adequate low to enable real-time sound simulation for moving listeners. In addition to this, several approaches have been proposed to accelerate the initial rendering, in order to improve this issue [92].

3) Hybrid Method

As outlined in the above literature review, classical methods for sound propagation are typically limited for realistic results for a number of practical interactive applications. Particularly, wave-based methods are adapted to the lower frequencies and relatively small domains but are not sensitive to the complexity of the domain. On the other hand, geometric methods are dependent on the number of successive reflections in the domain, and thus to their complexity [96]. For this reason, a number of hybrid methods have been developed and suggested by combining the classical sound propagation methods. This model was chosen because it is one of the most rapid ways to take the advantages of the above methods, in order to generate realistic sound effects, including reflections, reverberations and succeed significantly reduced calculation times.

The significance of the hybrid method is confirmed by the more recent literature. One of the first studies in this field, Tsingos et al. [76] extended the beam tracing algorithm to construct propagation paths with diffraction, and they introduced a practical approximation to the diffracted field in shadow regions. Their evaluation clearly established that (1) beam tracing algorithm is an efficient and unaffected from aliasing phenomena to compute diffraction sequences in densely occluded acoustic scenes, (2) it is an effectively technique to produce early diffracting propagation paths and auralization in real-time and (3) diffraction greatly improves the quality of spatialized sounds in immersive virtual environments. After that, Sikora [97] reported a hybrid algorithm using beam tracing method, which was intended to solve the problem of the refraction, without sacrificing the accuracy and efficiency of beam tracing method. Furthermore, Tsingos et al. [15] presented an auralization framework which gave the opportunity to render scattering effects interactively thus providing a more compelling experience, based on the programmable graphics hardware for all geometric computations. Also, after validation examples, their results showed for the first time that the Kirchhoff approximation can be successfully used for off-line sound propagation in very complex scenes. Also, Lauterbach, Chandak and Manocha [5] combined two different methods in order to produce a innovative algorithm for real-time auralization in complex, dynamic virtual environments. Particularly, they used the ray tracing which gives realistic acoustic simulation in interactive environments and the frustum tracing for the volumetric representation.

Moreover, Stavrakis, Tsingos and Calamia [98] presented a novel graph-based topological sound propagation algorithm that can compute interactive reverberation effects in complex coupled environments, in which they used auditory masking and scalable Fourier domain processing to render a large number of reverberated components. Their work clearly had some limitations, but it was a springboard for

the development of interesting real-time sound propagation methods and it could find applications both for acoustical design and virtual environments. In the same token, the group of Taylor [99] presented an interactive algorithm which combines both geometric propagation techniques to compute the propagation paths and a ray-based underlying representation that is used to compute specular/diffuse reflections and edge diffraction. Similarly, Pohl and Stephenson [100] described a new idea of combining ray tracing with the radiosity method to a very efficient geometric simulation method including diffraction and scattering. After the evaluation of the proposed hybrid algorithm, they reached the conclusion that the computation time was reduced from exponential to linear growth with split-up of sound particles, but the method became ineffective in case of no split-up.

Additionally, Yeh et al. [4] presented a novel hybrid approach that combines both geometric and numerical acoustic techniques for interactive sound propagation in complex environments. Specifically, they used wave-based techniques to pre-compute the pressure field in the near-object regions and geometric propagation techniques in the far-field regions to simulate the sound propagation in a scene. They demonstrated that their system was able to simulate high-fidelity acoustic effects such as diffraction, scattering, low-pass filtering behind obstruction, reverberation, and high-order reflections in large, complex indoor and outdoor environments with a satisfactory realistic result. Also, the pressure computation requires orders of magnitude lower memory than standard wave-based numerical techniques. Moreover, Pelzer, Masiero, Vorländer [101] proposed a hybrid reproduction approach, in order to succeed a realistic and natural sounding high quality auralization of sound sources in enclosures, by using binaural technology including near-field effects for close sources and employment of individual head-related transfer functions. In the same year, 2014, the group of Schissle [10] developed a hybrid algorithm with the merger of radiosity and path tracing techniques. They managed to address the problem of interactive sound propagation and rendering in large-scale virtual environments which are composed of multiple moving sources and objects. They demonstrated an order of magnitude performance improvement over previous methods, through the performance of their method in complex indoor and outdoor environments.

Furthermore, in [102] a novel algorithm was recommended to accurately solve the wave equation for dynamic sources and listeners using a combination of pre-computation techniques and GPU-based runtime evaluation. It was proved a significant improvement in runtime memory comparing with prior wave-based techniques which were applied to large scenes with moving sources. Likewise, Charalampous and Michael [103] introduced a hybrid sound propagation algorithm in which image source method was used to calculate sound reflections from specular surfaces and a prioritized ray tracing algorithm for fast detection and evaluation of valid image sources from the tree of candidate image sources. After the comparison with other algorithms, such as ray tracing algorithm and best first image source algorithms, they demonstrated that acoustical results were improved in most cases. A further important publication is from Podkosova group [104] which developed a hybrid sound model based on the image source method and the secondary sound sources for late reflections and reverberation. The motivation of this approach is that a complex real-time hybrid model enhances task performance in 3D audio games when compared to a basic model. In addition to the previous work, Rungta et al. [105] developed a coupled sound synthesis-propagation algorithm that can generate realistic sound effects for computer games and virtual reality, by combining modal sound synthesis, sound radiation, and sound propagation. In order to perform the sound propagation, they used the fast ray tracing technique to compute the impulse responses using perceptual Hankel approximation. Add to this, with the use of 3D virtual complex indoor and outdoor scenes, they confirmed that the proposed method can handle a high degree of dynamism in term of source radiation and propagation in complex scenes.

Finally, Schissler and Manocha [35] from the University of North Carolina at Chapel Hill presented an interactive algorithm for sound propagation and rendering in complex, dynamic scenes with a large number of sources, which combined fast backward ray tracing from the listener with sound source clustering to compute propagation paths. They demonstrate their algorithm's performance on complex indoor and outdoor scenes with high acoustic complexity and observe significant speedups over prior algorithms.

B. Web 3D Spatial Sound

Many approaches for Web 3D applications aimed at realistic visualization of the scene. As in the above desktop embodiments so in web applications, spatial sound can offer further details to a 3D graphic world.

As a result, the first attempt took place with the use of the <bgsound> tag, in which only background music could be contained in a web page and was not being supported from all browsers. After that, flash was the first cross-browser way of audio on the Web, but a key limitation of this research was the requirement of the plugins installation. Moreover, the focus of following research was concentrated on the element <audio> in HTML5, which could avoid the plugins, but was not designed for sophisticated and complex applications [106], [107]. Particularly, the element <audio> is inferior to apply filters to the sound signal and access the raw PCM data. Furthermore, the orientation both of listener and sound source is not available; as a result the proposed method cannot be readily used in practice. Lastly, it does not afford low-latency precise-timing model, which is very important in order to develop interactive applications, with fast auditory response to user actions [108]. Thus, it is not adequate for a 3D interactive web scene with demanding sound design.

Under these circumstances, several alternatives have been proposed, in order to establish an effective API, which attends to overcome the most of these limitations. One of the most interesting approaches to this issue is Web Audio API, which has been proposed by Mozilla Foundation.

Indeed, it was not a coincidence that Web Audio API has gained much attention from researchers in the last years. In particular, the literature demonstrates a variety of studies which utilize Web Audio API, in order to accomplish the sound in browser. For instance, Choi and Berger [109] developed a JavaScript library which is based on the Web Audio API in order to facilitate music in the web environment, bypassing underlying tasks and augmenting useful features. Additionally [110], an innovator framework was described which used the Web Audio API to render object-based 3D audio in a web browser without requiring plugins. Similarly, Rawlinson et al. [111] have also presented an audio feature extractor library using Web Audio API. Their study was a lightweight implementation which was characterized by flexibility and adaptability to introduce audio in a web system. Furthermore, a similar work was introduced by Kleimola and Larkin [112], which included audio effects for web browsers and built on top of Web Audio API. The strong point of this work was the directly loading from the open web without manual installations. Moreover, Pendharkar et al. [113] proposed a new engine which was ported from Adobe's Flash platform to Web Audio API. The main objective of this study was to overcome the differences of the architecture between Adobe's Flash and Web Audio API (the most well-known techniques for web auralization in literature). As a result, many applications had not to be redesigned to work both of them. Another approach, which based on the Web Audio API, was analyzed by Schnell group [114]. This module analyzed a novel solution for the synchronization, schedule and aligning of the

audio playback in the internet environments. Finally, Mahadevan et al. [115] described a novel learning environment which offered a web platform for teaching computer science through algorithmic music composition.

This fact seems to be justified because Web Audio API is open source and be supported from the most browsers. Except that, it offers multi-channel audio and high-level sound abilities as filters, delay lines, amplifiers, spatial effects (such as panning). Also, audio channels can have 3D distribution according to the position, speed or direction of the viewer and the sound source.

Different from the above publications, the first systematic study for the implementation of spatial sound in web presented from the group of Stamoulias [116] in which they enriched the X3DOM framework (an open source JavaScript framework, used to create declarative 3D scenes in Web pages) with spatial sound features, using both the X3D (a royalty-free ISO standard XML-based file format for representing 3D computer graphics in web) sound nodes and the structure of Web Audio API.

On the other hand, due to the continuous interest of high degree of realistic in web 3D environments, neither the Web Audio API nor any other approach enrichment them with immersive sound characteristics. Particularly, characteristics such as surface reflection, reverberation, physical phenomena including interference and diffraction have not yet been included in web 3D scenes, even though they play a major role in the representation of a realistic 3D sound. Consequently, it is understandable that there has been little discussion on interactive web 3D immersive environment with the addition of spatial sound effects by taking into account the geometry of the scene.

III. Summary and Outlook

This review has provided a summary of the most significant research taking place in the field of the spatial sound propagation. Specifically, we tried to present the techniques and the algorithms which have been designed in order to provide spatial sound for dynamic and interactive environments, both for the sound propagation and the audio rendering.

After this study, we observed that there is a vast amount of literature on the auralization area [117]. However, this is not particularly surprising given the fact that audio technology has reached a point where algorithms, hardware, and auditory display technology are becoming standard components and be applicable in many fields (video games development, virtual reality, acoustics engineering and other disciplines).

A further conclusion that emerges from this research is that most of the researchers proposed geometric sound propagation algorithms in order to succeed realistic auralization in large scenes with a high number of objects. This fact is justified on the grounds that geometric techniques can be used for fast computation of propagation paths from a source to a listener and takes into account the most of the physical sound characteristics such as specular reflections, diffuse reflections and edge diffraction. In practice, this approach can give a realistic impression of a dynamic sound environment in real time.

Except the geometric proposed algorithms, a growing body of literature has proposed hybrid methods for the sound propagation, as effective techniques. A satisfactory explanation for this outcome can be the fact that this method combines different algorithms and thus takes the advantages of all of them. For the same reason, hybrid method can overcome the most of the major drawbacks that should be addressed during the spatial sound synthesis.

To give an illustration of the results which are generated by this research, we summarize the principal advantages and features of each sound propagation algorithm, in Table I. The strong point of this table lies in the fact we recommend the method that best suits in any case of acoustic approaches, taking to account a significant body of literature.

Namely, Table I compares the main categories of sound propagation methods and indicates the advantages/disadvantages each of them. Initially, it presents the Acoustic Wave Equation Methods - FEM, BEM, FDTD - which are the most prominent numerical techniques for solving the wave equation. In general these methods are widely used for the reason that they can accurately simulate all acoustic effects with low computational complexity. Specifically, FEM and BEM have traditionally been employed mainly for the steady-state frequency domain response, as opposed to a full time domain solution of the wave equation. As a result, FEM is suitable for the interior and BEM for the exterior scattering problems. The FDTD algorithm, on the other hand, produces better suited impulse responses to auralization than the above methods and it has become a commonly used algorithm in room acoustic modeling. However, the requirements of Acoustic Wave Equation Methods increase significantly for complex scenes, so under those circumstances, they are not recommended to model the sound in interactive virtual environments and are perceived as too slow for real time sound rendering.

Furthermore, the next group of sound rendering algorithms, which is described in Table I, is the category of Geometric Methods. Particularly, the first mentioned method - Image Sources - is a technique that provides accurate results, as it detects all the possible sound reflections in a dynamic scene, as a result it is suitable for sound propagation in indoor and outdoor spaces with geometrically reflecting boundaries and without aliasing issues. Additionally, Ray Tracing is the second geometric method. This algorithm is of widespread interest due to the fact that it handles dynamic scenes, taking to account demanding physical characteristics (such as diffuse reflection) with simplicity and generality. In the same fashion, beam tracing method belongs in the same sound propagation category. It benefits from the fact that it is the fastest commonly used geometric room acoustics modeling technique and can handle moving listener. Also, in comparison with other methods, like ray tracing, beam tracing clearly has an advantage as a deterministic method. Finally, the radiosity is suitable for the simulation of sound propagation in urban auditory environments and predicts room sound fields with some accuracy. Despite the fact that the Geometric Methods have many advantages over Acoustic Wave Equation Methods, they have failed to be applicable to interactive reverberation effects and not to be susceptible to aliasing errors. So, taken together, Geometric Methods are distinguished by the benefits of effectiveness, speed and the efficiency for the computation of complex physical sound phenomena, such as reflections and diffractions. Despite the fact that they have many advantages over Acoustic Wave Equation Methods, they have failed to be applicable to interactive reverberation effects and not to be susceptible to aliasing errors.

The last sound propagation category which is highlighted in Table I, the Hybrid Methods, has been suggested in order to overcome the limitations of the previous methods. In particular, these algorithms can be readily used in practice for realistic sound effects, including reflections, reverberations and succeed significantly reduced calculation times; also, they solved the problem of the refraction in dispersive environment. As a result, they can be used in computer games and virtual reality to generate realistic sound effects.

As was mentioned, more details on this issue are given in Table I.

In conclusion, it is evident from this study that the field of the spatial sound propagation is still required research and development. There has also been much progress in the auralization during the past decade, but the real challenge is the sound modeling in dynamic environments in which we can interactively change everything by including the geometry and materials in real time.

TABLE I. Units for Magnetic Properties

Algorithms		Characteristics	
Acoustic Wave Equation Methods	FEM	FEM takes into account the most of fundamental wave phenomena. As a result, it is very accurate and can be used in dispersive environments. But a number of potential limitations cannot be easily overcome in order to be applicable in complex 3D environments.	• can accurately simulate all acoustic effects • are general and simply • are highly accurate But • cannot be used for large-scale models or high frequencies • memory requirements increase significantly for large virtual environments
	BEM	BEM is better suited for open scenes than FEM.	
	FDTD	FDTD has become a commonly used algorithm to simulate sound in acoustic room, because it produces better suited impulse responses to auralization than FEM and BEM, which typically calculate frequency domain responses.	
Geometric Methods	Image Sources	Image Sources used to calculate the sound propagation in indoor and outdoor spaces with geometrically reflecting boundaries and there are no aliasing issues, especially for dynamic scenes.	• are used to compute early reflections and diffractions in static scenes • can handle complex geometry • quick and effective But • are too computationally intensive to be applicable to interactive reverberation effects • are not unaffected by aliasing errors and may need a very high density of samples to overcome those problems
	Ray Tracing	Ray tracing characterized by simplicity. Also, the speed and the efficiency on GPU hardware are additional strong points of this algorithm. It handles dynamic scenes and the diffuse reflection efficiently.	
	Beam Tracing	It is currently considered as the fastest commonly used geometric room acoustics modeling technique. It can handle moving listener, which offers more realistic results. However, current algorithms take large pre-processing time and are not directly applicable to dynamic scenes with moving objects.	
	Radiosity	It is used in order to simulate sound propagation in urban.	
Hybrid Methods			• realistic sound effects, including reflections, reverberations and succeed significantly reduced calculation times • can simulate sound refraction in dispersive environment • can generate realistic sound effects for computer games and virtual reality

Further work needs to be done in the development of sound propagation algorithms which are based on geometric acoustics and take into account physical phenomena in a web 3D virtual scene. Previous research can only be considered a first step to include not only spatial information, but also physical characteristics of the sound propagation in a browser. To fill this literature gap, the focus of recent research should be on synthesizing and processing high quality audio in web environments. In other words, many acoustic effects including surface reflection, reverberation, physical phenomena such as interference and diffraction, the absorption, coefficient of materials should be taken into account, in order to increase the realism of the sound in a web 3D environment. This may be considered a promising aspect of the auralization to expand the field of immersive sound beyond the limits of the current web 3D technology.

References

[1] M. Naef, O. Staadt and M. Gross, "Spatialized audio rendering for immersive virtual environments", VRST '02 Proceedings of the ACM symposium on Virtual Reality Software and Technology, pp. 65-72, 2002.

[2] M. Pallot, P. Daras, S. Richir and E. Loup-Escande, "3D-live: live interactions through 3D visual environments", Proceedings of the 2012 Virtual Reality International Conference, article No. 23, pp 8 2012.

[3] F. Hamid and B. Kapralos, "A Review of Spatial Sound for Virtual Environments and Games with Graphics Processing Units", The Open Virtual Reality Journal, Volume 1, pp. 8-17, 2009.

[4] H. Yeh, R. Mehra, Z. Ren, L. Antani, D. Manocha and M. Lin, "Wave-ray coupling for interactive sound propagation in large complex scenes", ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia 2013 TOG Homepage archive, Volume 32, pp. 165, Issue 6, 2013.

[5] C. Lauterbach, A. Chandak and D. Manocha, "Interactive sound rendering in complex and dynamic scenes using frustum tracing", IEEE Transactions on Visualization and Computer Graphics, Volume. 13, Issue 6, pp. 1672 – 1679, 2007.

[6] A. Sterling and M. C. Lin, "Interactive modal sound synthesis using generalized proportional damping", I3D '16 Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, pp. 79-86, 2016.

[7] M. Labschütz, "Realistic real-time audio rendering in virtual environments", cg.tuwien.ac.at, 2011.

[8] M. Vorländer, "Auralization: Fundamentals of Acoustics", Modelling, Simulation, Algorithms, and Acoustic Virtual Reality, Springer-Verlag, Berlin, pp. 335, ISBN: 978-3-540-48829-3, 2008.

[9] T. Funkhouser, N. Tsingos, and J. M. Jot, "Survey of Methods for Modeling Sound Propagation in Interactive Virtual Environment Systems", Citeseer, 2003.

[10] C. Schissler, R. Mehra and D. Manocha, "High-order diffraction and diffuse reflections for interactive sound propagation in large environments", ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH, Volume 33, pp 12, Issue 4, 2014.

[11] R. Mehra and D. Manocha, "Wave-based sound propagation for VR applications", VR Workshop: Sonic Interaction in Virtual Environments (SIVE), pp 41-46, 2014.

[12] N.Atalla and R.J.Bernhard, "Review of numerical solutions for low-frequency structural-acoustic problems", Applied Acoustics, Volume 43, Issue 3, pp 271-294, 1994.

[13] KJ Bathe, "Finite element procedures". Publisher: Prentice Hall, Pearson Education, Inc., First Edition, ISBN 978-0-9790049-5-7, 1996.

[14] L. Savioja, T. Rinne and T. Takala, "Simulation of room acoustics with a 3D finite difference mesh", Proceedings of Intl. Computer Music Conf. (ICMC94), pp. 463–466, 1994.

[15] N. Tsingos, C. Dachsbacher, S.Lefebvre and M. Dellepiane, "Instant sound scattering", EGSR'07 Proceedings of the 18th Eurographics conference on Rendering Techniques, pp. 111-120, 2007.

[16] D. Quinn, "A finite element method for computing sound propagation in ducts containing flow", 5th Aeroacoustics Conference, https://doi.org/10.2514/6.1979-661, 1979.

[17] L. L. Thompson, "A review of finite-element methods for time-harmonic acoustics", The Journal of the Acoustical Society of America, Volume. 119, pp 1315-1330, Issue 3, 2006.

[18] Y.T. Chou and S.Y. Hsia, "Numerical Analysis of Indoor Sound Quality Evaluation Using Finite Element Method", Mathematical Problems in Engineering, Volume 2013, pp. 8, 2013.

[19] R. Mehra, N. Raghuvanshi, L. Antani, A. Chandak, S. Curtis and D. Manocha, "Wave-based sound propagation in large open scenes using an equivalent source formulation", ACM Transactions on Graphics (TOG), Volume 32, pp 13, Issue 2, 2013.

[20] L. Savioja, "Modeling techniques for virtual acoustics", Ph.D. dissertation, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Helsinki, Finland, 1999.

[21] R. Mehra, R, L. Antani, S. Kim and D. Manocha, "Source and Listener Directivity for Interactive Wave-Based Sound Propagation", IEEE Transactions on Visualization and Computer Graphics, Volume 20, Issue 4, pp. 495 – 503, 2014.

[22] L.H. Chen and D.G. Schweikert, "Sound Radiation from an Arbitrary Body", The Journal of the Acoustical Society of America, Volume 35, pp 1626-1632, Issue 10, 1963.

[23] K. A. Cunefare and G. Koopmann, "A boundary element method for acoustic radiation valid for all wavenumbers", The Journal of the Acoustical Society of America, Volume 85, pp 39, Issue 1, 1989.

[24] J. Zhenlin M. Qiang and Z. Zhihua, "Application of the Boundary Element Method to Predicting Acoustic Performance of Expansion Chamber Mufflers With Mean Flow", Journal of Sound and Vibration, Volume 173, Issue 1, pp 57-71, 1994.

[25] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation", The Journal of the Acoustical Society of America, Volume 110, pp 2440-2448, Issue 5, 2001.

[26] W. Haijun, L. Yijun and J. Weikang, "A fast multipole boundary element method for 3D multi-domain acoustic scattering problems based on the Burton–Miller formulation", Engineering Analysis with Boundary Elements, Volume 36, pp 779-788, Issue 5, 2012.

[27] R. Blumricha and D. Heimann, "A linearized Eulerian sound propagation model for studies of complex meteorological effects", The Journal of the Acoustical Society of America, Volume 112, pp 446, Issue 2, 2002.

[28] E.M. Salomons, R. Blumrich and D. Heimann, "Eulerian time-domain model for sound propagation over a finite-impedance ground surface. Comparison with frequency-domain models", Acta Acustica united with Acustica, Volume 88, pp 483-492, 2002.

[29] K. Heutschi, M. Horvath and J. Hofmann, "Simulation of Ground Impedance in Finite Difference Time Domain Calculations of Outdoor Sound Propagation", Acta Acustica united with Acustica, Volume 91, pp 35-40(6), 2005.

[30] V.E. Ostashev, D.K. Wilson, L. Liu, D.F. Aldridge, N.P. Symons and D. Marlin, "Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation", The Journal of the Acoustical Society of America, Volume 117, pp 503-517, Issue 2, 2005.

[31] S. Van Duyne and J. O. Smith. "The 2-d digital waveguide mesh", Applications of Signal Processing to Audio and Acoustics, pp 177–180 (1993).

[32] L. Savioja, T. Lokki and V. Välimäki, "The interpolated 3-D digital waveguide mesh method for room acoustic simulation and auralization", Conference: Joint Baltic-Nordic Acoustical Meeting, Volume 2, pp 48-52, 2002.

[33] D. T. Murphy, "The digital waveguide mesh: Applications in acoustic modeling and current research directions", The Journal of the Acoustical Society of America, Volume 116, pp 2562-2562 ,Issue 4, 2004.

[34] D. Manocha and M.C. Lin, "Interactive sound rendering", Computer-Aided Design and Computer Graphics, 2009.

[35] C. Schissler and D. Manocha, "Interactive sound propagation and rendering for large multi-source scenes", ACM Transactions on Graphics (TOG), Volume 36, pp 4, Issue 1, 2017.

[36] N. Raghuvanshi, J. Snyder, R. Mehra, M. Lin and N. Govindaraju, "Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes", ACM Transactions on Graphics, Volume 29, Issue 4, 2010.

[37] T. Lokki, L. Savioja, R. Vaananen, J. Huopaniemi and T. Takala, "Creating interactive virtual auditory environments", IEEE Computer Graphics and Applications, Volume. 22, pp 49 – 57, Issue 4, 2002.

[38] T. Lokk and L. Savioja, "State-of-the-art in auralization of concet hall models - What is still missing?", Baltic-Nordic Acoustics Meeting, 2008.

[39] E.; Tavazzi and F. Avanzini, "Distance rendering and perception of nearby virtual sound sources with a near-field filter model", Applied Acoustics, Volume 115, pp 61–73, 2017.

[40] C. Lauterbach, A. Chandak and D. Manocha, "Interactive sound rendering in complex and dynamic scenes using frustum tracing", IEEE Transactions on Visualization and Computer Graphics, Volume 13, pp 1672-1679, Issue 6, 2007.

[41] L. Savioja, "Real-time 3D finite-difference time-domain simulation of low-and mid-frequency room acoustics", Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria, 2010.

[42] N. Raghuvanshi and J. Snyder, "Parametric wave field coding for precomputed sound propagation", ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH, Volume 33, Issue 4, pp 11, 2014.

[43] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi and J.West, "A beam tracing approach to acoustic modeling for interactive virtual environments", SIGGRAPH '98 Proceedings of the 25th annual conference on Computer graphics and interactive techniques, pp 21-32, 1998.

[44] L. Savioja and U.P. Svensson, "Overview of geometrical room acoustic modeling techniques", The Journal of the Acoustical Society of America, Volume 138, pp 708-730, Issue 2, 2015.

[45] S. Lauri, H. Jyri, L. Tapio and V. Ritta, "Creating interactive virtual acoustic environments", JAES, Volume 47, Issue 9, pp 675-705, 1999.

[46] B.M. Gibbs and D.K. Jones, "A Simple Image Method for Calculating the Distribution of Sound Pressure Levels within an Enclosure", Acta Acustica united with Acustica, Volume 26, pp 24-32, 1972.

[47] F. Santon, "Numerical prediction of echograms and of the intelligibility of speech in rooms", The Journal of the Acoustical Society of America, Volume 59, pp 1399-1405, Issue 6, 1976.

[48] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics", Acoustical Society of America, Volume 65, pp 943-950, Issue 4, 1979.

[49] J. Borish, "Extension of the Image Model to Arbitrary Polyhedra", The Journal of the Acoustical Society of America, Volume 75, pp 1827-1836, 1984.

[50] H. Lee and B-H. Lee, "An efficient algorithm for the image model technique", Applied Acoustics, Volume 24, pp 87-115, Issue 2, 1988.

[51] M. Vorländer, "Simulation of the transient and steady state sound propagation in rooms using a new combined ray tracing/image source algorithm", The Journal of the Acoustical Society of America, Volume 86, pp 172-178, 1989.

[52] R. Heinz, "Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail", Applied Acoustics, Volume 38, Issues 2–4, pp 145-159, 1993.

[53] E.A. Lehmann and A.M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model", The Journal of the Acoustical Society of America, Volume 124, pp 269-277, Issue 1, 2008.

[54] S. G. McGovern, "Fast image method for impulse response calculations of box-shaped rooms", Applied Acoustics, Volume. 70, Issue 1, pp 182–189, 2009.

[55] O. André, C. Guilherme, D. Paulo, M. D. Thomas, V. José, M. Catarina and S. Jorge, "Real-time dynamic image-source implementation for auralization", Proceedings of the 16th International Conference on Digital Audio Effects. 16th International Conference on Digital Audio Effects, pp 368-372, 2013.

[56] P. Charalampous and D. Michael, "Tree Traversal Algorithms for Real Time Sound Propagation Calculation", 55th International Conference: Spatial Audio, pp 4-1, 2014.

[57] P. Svensson, "The Early History of Ray Tracing in Room Acoustics", Modern Acoustics and Signal Processing, chapter 2, Part of the series Modern Acoustics and Signal Processing, pp 15-31, 2015.

[58] A. Krokstad, S. Strom and S. Sørsdal, "Calculating the acoustical room

response by the use of a ray tracing technique", Journal of Sound and Vibration, Volume 8, pp 118-125, Issue 1, 1968.

[59] A. Kulowski, "Error investigation for the ray tracing technique", Applied Acoustics, Volume 15, pp 263-274, Issue 4, 1982.

[60] A. Kulowski, "Algorithmic representation of the ray tracing technique", Applied Acoustics, Volume 18, pp 449-469, Issue 6, 1985.

[61] H. Lehnert, "Systematic errors of the ray-tracing algorithm", Applied Acoustics, Volume 38, pp 207-221, Issues 2–4, 1993.

[62] D.V. Maercke and J. Martin, "The prediction of echograms and impulse responses within the Epidaure software", Applied Acoustics, Volume 38, pp 93-114, Issues 2–4, 1993.

[63] K.M. Li, S. Taherzadeh and K. Attenborough, "An improved ray-tracing algorithm for predicting sound propagation outdoors", The Journal of the Acoustical Society of America, Volume 104, pp 2077-2083, Issue 4, 1998.

[64] W. Mueller and F. Ullmann, "A scalable system for 3D audio ray tracing", Multimedia Computing and Systems – Italy, IEEE, 1999.

[65] A. Alpkocak and K. Sis, "Computing Impulse Response of Room Acoustics Using the Ray-Tracing Method in Time Domain", Archives of Acoustics, Volume 35, pp 505–519, Issue 4, 2010.

[66] M. Dreher, G. Dutilleux and F. Junker, "Optimized 3D ray tracing algorithm for environmental acoustic studies", Proceedings of the Acoustics 2012 Nantes Conference, 2012.

[67] M. Okada, T. Onoye and W. Kobayashi, "A Ray Tracing Simulation of Sound Diffraction Based on the Analytic Secondary Source Model", IEEE Transactions on Audio, Speech, and Language Processing, Volume 20, pp 2448 – 2460, Issue 9, 2012.

[68] M. Taylor, A. Chandak, Q. Mo, C. Lauterbach, C. Schissler and D. Manocha, "Guided Multiview Ray Tracing for Fast Auralization", IEEE Transactions on Visualization and Computer Graphics, Volume 18, pp 1797 – 1810, Issue 11, 2012.

[69] Q. Mo, H. Yeh and D. Manocha, "Tracing Analytic Ray Curves for Light and Sound Propagation in Non-linear Media", IEEE Transactions on Visualization and Computer Graphics, Volume 22, pp 2493 – 2506, Issue 11, 2016.

[70] Q. Mo, H. Yeh, M. Lin and D. Manocha, "Analytic ray curve tracing for outdoor sound propagation", Applied Acoustics, Volume 104, pp 142–151, 2016.

[71] C. Schissler and D. Manocha, "Adaptive impulse response modeling for interactive sound propagation", Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, pp 71-78, 2016.

[72] P. Charalampous and D. Michael, "Sound propagation in 3D spaces using computer graphics techniques", Virtual Systems & Multimedia, 2015.

[73] T. Lewers, "A combined beam tracing and radiant exchange computer model of room acoustics", Applied Acoustics, Volume 38, pp 161-178, Issues 2–4, 1993.

[74] T. Funkhouser, P. Min and I. Carlbom, "Real-time acoustic modeling for distributed virtual environments", Proceedings of the 26th annual conference on Computer graphics and interactive techniques, pp 365-374, 1999.

[75] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi and J. West, "Interactive Acoustic Modeling of Complex Environments", The Journal of the Acoustical Society of America, Volume 105, pp 1357-1358 , Issue 2, 1999.

[76] N. Tsingos, T. Funkhouser, A. Ngan and I. Carlbom, "Modeling acoustics in virtual environments using the uniform theory of diffraction", Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp 545-552, 2001.

[77] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E West, G. Pingali, P. Min and A. Ngan, "A beam tracing method for interactive architectural acoustics", The Journal of the Acoustical Society of America, Volume 115, pp 739-756, Issue 2, 2003.

[78] R. Ajaj, L. Savioja and C. Jacquemin, "Software platform for real-time room acoustic visualization", Proceedings of the 2008 ACM symposium on Virtual reality software and technology, pp 247-248, 2008.

[79] S. Laine, S. Siltanen, T. Lokki and L. Savioja, "Accelerated beam tracing algorithm", Applied Acoustics, Volume 70, Issue 1, pp 172–181, 2009.

[80] F. Antonacci, M. Foco, A. Sarti and S. Tubaro, "Fast Tracing of Acoustic Beams and Paths Through Visibility Lookup", IEEE Transactions on Audio, Speech, and Language Processing, Volume 16, Issue 4, pp 812 – 824, 2008.

[81] F. Antonacci, A. Sarti and S. Tubaro, "Two-Dimensional Beam Tracing from Visibility Diagrams for Real-Time Acoustic Rendering", Proceeded of the 7th Int. Conference on Digital Audio Effects, 2010.

[82] T. Funkhouser, J.-M. Jot and N. Tsingos, "Sounds Good to Me!" Computational Sound for Graphics, Virtual Reality, and Interactive Systems, SIGGRAPH 2002 Course Notes, 2002.

[83] P. Svensson and U.R. Kristiansen, "Computational Modelling and Simulation of Acoustic Spaces", 22nd International Conference: Virtual, Synthetic, and Entertainment Audio, 2002.

[84] A. Rajkumar, B. F. Naylor, F. Feisullin and L. Rogers, "Predicting RF coverage in large environments using ray-beam tracing and partitioning tree represented geometry", Wireless Networks, Volume 2, pp 143–154, Issue 2, 1996.

[85] E. Deines, M. Bertram, J. Mohring, J. Jegorovs, F. Michel, H. Hagen and G.M. Nielson, "Comparative Visualization for Wave-based and Geometric Acoustics", IEEE Transactions on Visualization and Computer Graphics, Volume. 12, pp 1173 – 1180, Issue 5, 2006.

[86] Z. Yamauti, "The light flux distribution of a system of interreflecting surfaces", J. Opt. Soc. Am., Volume 13, pp 561–571, Issue 5, 1926.

[87] E.M. Nosal, "Room sound field prediction by acoustical radiosity master thesis", 2002.

[88] J. Shi, A. Zhang, J. Encarnacao and M. Göbel, "A modified radiosity algorithm for integrated visual and auditory rendering", Computers & Graphics, Volume 17, pp 633-642, Issue 6, 1993.

[89] N. Tsingos and J.D. Gascuel, "A general model for the simulation of room acoustics based on hierarchical radiosity", SIGGRAPH'97 technical sketch, 1997.

[90] N. Tsingos and J.D. Gascuel, "Acoustic simulation using hierarchical time-varying radiant exchanges", iMAGIS-GRAVIR/IMAG – Citeseer, 1997.

[91] E.M. Nosal and M. Hodgson, "Preliminary experimental validation of the radiosity technique for predicting room sound fields", Citeseer.

[92] E.M. Nosal, M. Hodgson and I. Ashdown, "Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms", The Journal of the Acoustical Society of America, Volume 116, pp 970-980, Issue 2, 2004.

[93] J. Kang, Application of radiosity method in acoustic simulation, ICA, 2004.

[94] S. Siltanen, T. Lokki, S. Kiminki and L. Savioja, "The Room Acoustic Rendering Equation", The Journal of the Acoustical Society of America, Volume 122, pp 1624-1635, Issue 3, 2007.

[95] R.T. Muehleisen, "Six Surface Steady State Acoustic Radiosity", Architectural Engineering Conference (AEI), 2008.

[96] S.M. Pasareanu, "A numerical hybrid method for modeling outdoor sound propagation in complex urban environments", master thesis, URL: http://hdl.handle.net/10919/47601, 2014.

[97] M. Sikora, "Algorithms for simulations of sound and ultrasound propagation in complex dispersive environments", Proceedings Elmar 2004. 46th International Symposium, 2004.

[98] E. Stavrakis, N. Tsingos and P. Calamia, "Topological Sound Propagation with Reverberation Graphs", Acta Acustica united with Acustica, Volume 94, pp 921-932, Number 6, 2008.

[99] M. Taylor, A. Chandak, L. Antani and D. Manocha, "RESound: Interactive Sound Rendering for Dynamic Virtual Environments", MM '09 Proceedings of the 17th ACM international conference on Multimedia, pp 271-280, 2009.

[100] A. Pohl and U.M. Stephenson, "A Combination of the Sound Particle Simulation Method and the Radiosity Method", Building Acoustics, Volume 18, pp 97-122, Issue 1/2, 2011.

[101] S. Pelzer, B. Masiero and M. Vorländer, "3D Reproduction of Room Auralizations by Combining Intensity Panning", Crosstalk Cancellation and Ambisonics, Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, pp 182-188, 2014.

[102] R. Mehra, A. Rungta, A. Golas, M. Lin and D. Manocha, "WAVE: Interactive Wave-based Sound Propagation for Virtual Environments", IEEE Transactions on Visualization and Computer Graphics, Volume 21, pp 434 – 442, Issue 4, 2015.

[103] P. Charalampous and D. Michael, "Improved hybrid algorithm for real time sound propagation using intelligent prioritization", Electrotechnical Conference (MELECON), 18th Mediterranean, 2016.

[104] I. Podkosova, M. Urbanek and H. Kaufmann, "A Hybrid Sound Model

for 3D Audio Games with Real Walking", Proceedings of the 29th International Conference on Computer Animation and Social Agents, pp 189-192, 2016.
[105] A. Rungta, C. Schissler, R. Mehra, C. Malloy, M. Lin and D. Manocha, "SynCoPation: Interactive Synthesis-Coupled Sound Propagation", IEEE Transactions on Visualization and Computer Graphics, Volume 22, pp 1346 – 1355, Issue 4, 2016.
[106] B. Smus, "Web Audio API". O'Reilly Media, Inc, 2013.
[107] R. Spuy, "Sound with the Web Audio API. In Advanced Game Design with HTML5 and JavaScript". IEEE, 2015.
[108] B. Smus, "Retrieved from Developing Game Audio with the Web Audio API", Available online: http://www.html5rocks.com/en/tutorials/webaudio/games/ (accessed May 2015).
[109] H. Choi and J. Berger, "WAAX: Web Audio API eXtension", International Conference on New Interfaces for Musical Expression, KAIST, Daejeon, Korea, 2013
[110] C. Pike, P. Taylour and F. Melchior, "Delivering Object-Based 3D Audio Using The Web Audio API And The Audio Definition Model", WAC - 1st Web Audio Conference, 2015.
[111] H. Rawlinson, N. Segal and J. Fiala, "Meyda: an audio feature extraction library for the Web Audio API", WAC - 1st Web Audio Conference, 2015.
[112] J. Kleimola and O. Larkin, "Web Audio Modules", 12th Sound and Music Computing Conference, Maynooth, Ireland, 2015.
[113] C. Pendharkar, P. Bäck and L. Wyse, "Adventures in scheduling, buffers and parameters: Porting a dynamic audio engine to Web Audio", WAC - 1st Web Audio Conference, 2015.
[114] N. Schnell, V. Saiz, K. Barkati and S. Goldszmidt, "Of Time Engines and Masters An API for Scheduling and Synchronizing the Generation and Playback of Event Sequences and Media Streams for the Web Audio API", WAC - 1st Web Audio Conference, 2015.
[115] A. Mahadevan, J. Freeman and B. Magerko, "EarSketch: Teaching computational music remixing in an online Web Audio based learning environment", WAC - 1st Web Audio Conference, 2015.
[116] A. Stamoulias, E. Lakka and A.G. Malamos, ""Wrapping X3DOM around Web Audio API", International Journal of Interactive Multimedia and Artificial Intelligence, Volume 3, pp 36-46, Number 5, 2015.
[117] M. Cohen, J. Villegas and W. Barfield, "Special issue on spatial sound in virtual, augmented, and mixed reality environments", Virtual Reality, Volume 19, pp 147–148, Issue 3, 2015.

E. Lakka

Eftychia Lakka was born in 1982. She received her BSc degree in Computer Science from the University of Ioannina, Greece (2006) and her M.Sc. in Synthesis of Images and Graphic Designs from the University of Limoges, France (2011). Since October 2016, she has started the MPhil/PhD in School of Computing and Mathematics within the Faculty of Computing, Engineering and Science in the University of South Wales, UK. She has been as Research Associate at the Advanced Knowledge, Image & Information Systems Laboratory, dept. of Informatics, T.E.I. of Athens (2010-2014) and at the Multimedia Content Laboratory, dept. of Informatics Engineering, T.E.I. of Crete (2014- 2015). Since April 2016, she has joined Information Technologies Institute of Centre for Research and Technology as research assistant. Her research interests include 3D computer games, 3D sound propagation, graphics and multimedia web applications.

A. G. Malamos

Athanasios G. Malamos was born in 1969. He received his BSc degree in Physics from the University of Crete (1992) and his PhD from the Technical University of Crete in 2000. From 1997 to 2002 he was a Research Assistant and a researcher in the ICCS National Technical University of Athens. Since 2002 he is with the Technological Educational Institute of Crete at the Department of Informatics Engineering as an Assistant Professor (2002-2006) and as an Associate Professor (2006 until present). He is the coordinator of the Multimedia Content Lab, and has supervised many research projects granted with EU and National research funds. He has served as program committee member and reviewer for several international conferences and workshops. Dr. Malamos is a reviewer for IEEE, Springer as other international journals. He is member of the IEEE Computer Society, ACM SIGGRAPH and the WEB3D consortium. His research interests include multimedia semantics, AI and graphics.

K. G. Pavlakis

Konstantinos G. Pavlakis got his B. Sc. in Physics, the M. Sc. on space physics and the Ph. D. on radiation transfer and astrophysical Masers from the department of Physics, University of Crete. He is currently a lecturer in the field of general physics in the department of Mechanical Engineering of the Technological Education Institute of Crete. His scientific interests include modeling of radiation transfer in atmospheres, the study of climate change, and environmental-atmospheric physics. He has published 27 scientific papers in refereed journals and 11 papers in conference proceedings taking part in research projects as a researcher in the University of Leeds, the University of Crete, the Foundation of Research and Technology and the Technological Education Institute of Crete.

J. A. Ware

J. Andrew Ware is a Professor of Computing at the University of South Wales where he is engaged in both teaching and research. His main research area is Applied Artificial Intelligence, much of which is carried out in collaboration will business and industrial partners. Andrew has successfully supervised over 30 PhD students. He is a Fellow of the British Computer Society and Senior Fellow of the Higher Education Academy. Andrew is also Regional Director of Technocamps, an innovative and progressive initiative aimed at increasing the number of young people studying Computer Science at school and university.

TV Series and Social Media: Powerful Engagement Factors in Mobile Video Games

Fernando de Rada[1], Asuncion Mochon[2]*, Yago Saez[3]

[1] Camilo Jose Cela University, Madrid (Spain)
[2] Applied Economics and Economic History Department, UNED University, Madrid (Spain)
[3] Artificial Intelligence Department, Carlos III University of Madrid, Madrid (Spain)

Received 4 October 2018 | Accepted 30 October 2018 | Published 5 November 2018

Abstract

The free-to-play business model has become hegemonic in the mobile video game industry, displacing the traditional paid content model that was the norm until the appearance of manufacturers' app stores. Companies attempt to monetize these games by means of in-game micro-transactions and in-game advertising; thus, it is essential to acquire an enormous number of users because only a small percentage will ultimately make any purchases. To keep players engaged, companies typically put in place marketing and design strategies derived from behavioral telemetry, to maintain a grip on players. We propose an innovative approach, focusing our attention on the impact of having a video game based on a famous TV series. Furthermore, we analyze the effect of social networks on game metrics. The outcome indicates that developing a game based on a TV series and integrating social media with the gameplay improve and reinforce the user's activation, retention and monetization.

Keywords

Free-To-Play, Video Games, Mobile Games, Game Analytics, Monetization, Customer Retention, Character Identification, TV Series, Social Media, Facebook, Digital Marketing.

DOI: 10.9781/ijimai.2018.11.002

I. Introduction

The mobile video game industry has undergone a deep transformation since the launch of manufacturers' app stores. The barriers to market entry for games and applications have been removed, building up a global and highly competitive space where free-to-play video games have created a multi-million-dollar market based on the sales of virtual goods through in-game micro-purchases. Several companies such as King.com and Supercell report daily incomes of approximately $5M and several hundred million Daily Active Users (DAUs) merely from one or two of their free-to-play mobile games [1]-[2]. However, with approximately 500 games launched per day on the iOS digital app store [3], there is an obvious discoverability issue for any brand new content published on it, which hampers the organic growth of its installed base. In such a context, large companies invest heavily in acquiring new players for their games through different in-game advertising networks, which typically work on a best-bid-per-installed-user basis, causing user acquisition costs to skyrocket [4]. Considering that only approximately one to three percent of players will eventually become paying customers [5]-[6], combined with the fact that acquiring each new user can cost in the range of $3.5 [7], companies are forced to maximize the income from paying users and, simultaneously, to retain non-paying users as much as possible to convert them or, at worst, to capture some value from them through in-game advertising, their word-of-mouth recommendations, or being cross-linked to other games in the company's portfolio [8].

For this model to be financially viable, it is essential to implement numerous adjustments in the game parameters through different iterations planned over time to finesse the user experience and improve business key metrics. These adjustments typically derive from the analysis of the behavioral telemetry that is recorded in all installed games and that is sent and consolidated on the server side [9].

In this research, we present a novel study, analyzing the effect of a mobile game based on a TV series using real data. To the best of our knowledge, to date, no one has conducted this type of analysis. The mobile game Red Eagle Origins is based on a very famous Spanish TV series: Red Eagle. It was first released in 2009 and is still on the air, already amounting to nine seasons. Although to a certain extent it has been internationally broadcast, there remain many territories where the TV series has not arrived yet, with the video game having been released in such territories. The video game was first released in 2014 for iOS and Android, with over 315.000 downloads (until January 2017), including players from Spain and many other countries. It seems clear that there is going to be a wide gap between the audience exposed to the TV series and the rest of the players in terms of their knowledge of the personality traits and attractiveness of the main characters, the complexity and diversity of the plot, and the setting and historical context. Therefore, first, we propose comparing the metrics gathered from both sets, measuring the impact in terms of user activation, retention, and monetization. Our initial hypothesis is that the players exposed to the TV series are going to have more retention and monetization, though we are uncertain about the real consequences. Second, we aim to analyze the impact of social networks on user behavior comparing regular users and users logged on Facebook. The hypothesis to be tested is whether social networks can help obtain higher-quality users in terms of their retention metrics.

This paper is structured as follows. Section II presents the state of the art. Section III describes the TV series, Facebook and game

* Corresponding author.
E-mail address: amochon@cee.uned.es

features. The dataset is explained in Section IV. The results comparing Spanish players versus the rest of the world are presented in Section V. Section VI presents the results comparing users involved and non-involved in social networks, such as Facebook. Section VII analyzes the business impact of the outcome and identifies future marketing strategies. Finally, the conclusions are presented in Section VIII.

II. Related Work

Key metrics in the game industry include the following variables: activation, retention and monetization; thus, these features have been widely studied in the literature. However, what makes players engaged? First, we should review this topic to understand the factors that have been identified as influencing players' activation and, ultimately, their retention and monetization.

Finally, we review different concepts such as immersion and character identification in the context of video games, attempting to determine the extent to which these factors have been postulated by studies influencing player enjoyment and engagement.

A. Engagement and Users' Retention and Monetization

Identifying player motivations is one of the important aspects to be considered with respect to game design [10]. Self Determination Theory (SDT) proposes some general human needs, namely, competence (the need to participate in activities in which we feel capable and effective), autonomy (the need to experience freedom in the activities that we choose), and relatedness (the need to feel connected to other people) [11]. Different motivational models for engagement in video games have been proposed by studies applying SDT principles. Some authors focus on the fundamental psychological needs derived from SDT that gameplay may or may not fulfill, suggesting that "the SDT model of need satisfaction predicts sustained engagement over time, as well as the short-term effects of game activities on players' well-being" [12]. Vallerand et al. conceptualized the term passion [13], which was subsequently used by Wang et al. to examine the underlying psychological processes in digital gaming, with harmonious passion being signaled as the pursuit of engagement in an activity by choice and in harmony with other activities and being strongly associated with intrinsic motivation and positive affect [14].

However, considering the interactive nature of game playing, some other components are necessary to build up an optimum game experience [15]. Information Interaction Theory addresses the communication between the user and the computer interface, with the latter being the main vehicle for the transmission of user experiences [16]. O'Brien and Toms derived from this theory some factors that may be relevant to engagement: affective appeal, challenge, feedback, and perceived control [17].

Perhaps the most influential research on what makes interactive experiences enjoyable was conducted by Csikszentmihalyi in regard to his theory of flow [10], which was originally developed to describe the subjective, emotional state of optimal pleasure that arises when someone is absorbed in any activity that is assumed to be valuable. The idea that optimal experience requires a good balance between the challenges of the game and the skills of the user is central to flow. In addition, such experience should be intrinsically rewarding and immersive, involving a sense of personal control and a high degree of concentration and having clear goals and immediate feedback [18]. Flow theory has been applied by other studies, including [19], [20], and [21], among others.

Uses and Gratifications (U&G) theory has been applied by [22] to group the reasons that individuals hold for engaging in videogames. An analysis of data derived from focus group interview sessions led to six dominant dimensions that include: arousal, challenge, competition, diversion, fantasy, and social interaction.

On the other hand, some studies have signaled that aesthetics is an important element of engagement [23]; aesthetics has also been linked to the usability and visual appearance of the user interface as well as to the skills and needs of users [24].

[25] elaborated the importance of competitive elements as determinant of enjoyment in playing computer games. They concluded that the wish to be challenged and to compete with others, with the game AI, or even with one's own previous achievements is most likely the single most important motive for interactively entertaining oneself.

With regard to the instruments for measuring engagement, some attempts to standardize the factors that have a significant influence have been made. [26] developed a questionnaire to measure engagement in video games and to empirically test the extent to which such questions could be used to construct a quantitative measure. [27] developed a self-report instrument of user engagement: the User Engagement Scale (UES). [28] extended this work, investigating the use of the UES in the context of game-based environments.

In regard to user retention, presently, given the high competition and low discoverability of the mobile video game market, it is much more expensive to acquire a new user than to retain an older user [29]. Thus, churn (or the number of players dropping out) prediction has become an interesting area of study. [30] used time series' feature representation based on frequency analysis, from the login records extracted from the real data of an online game, for churn prediction modeling. [8] defined the high-value player segment and formulated churn prediction for two live casual social games using in-game virtual currency as an incentive to retain players. Their results showed that giving in-game currency for free does not have an impact on user retention and suggested that players can only be retained by improving their gameplay experience before the churn event occurs. [31] developed a model for churn prediction using data from five free-to-play commercial games across mobile and web-based social-online platforms, defining two different formal models of churn prediction based on game-agnostic features. [9] also used behavioral telemetry data from commercial games, extracting information on how player engagement evolves over time, identifying common patterns and predicting when players' interest in the game is declining. They found that the average player's interest in playing these games follows a decreasing power-law curve.

Finally, [32] used telemetry data from a commercial sports game to encode gameplay patterns for specific players as feature vectors and applied regression to model player retention. Their conclusions signaled that providing players with the correct challenge is a key feature involved in retention.

B. Immersion and Character Identification

Immersion has been described by many authors as an engaging experience in which players lose awareness of the real world [26]. [33] add some extra characteristics to the previous definition, namely, a distorted sense of time, a high sense of control or a strong involvement perception. Some other authors consider immersion to be specifically related to the psychological experience of engaging with a computer game [21], even if it does not provide the most optimal experience.

Regarding the motivations why people should play video games, immersion has been widely signaled as one of the most important. Bartel's Player Types [34] are a well-known player's taxonomy that has been reviewed by many studies. A good example is an empirically grounded study on player motivations in some popular online games [35], in which a factor analytic approach was taken to group different player motivations into three broad categories, *social, achievement and immersion*, with the last category being described as a sum of the desire to explore, create a persona linked to a story, customize the appearance

of the player's avatar, and escape from real-life stress and problems.

The notion of *identification* with media characters has been widely discussed in media research, well before the emergence of video games. The audience members of a film or TV series or the readers of a novel often become absorbed in the plot and identify with the characters portrayed. There are two different modes of reception typically described by identification researchers: dyadic and monadic. In the dyadic or spectatorship model, media users perceive a social distinction between themselves and the media characters: "*viewers observe characters, evaluate them and respond in specific emotional ways*" [36]. In this respect, [37] holds that televised drama is entertaining because of the emotional response patterns of viewers; *transportation theory* [38] acknowledges a strong sense of connection between media users immersed in mediated narratives with characters that are encountered repeatedly over time. Nevertheless, in the monadic approach, audience members experience the interpretation of the plot as though the events were happening to them "*in a process that consists of increasing loss of self-awareness and its temporary replacement heightened emotional and cognitive connections with character*" [39].

Some studies related to character identification in video games have found a strong correlation between avatar identification and game enjoyment [40]; additionally, [41] find that players of online games who perceive a smaller psychological difference between their avatar and themselves are generally more satisfied and show better retention metrics. Players do not perceive the main character that they are controlling as a distinct social entity: instead, they experience a merging of their own self and the protagonist [36]. This merging process is strongly induced by cognition-based and social dimensions of self-perceptions [39], such as goals, attractiveness, attitudes, successfulness, and respect by others. Players continuously receive and produce information about such dimensions of character attributes. As in interpersonal relationships, in which duration or familiarity is important to the scope of the relationship [42], the longer the audience is exposed to a character, the more likely it will be able to imagine being that character.

Thus regarding our study, it seems reasonable to expect a higher level of character identification in the Spanish subset of players who have been repeatedly exposed to the TV series. This character identification would have an important impact on the game metrics.

III. Tv Series, Social Network and Game Features

A. Red Eagle: The TV Series

Red Eagle is one of the most successful TV series ever broadcast in Spain. This adventure story, frequently peppered with dramatic elements such as underhanded intrigues and torrid romances, is set in the Spanish "Gold Century" around 1660, occurring in the streets of Madrid. The story is based on the wanderings of "Gonzalo de Montalvo", a school teacher who maintains a double life, given that he also embodies Red Eagle, a sort of masked hero. Cared for to its last detail, it has been one of the most expensive TV series ever produced in Spain, with a budget reaching more than a million euro per episode in its early stages [43]. Since its premiere in February 2009, it has been on air for as many as nine seasons, with an average of more than four million spectators and an average audience share of more than 21%[1].

B. Red Eagle Origins: The Game

Red Eagle Origins is the official licensed mobile video game that brings in the best-loved characters of the TV series. The atmosphere of the TV series is recreated by a high-end 3D setting, where a third-person camera keeps the main character always centered on the screen, allowing players to directly control his/her movements in a highly immersive and interactive game experience. The game was released in February 2014 for Android and iOS platforms and quickly reached very good consumer ratings on both (4.73 and 4.20 out of 5 on iOS and Android, respectively)[2]. Upon its release, it also ranked 5th on the Spanish App Store charts in the action games category and reached 5th on the Google Play action games chart, this time coinciding with the beginning of the new season of the TV series in September 2014. Between its release and January 2017, the video game has reached more than 315.000 downloads. Table I depicts the downloads by continent for the analyzed period, differentiating Spanish players. The game has a strong penetration in Africa and the Middle East, mainly in the United Arab Emirates and Egypt, because of the agreement reached by the publisher with the Arabic media group MBC for distribution in its main area of influence.

TABLE I. Downloads by Continent for the Analyzed Period

Continent	Downloads	%
Spain	1120	12%
Rest of the world	8198	88%
Europe	2348	25%
Africa	4887	52%
America	830	9%
Others	133	1%
Total	9318	100%

The game is classified as an "endless runner" type, in which the player-controlled avatar cannot stop its forward momentum, in this case, through the charming streets of 17th-century Madrid. Players have to dodge many static and moving obstacles through complicated zones in a fugue that will become increasingly challenging due to its hectic pace. The game difficulty grows gradually as both the complexity of the scenes and the race speed are increased step by step.

The scene consists of a hierarchical structure of blocks that can be can swapped round or combined with each other, forming an infinite, but versatile, circular path. The structure with a higher hierarchy is named a "Round". Three different Rounds are concatenated, with the last Round acting as an infinite loop for those players who reach that far. The Rounds are composed of five Zones: The Village, The Forest, The Castle, The Village Combat Area and The Castle Combat Area (Table II).

TABLE II. Composition of Rounds and Zones

Round 1	Round 2	Round 3
The Village	The Village	The Village
The Village Combat Area	The Forest	The Village Combat Area
The Forest	The Castle	The Forest
The Castle	The Castle Combat Area	The Castle
The Forest	The Forest	The Castle Combat Area
		The Forest

During the race, players must not only dodge all types of obstacles but also need to collect some useful objects that will enable their main character to face the different challenges that will be presented. One of these challenges that represents an important novelty in this endless runner genre is the introduction of a "combat area". The character is automatically moved through a narrow hallway, and the player has to defeat the enemies that appear, using different weapons that his/her

1 Statistics provided by Globomedia, S.L.

2 Gathered from iOS and Google official sites in February and September 2014.

avatar can improve throughout the game, without causing damage to any innocent passerby. Fig. 1 depicts a screenshot of the video game.

Fig. 1. Screenshots of the third-person game: running action and combat area scenes.

Regarding monetization, since Red Eagle Origins is a free-to-play video game, incomes are based on micro-payments and in-game advertising. First, there is a virtual currency inspired by the Spanish coins of the period ("maravedí"), which players can collect during the race and which are mainly intended to acquire consumables. Then, very engaged players can use real currency to buy customization elements, a different main character with which they tend to identify, or more valuable objects ("power-ups") that let them progress faster in the game.

Further income is derived from in-game advertising. Following each ended game, three different types of ads may be shown: a rewarded video, an interstitial ad, or an ordinary video. Rewarded videos are short pieces typically advertising other games that give a certain amount of virtual currency or any other virtual good to players who complete its visualization. Interstitial and ordinary video ads are full screen images or videos that allow players to directly download the advertised app or game. Fig. 2 shows a screenshot of virtual and real currency in the game.

Fig. 2. Screenshots of the in-game store: virtual and real currency

C. Red Eagle Users in Facebook

Regarding its social features, the game is integrated with Facebook; thus, players can invite their friends to download the video game from the game itself and compete with them. When logged in and during their run, players will be able to check the farthest point that their Facebook friends have reached, which may become an interesting engaging factor. Moreover, the plot of the game revolves around some scrolls that are randomly dispersed along the way and that contain valuable details of the main characters of the TV series. Players will be able to collect them during gameplay and swap duplicates with their Facebook friends. The objective of this social feature is to improve user acquisition and foster player engagement.

IV. Red Eagle Origins Data Set

The Red Eagle Origins analytics engine, developed by Wildbit Studios, is a two-fold distributed instance. On one hand, it runs in each client installation, collecting local data. On the other hand, there is a single server environment that synchronizes the data with all client instances, acting as a global data warehouse. Synchronization is bi-directional, which means not only that the server is able to receive all behavioral data from client installations but also that such client instances may be updated with the specific configuration parameters that the game will be applying from that moment forward.

The analytics engine can gather behavioral data even when the device is offline because such data are stored first in a local database. In this manner, the system can locally store an enormous amount of both transactional and operational data for each user, covering all events occurring in each game session. All local data are regularly uploaded to the server side, where everything is consolidated in one dataset that currently stores approximately 20 million user variables from more than 9.3 million games and more than 5.6 million sessions.

The data can be classified into two large clusters: individual user data and aggregated metrics.

A. Individual User Data

The user data are divided into five groups, depending on the goals pursued:

1. *Identification:* Allows the system to link each user dataset with a unique player, including information such as: installation id, country, language, installation time or operating system.
2. *Activation:* Information related to the players' activity during the first week or after a certain number of games played to identify whether they have had a minimum exposure to the game features. Considering the typology of this video game, which possesses a straightforward mechanics, we have fixed a minimum threshold of 10 games played to evaluate users' activation. The variables included in this group are: installation date, last activity time, number of games played or retention rate (percentage of returning users measured at a regular interval, typically weekly or monthly).
3. *Engagement / Retention:* Addresses the user's engagement level, measuring player frequency, progression and uses of the game features. This data set includes the largest number of variables, among which are: the retention rate for longer periods (typically 30 days), recency (time since the last entry into the game), number of sessions and games played, number of total meters travelled, maximum meters travelled in a single game (high score), experience level reached, number of scrolls, calices and coins collected, number of skills used, and characters unlocked.
4. *Monetization:* This group includes key variables to measure the game financial trend. All transactional data are recorded from both the virtual economy and the real economy, such as: virtual and real currency expenditure, virtual items sold, skills improved, and the number of ads shown and clicked by players.
5. *Social features or viralization:* Comprises all variables related to the integration with Facebook such as number of players who have logged in, number of shares, and number of scrolls shared, among others.

B. Aggregated Metrics

There are many aggregated metrics that are built based on all the gathered user data and cooked on the server side. Some of these metrics are typically taken as Key Performance Indicators (KPIs) in the video game and other digital industries. Some of these KPIs are: number of DAUs, Monthly Active Users (MAUs), the Conversion Rate (CVR) to become a paying user, Average Revenue Per User (ARPU), Average Revenue Per Paying User (ARPPU), or Revenue Per Mille (RPM), related to in-game ads.

The dataset used in the present study includes players who installed the game for the first time from May 5 to July 15, 2016, and registered their behavioral data until September 2, 2016. This time frame allows for at least 49 days of operation for those users who installed the game on July 15, 2016. Table III shows a summary of the raw dataset, which includes 9318 players, of whom 1120 are Spanish players and 8198 are from the rest of the world.

TABLE III. Summary of Raw Dataset

Dataset	Spanish players	Rest of the world
Installation interval	May 5 to July 15, 2016	May 5 to July 15, 2016
Time frame for the experiment	May 5 to September 2, 2016	May 5 to September 2, 2016
Minimum period of operation	49 days	49 days
# of players	1120	8198

As we have observed, the dataset includes more than several hundred attributes pertaining to five different groups. We have identified some appropriate variables from the *identification, activation, engagement* and *monetization* groups that allow us to focus on the meaningful data to compare both subsets. Table IV summarizes the final attribute set for the experiment.

TABLE IV. Final Attribute Set

User Data Group	Final Attribute Set
Identification	Installation ID, Country, Installation time
Activation	# games played, Last activity time
Engagement	# games played, Maximum meters travelled, Total meters travelled, Lifespan
Ads monetization	# in-game ads shown, # clicks on ads, # ads unavailability

V. TV Series Impact

In this section, we evaluate whether there is a significant difference between the audience exposed to the TV series and the rest of the players in terms of the initial interest that they show in the game (*activation)*, in addition to their *retention* and *monetization* metrics.

A. Activation

We consider a player to be activated after he/she has played 10 games. We have assessed this threshold evaluating the average maximum number of meters (high score) reached by players who have played fewer than 10 games, which turns out to be 772.9 meters for the Spanish subset and 763.8 for the rest of the world. Considering that the first zone – "The Village" – has a total length of 1012 meters, the players of both segments have covered, on average, approximately three-fourths of the total distance, in this manner being fairly exposed to the main game mechanics.

Table V summarizes the churn metrics for each subset measuring the following activation variables: number of players who have played fewer than 10 games and number of players who dropped out before 7 days (R7). The results show that the rest of the world segment has a higher rate of early dropout or churn than the Spanish segment in terms of games (79% versus 61%) and number of days played (89% versus 80%). This outcome suggests that having players exposed to the TV series has a positive effect on the players' activation.

TABLE V. Churn Metrics (Early drop)

	Spanish players	Rest of the world players	Total
Total # players	1120	8198	9318
# players games < 10	683	6474	7157
Percentage	61%	79%	77%
# players drop out before R7	899	7267	8166
Percentage	80%	89%	88%

B. Engagement/Retention

For an evaluation of engagement and retention, we focus on only those players who can be considered *activated,* i.e., those who have played more than 10 games. For the retention metrics, we have selected four variables that may indicate the extent to which players are engaged in the game: number of games played, maximum meters travelled, total meters per user and lifespan. Table VI shows the results for these variables.

The results show that users exposed to the TV series, that is, Spanish players, have a higher retention rate for all variables tested: they play more games, have a higher maximum number of meters, travel more meters, and play for more days than players from the rest of the world. These variables do not follow a normal distribution[3]; thus, we have used nonparametric tests to check whether the differences between both sets are significant. Because we do not want our test to be sensitive to differences in the general shapes of the distributions, we discard the Kolmogorov-Smirnov (KS) test and, instead, perform the Mann-Whitney U test, which tests for differences in the location of two samples [44]. The results obtained indicate that the differences are significant for all variables ($p < 0.05$).[4]

We have graphically compared the normalized data by means of their density plots; see Fig. 3. The density plot for the number of games played is depicted in Fig. 3.a. It shows that the number of players who drop out early is higher in the rest of the world set. Nevertheless, as we move toward the right-hand side of the curve, the Spanish segment density plot is always above, showing the stochastic dominance that the U test revealed. Similarly, at the beginning, the rest of the world density plot is above the Spanish set for the maximum number of meters (Fig. 3.b), total meters (Fig. 3.c) and lifespan (Fig. 3.d), but as we spread toward the right-hand side of the figure, the Spanish plot is always above. These figures confirm that players exposed to the TV series are more engaged, most likely because they identify with the TV characters, increasing the retention ratio.

This difference in players' engagement is ultimately determined by the retention rate curve, which measures what percentage of players return to the game throughout the days since their installation date (Fig. 4).

Again, performing a Mann-Whitney U test, we can verify that the differences between both sets are significant, where a *p-value* = 0.0001063 confirms the dominance of the Spanish set in terms of the retention rate. Typically, the retention rate is measured at days one

3 It has been verified performing a Shapiro-Wilk normality test, obtaining p-values well below 0.05 for both segments in all variables.

4 Number of games p = 7.98e-13, Max meters p = 1.313e-15, Total meters p = 2.2e-16 and Lifespan p = 1.711e-05.

(R1), seven (R7), and thirty (R30), with R1 accounting for a very early drop out, R7 being the activation threshold, and R30 being the long-run reference metric. Although we observe similar values at R1 for both sets (R1_ES = 75.5%; R1_REST = 74.1%), as we progress throughout the number of days since installation, the retention values tend to diverge, with differences of approximately 13% and 12% for R7 and R30, respectively, as shown in Table VII.

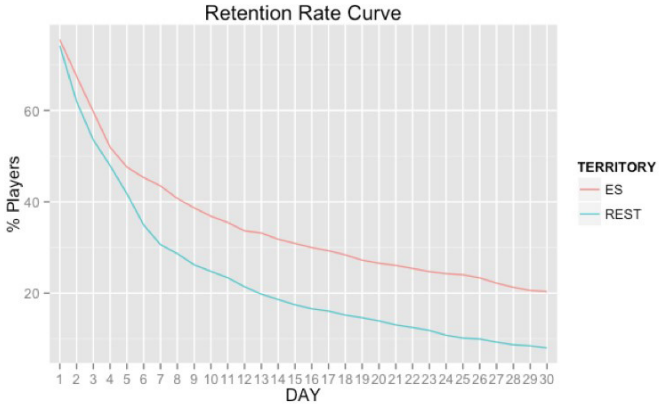


Fig. 4. Graphic comparison of the retention rate curves.

TABLE VII. RETENTION RATE

	Spanish players	Rest of the world players
R1	75.5%	74.1%
R7	43.5%	30.6%
R30	20.4%	8%

C. Ads Monetization

An important piece of information for the mobile video game industry is knowing whether the value of acquired users can also depend on factors such as character identification by means of a TV series. Therefore, in this section, the impact of being exposed to the TV series is measured in terms of monetization.

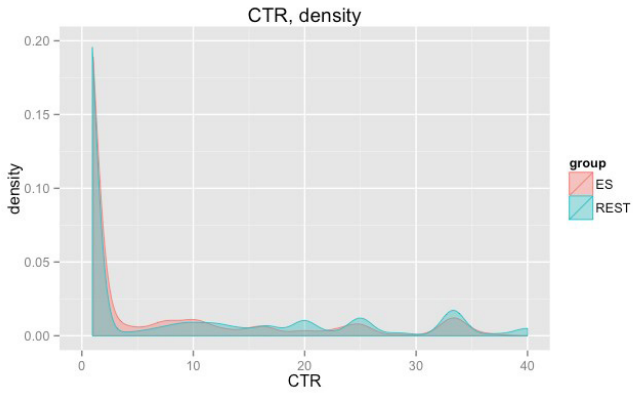
Ad performance is typically measured by impressions through the RPM metric[5]. Nevertheless, in the Red Eagle Origins video game, the ad network provider is Chartboost, which measures advertisers in terms of bid[6] per click or install[7] (instead of per impression). Chartboost RPM is a combination of an advertising campaign's Install Rate (IR)[8], Click-Through-Rate (CTR)[9], and bid. In this section, we include the CTR variable to measure ad performance.

Furthermore, as a publisher, the goal is to maximize earnings by showing as many ads as possible. Hence, the first variable analyzed is the number of ads shown. We could simply count the number of times any ad is shown to a player, but there is not always an ad available when the game is ready to show it. Nevertheless, there is a variable in the dataset that measures the number of times that an ad request is not fulfilled per player. Thus, we establish a new variable that adds up the number of ads shown plus the number of times that an ad is requested and not fulfilled. This variable represents the *ads chance* and has allowed us to make a comparison between both segments in terms of their ad potential revenue.

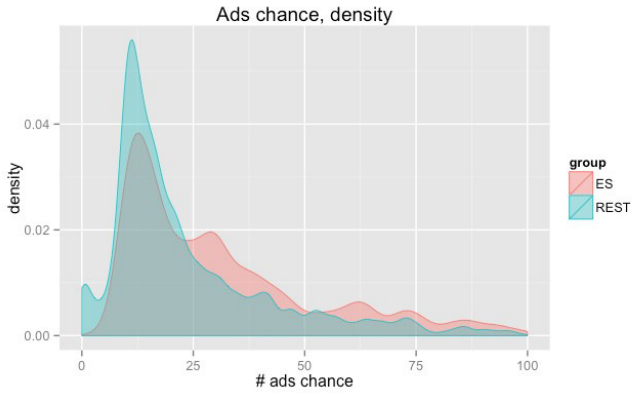
TABLE VIII. ADVERTISING METRICS

	Spanish players	Rest of the world players
CTR		
Mean	7.02	11.82
5% Trimmed Mean	4.04	7.95
Standard Deviation	17.57	23.67
Ads Chance		
Mean	65.41	34.09
5% Trimmed Mean	47.87	25.47
Standard Deviation	113.95	64.77

Table VIII shows the Ads Chance and CTR variables by set. We perform a U test to observe whether there is a significant difference between the two segments in regard to the Ads Chance variable, and the test confirms that the Spanish region shows a greater proportion of the ad impacts available, with a *p-value* < 2.2e-16. We also check how both segments behave in terms of CTR. We perform another U test, and with a *p-value* = 0.00081, the test confirms that contrary to the other variables measured, the rest of the World segment is more willing to click on the ads displayed. Fig. 5 graphically shows these results. This interesting finding may be motivated by several reasons such as the quality of the ad server or a different user behavior based on the user's citizenship.



a. CTR



b. Ads chance

Fig. 5. Densities of monetization variables for Spanish players (ES) and the rest of the world players (REST).

5 Earnings that accrue for every 1000 impressions received.

6 Advertisers' with higher bids take priority in showing their ads on the publishers' available inventory.

7 Mobile networks commonly bill advertisers on a CPI (Cost per Installation) basis.

8 The Install Rate of an ad campaign means how many clicks on an ad it takes to lead to an installation of an app.

9 The ratio of users who click on a specific link to the number of total users who view an ad.

VI. Social Network Impact

In this section, we have divided the data set into two subsets: players logged on Facebook and players non-logged on Facebook. Users' behavior for these two sets have been compared in terms of activation, retention and monetization.

A. Activation

The churn metrics for each subset are depicted in Table IX. The results show that users who are logged on Facebook are more likely to become active users. They have lower early drop metrics both in terms of number of games (63% versus 78%) and dropping out within the first 7 days (83% versus 88%).

TABLE IX. Churn Metrics (Early drop)

	Logged on Facebook	Not logged on Facebook	Total
Total # players	631	8687	9318
# players games < 10	396	6761	7157
Percentage	63%	78%	77%
# players drop out before R7	522	7645	8167
Percentage	83%	88%	88%

B. Engagement/Retention

Table X shows the results for both subsets in terms of: number of games played, maximum meters travelled, total meters per user and lifespan. For all variables included, the users logged on Facebook turn out to be higher-quality users because, on average, they play more games, have a higher maximum meter score, have more total meters, and have a longer lifespan. None of these variables follows a normal distribution; thus, non-parametric tests have been performed to check whether the differences are significant. The Mann-Whitney U test indicates that the differences in both sets are significant[10].

TABLE X. Retention Metrics

	Logged on Facebook	Not logged on Facebook
Total players	631	8687
Activated players (games ≥ 10)	235	1926
%	37%	22%
# games played		
Mean	23.19	11.51
5% Trimmed Mean	11.44	6.17
Standard Deviation	88.15	36.26
Max meters		
Mean	1418.7	961.1
5% Trimmed Mean	1265.03	883.64
Standard Deviation	1267.6	934.53
Total meters		
Mean	23392	8457
5% Trimmed Mean	8456.62	3510.62
Standard Deviation	134993.8	44499.74
Lifespan (days)		
Mean	6.28	3.89
5% Trimmed Mean	3.76	1.89
Standard Deviation	14.54	12.56

Fig. 6 depicts a graphical comparison of the normalized data by means of their density plots in terms of the number of games played (Fig. 6.a) maximum meters reached (Fig. 6.b), total meters travelled (Fig. 6.c) and lifespan (Fig. 6.d). These figures confirm that players logged on Facebook are likely to be more involved in the game, most likely because they are enabled to compete with their friends and to share and swap duplicate scrolls with each other.

As with our analysis of the TV series impact above, we compare the retention rate curves of Facebook vs. non-Facebook players, and we observe that there is a significant leap between both curves (Fig. 7).

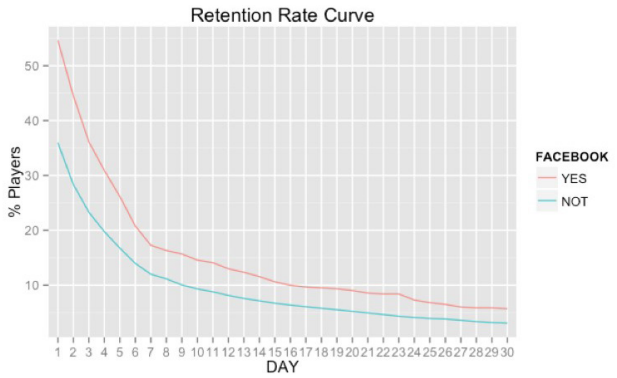


Fig. 7. Graphic comparison of the retention rate curves from players logged vs. not logged on Facebook.

Performing a Mann-Whitney U test, we verify that these differences are significant, with a *p-value* = 0.001377 that confirms a higher retention rate profile for the Facebook segment. If we take a more detailed look at the standard retention metrics, i.e., R1, R7 and R30, we observe that players logged on Facebook are much more willing to pass the day one retention threshold, giving the game a chance; in the R7 metric, there is also a good advantage for Facebook players, and the percentage of Facebook users who continue to play in the long-term retention metric, R30, is double that of those not logged on Facebook, as shown in Table XI.

TABLE XI. Retention Rate

	Logged on Facebook	Not logged on Facebook
R1	54.7%	35.9%
R7	17.3%	12%
R30	5.7%	3.1%

C. Ads Monetization

As with our above analysis of the players who have been exposed to the TV series, we can assess the ads monetization metrics for players logged on Facebook vs. those who are not. Doing so allows us to make a comparison between both segments in terms of their ad potential revenue. Table XII shows the outcome of the CTR and Ads Chance measured variables.

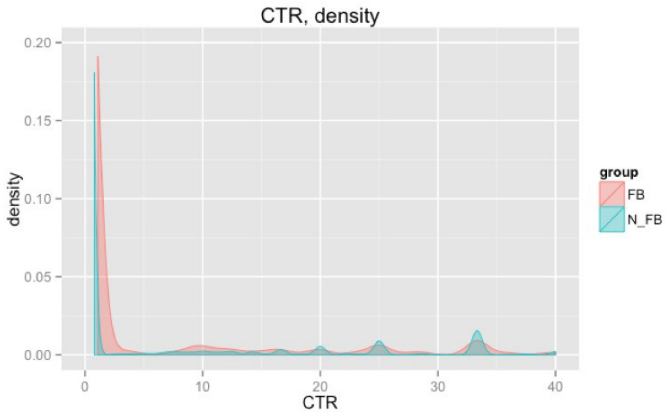
TABLE XII. Advertising Metrics

	Logged on Facebook	Not logged on Facebook
CTR		
Mean	7.97	6.17
5% Trimmed Mean	3.74	2.19
Standard Deviation	22.3	20.21
Ads Chance		
Mean	23.01	10.19
5% Trimmed Mean	10.62	5.24
Standard Deviation	94.21	33.94

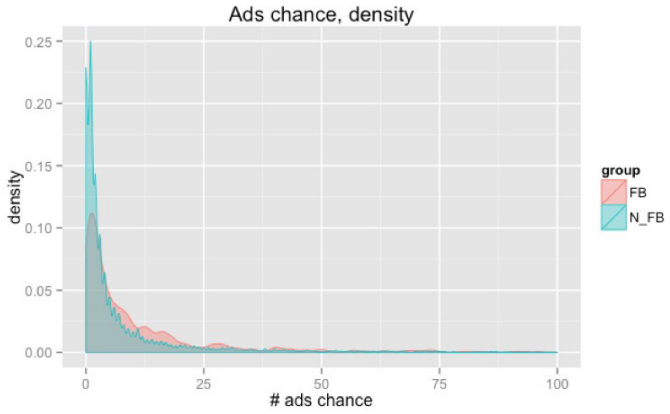
We perform a U test to observe whether there is a significant difference between the two segments in regard to the Ads Chance

10 The p-values obtained for all of these variables fall below 2.2e-16.

variable, and the test confirms that the Facebook segment shows a greater proportion of the available ad impacts, with a *p-value* < 2.2e-16. In terms of CTR, we perform another U test, with a *p-value* = 4.397e-05. The test confirms that the Facebook segment is more willing to click on the ads displayed. Fig. 8 graphically shows these results.



a. CTR



b. Ads chance

Fig. 8. Densities of monetization variables for users logged and not logged on Facebook.

VII. Business Impact

In this section, we discuss the results obtained from the business value point of view. We check to what extent the differences observed in the behavior of each segment of players imply differences in revenues. To do so, we have focused on assessing the income derived from in-game advertising during the time frame of the experiment. The variable selected to compare both segments is the Life Time Value (LTV), computed as the product of the Average Revenue Per Daily Active User (ARPDAU)[11] times the Lifespan.

Table XIII compares the TV series impact over these variables. We observe a great difference in between both segments: Spanish players show a LTV increase of 192% compared to the Rest of the world.

11 The ARPDAU is a standard metric in the industry and allows game designers and managers to follow the income evolution on a daily basis, frequently in order to monitor the changes applied on the fly to one or some of the game parameters [45]. It is calculated as: ARPDAU= RPM*ADAO/1000, where the average RPM for such period was $1.09 and the Average Daily Ads Opportunity (ADAO) per active user is calculated as ADAO = AdsOpp / Lifespan.

TABLE XIII. TV Series Business Impact

	Spanish players	Rest of the world players
Lifespan in days (mean)	18.15	9.19
Ads Opp (mean)	65.41	34.09
ADAO	3.604	3.709
ARPDAU	0.0039	0.0040
LTV	0.071	0.037

When the same calculations are performed to compare those players logged and not logged in Facebook, we get an even more significant increase of 226% on LTV of logged users, see Table XIV.

TABLE XIV. Social Media Business Impact

	Logged on Facebook	Not logged on Facebook
Lifespan in days (mean)	6.28	3.89
Ads Opp (mean)	23.01	10.19
ADAO	3.664	2.619
ARPDAU	0.0040	0.0028
LTV	0.025	0.011

We have seen in previous sections how being exposed to the TV series, or being logged on Facebook when playing the game, both have a significant positive impact on the retention metrics, increasing the users' lifespan. From the manager perspective, in the free-to-play business model the goal is to capture as much value as possible from users that have got the video game for free. So, maximizing the retention metrics could have a two fold benefit: on the one hand, the longer the players stay in the game, the higher opportunities they will have to bring their friends to share the same experience, reducing in turn the marketing User Acquisition Cost (UAC) through this "viralization" mechanism. On the other hand, we have seen that a longer lifespan translates into greater LTV derived from higher exposure times to in-game advertising, increasing final revenues. With the analysis done, we propose the following marketing strategies that could be implemented in free-to-play games:

- A video game company could be interested in achieving a licensing agreement with the owners of a TV series whose main audience overlaps with the video game target, as our experiment reveals that the LTV of players will get an increase of 192%.
- Use incentives to encourage players to log on Facebook or similar social network, as their LTV shall exceed 226% of not logged users. These incentives could be giving some virtual currency or digital goods for free.

VIII. Conclusions

In this study, we analyze the free-to-play video game industry from a novel perspective. We measure the impact of TV series and social networks on the main game metrics: activation, retention and monetization. To meet this challenge, a database of real users involved in the "Red Eagle Origins" game is used.

The first hypothesis to be tested regarding character identification involves checking whether being exposed to the TV series "Red Eagle" influences the initial interest that players show in the game (*activation)* and their *retention* and *monetization* metrics. The results indicate that Spanish players who have been exposed to the TV series are higher-quality users than players from the rest of the world who have not been exposed to the TV series.

The Spanish subset has a lower rate of "churn" or early dropout than the rest of the world. To some extent, this difference may be due to various cultural and socio-psychological factors regarding the affinity

with the historic theme and the game typology itself. However, again, the studies noted above suggest that being exposed or not to the TV series could be one of the determinant factors in the initial degree of interest shown by players. The results obtained confirm previous studies that suggest that being exposed or not to the TV series could be a determinant factor in the initial degree of interest shown by players.

The retention metrics show that the Spanish segment of players plays more games, stays in the game for longer and travels more meters than the rest of the world segment. These results seem to suggest that being exposed to the TV series is a key factor for retention because of the extra motivation that character identification provides to players.

Regarding monetization through in-game ads, the results show that as Spanish players have in general higher retention metrics, ads also have more chances to be displayed to them, generating more revenue, meaning that Spanish players have more *business value*. Although it has been verified that the rest of the world segment has a higher CTR, this difference in behavior may be due to various factors, such as the suitability of the displayed ads or the average willingness of players from different territories to click on any ad. Regardless of the cause, it should not depend on being exposed or not to the TV series, given that the ads offered are related to other games or apps, and in this case, CTR can be interpreted as more of a geographical or socio-cultural dependent variable.

The second hypothesis tested in the present work concerns whether integrating a social network into the game mechanics may have an impact on user behavior, improving the quality of players that make use of these social capabilities. In this analysis, the database has been divided into players logged and not logged on Facebook. The results reveal that there is a significant impact on all analyzed metrics.

Players logged on Facebook are more involved in the game, given that they can compete with their friends and share and swap duplicate scrolls with each other. These factors have a positive impact in terms of activation and retention because they present lower dropout metrics and, therefore, higher retention curves. These users play more games, travel more meters and play for longer periods. In terms of monetization, these users also have higher business value because they show higher CTR and Ads Chance values.

Given the characteristics of the free-to-play video game industry, it is crucial to identify the variables that have a significant impact on the business model. In this study, we show that players exposed to a TV series linked to the video game identify with the main characters, improving the users' quality. Furthermore, players logged on social networks such as Facebook experience higher values in terms of activation, retention and monetization, thereby substantiating that playing and sharing with friends reinforce engagement.

References

[1] D. Takahashi, "Candy Crush Saga maker King reports strong Q1 with $569M in adjusted revenue and 61 cents-a-share profit," *Venturebeat.com*, 2015.

[2] J. Diaz, "Supercell Makes $5 Million Per Day Off Clash Of Clans," *AndroidHeadlines*, 2015.

[3] K. Graft, "500 games launched per day on iOS last year (and other digital sales facts)," *Gamasutra*, 2015.

[4] D. Takahashi, "User acquisition costs go up, up, and away in August," *Venturebeat.com*, 2015.

[5] T. McCalmont, "How Do I Know I Have a Healthy Game?," *Gamasutra*, 2013.

[6] M. Callaghan, N. McShane, and A. G. Eguíluz, "Using game analytics to measure student engagement/retention for engineering education," in *Remote Engineering and Virtual Instrumentation (REV), 2014 11th International Conference on*, 2014, pp. 297–302.

[7] M. Minotti, "User acquisition costs up 26% in February after January's drop, Fiksu finds," *Venturebeat.com*, 2016.

[8] J. Runge, P. Gao, F. Garcin, and B. Faltings, "Churn prediction for high-value players in casual social games," in *2014 IEEE Conference on Computational Intelligence and Games*, 2014, pp. 1–8.

[9] C. Bauckhage, K. Kersting, R. Sifa, C. Thurau, A. Drachen, and A. Canossa, "How players lose interest in playing a game: An empirical study based on distributions of total playing times," in *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, 2012, pp. 139–146.

[10] M. Csikszentmihalyi, *Flow: the psychology of optimal experience*. Harper Perennial, 1990.

[11] E. L. Deci and R. M. Ryan, "Self-determination theory: A macrotheory of human motivation, development, and health.," *Can. Psychol. Can.*, vol. 49, no. 3, p. 182, 2008.

[12] A. K. Przybylski, C. Scott, and R. M. Ryan, "A motivational model of video game engagement," *Rev. Gen. Psychol.*, vol. 14, no. 2, pp. 154–166, 2010.

[13] R. J. Vallerand *et al.*, "Les passions de l'ame: on obsessive and harmonious passion.," *J. Pers. Soc. Psychol.*, vol. 85, no. 4, p. 756, 2003.

[14] C. K. J. Wang, A. Khoo, W. C. Liu, and S. Divaharan, "Passion and intrinsic motivation in digital gaming," *CyberPsychology Behav.*, vol. 11, no. 1, pp. 39–45, 2008.

[15] B. Bostan and S. Öğüt, "In pursuit of optimal gaming experience: challenges and difficulty levels," in *Communication présentée à l'Entertainment= Emotion Conference (Benasque: Centro de Ciencias de Benasque Pedro Pascual (CCBPP))*, 2009.

[16] B. Shneiderman, "Direct manipulation for comprehensible, predictable and controllable user interfaces," in *Proceedings of the 2nd international conference on Intelligent user interfaces*, 1997, pp. 33–39.

[17] H. L. O'Brien and E. G. Toms, "What is user engagement? A conceptual framework for defining user engagement with technology," *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 6, pp. 938–955, 2008.

[18] E. A. Boyle, T. M. Connolly, T. Hainey, and J. M. Boyle, "Engagement in digital entertainment games: A systematic review," *Comput. Human Behav.*, vol. 28, no. 3, pp. 771–780, 2012.

[19] J. L. Sherry, "Flow and Media Enjoyment," *Commun. Theory*, vol. 14, no. 4, pp. 328–347, 2004.

[20] P. Sweetser and P. Wyeth, "GameFlow: A Model for Evaluating Player Enjoyment in Games," *Comput. Entertain.*, vol. 3, no. 3, pp. 3–3, 2005.

[21] C. Jennett *et al.*, "Measuring and defining the experience of immersion in games," *Int. J. Hum. Comput. Stud.*, vol. 66, no. 9, pp. 641–661, 2008.

[22] J. L. Sherry, K. Lucas, B. S. Greenberg, and K. Lachlan, "Video game uses and gratifications as predictors of use and game preference," *Play. video games Motiv. responses, consequences*, vol. 24, pp. 213–224, 2006.

[23] M. Jennings, "Theory and models for creating engaging and immersive ecommerce websites," in *Proceedings of the 2000 ACM SIGCPR conference on Computer personnel research*, 2000, pp. 77–85.

[24] B. Laurel, *Computers as theatre*. Addison-Wesley, 2013.

[25] P. Vorderer, T. Hartmann, and C. Klimmt, "Explaining the Enjoyment of Playing Video Games: The Role of Competition," in *Proceedings of the Second International Conference on Entertainment Computing*, 2003, pp. 1–9.

[26] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing," *J. Exp. Soc. Psychol.*, vol. 45, no. 4, pp. 624–634, 2009.

[27] H. L. O'Brien and E. G. Toms, "The development and evaluation of a survey to measure user engagement," *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, no. 1, pp. 50–69, 2010.

[28] E. N. Wiebe, A. Lamb, M. Hardy, and D. Sharek, "Measuring engagement in video game-based environments: Investigation of the User Engagement Scale," *Comput. Human Behav.*, vol. 32, pp. 123–132, 2014.

[29] J. Kawale, A. Pal, and J. Srivastava, "Churn prediction in MMORPGs: A social influence based approach," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, 2009, vol. 4, pp. 423–428.

[30] E. G. Castro and M. S. G. Tsuzuki, "Churn Prediction in Online Games Using Players' Login Records: A Frequency Analysis Approach," *IEEE Trans. Comput. Intell. AI Games*, vol. 7, no. 3, pp. 255–265, 2015.

[31] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage, "Predicting player churn in the wild," in *2014 IEEE Conference on*

Computational Intelligence and Games, 2014, pp. 1–8.

[32] B. G. Weber, M. John, M. Mateas, and A. Jhala, "Modeling Player Retention in Madden NFL 11.," in *IAAI*, 2011.

[33] P. Cairns, A. L. Cox, M. Day, H. Martin, and T. Perryman, "Who but not where: The effect of social play on immersion in digital games," *Int. J. Hum. Comput. Stud.*, vol. 71, no. 11, pp. 1069–1077, 2013.

[34] R. (MUSE L. Bartle, "Hearts, Clubs, Diamonds, Spades: Players who suit MUDs," *J. MUD Res.*, vol. 6, no. 1, p. 39, 1996.

[35] N. Yee, "Motivations for Play in Online Games," *CyberPsychology Behav.*, vol. 9, no. 6, pp. 772–775, Dec. 2006.

[36] K. Christoph, H. Dorothée, and V. Peter, "The video game experience as 'true' identification: A theory of enjoyable alterations of players' self-perception," *Commun. Theory*, vol. 19, no. 4, pp. 351–373, Nov. 2009.

[37] D. Zillmann, "The psychology of suspense in dramatic exposition," *Suspens. Conceptualizations, Theor. Anal. Empir. Explor.*, pp. 199–231, 1996.

[38] M. C. Green, T. C. Brock, and G. F. Kaufman, "Understanding media enjoyment: The role of transportation into narrative worlds," *Commun. Theory*, vol. 14, no. 4, pp. 311–327, 2004.

[39] J. Cohen, "Defining Identification: A Theoretical Look at the Identification of Audiences with Media Characters," *Mass Commun. Soc.*, vol. 4, no. 3, pp. 245–264, 2001.

[40] D. Hefner, C. Klimmt, and P. Vorderer, "Identification with the Player Character as Determinant of Video Game Enjoyment," *Entertain. Comput. - ICEC 2007*, vol. 4740, no. March 2016, pp. 39–48, 2007.

[41] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore, "The Life and Death of Online Gaming Communities : A Look at Guilds in World of Warcraft," *Distribution*, vol. In Press, pp. 839–848, 2007.

[42] R. B. Rubin and M. P. McHugh, "Development of parasocial interaction relationships," *J. Broadcast. Electron. Media*, vol. 31, no. 3, pp. 279–292, 1987.

[43] R. Gómez, "Los programas más caros de TVE," *El País*, 2013. [Online]. Available: http://cultura.elpais.com/cultura/2013/04/12/television/1365796815_355916.html.

[44] M. Coates, "Kolmogorov-Smirnov test vs Mann Whitney test for unpaired hypothesis testing?," *Analyticbridge*, 2009. [Online]. Available: http://www.analyticbridge.com/forum/topics/kolmogorovsmirnov-test-vs-mann.

[45] N. Lovell, "ARPDAU," *Gamesbrief.com*, 2012. [Online]. Available: http://www.gamesbrief.com/2012/09/arpdau/.

Fernando De Rada

Fernando De Rada is the Advanced Interfaces Director at Singular, an international provider of technology solutions that helps companies harness and leverage today's most cutting edge digital technologies to create value and grow. He received the degree in Physical Sciences from the Autonomous University of Madrid in 2003, and got his postgraduate Master in Image, Publicity and Corporate Identity from the UCJC in 2014. Over 25 years of experience as entrepreneur and senior executive in mobile, gaming and digital media, since the late 1980s has founded and managed different companies leading multidisciplinary teams with a verifiable track record of remarkable results. He has been managing several national research projects, in the fields of Big Data, healthcare, mobile and visual technologies. He is also member of the Spanish Interactive Arts and Sciences Academy, and has been awarded in 2014 with the Prize for a Professional Career by Retro Madrid and AUIC.

Asuncion Mochon

Asuncion Mochon holds a degree in Economics by Universidad Autonoma de Madrid and a PhD degree from UNED, where she got a tenure and is currently associate professor at the Department of Applied Economics and Economic History. She teaches graduate and undergraduate courses and has high experience in e-learning. Her research focuses on economic human behaviour and optimal strategies.

Yago Saez

Yago Saez received the degree in computer engineering in 1999. He got his Ph.D. in Computer Science (Software Engineering) from the Universidad Politecnica de Madrid, Spain, in 2005. Since 2007 till 2015 he was vice-head of the Computer Science Department from the Carlos III University of Madrid, where he got a tenure and is nowadays associate professor. He belongs to the Evolutionary Computation, Neural Networks and Artificial Intelligence research group (EVANNAI) and member of the IEEE Computational Finance and Economics Technical committee.

Removing Unclassified Hand Tremor Motion from Computer Mouse Input with Neural Networks

Stephen Mack*

Electrical Engineer, Lawrenceville NJ (USA)

Received 17 May 2018 | Accepted 3 October 2018 | Published 19 October 2018

Abstract

An artificial neural network based filter to remove unwanted tremor-induced motion in computer mouse input is presented and tested. A method to efficiently capture appropriate training data is shown to be important in the operation and training of the neural network filter. The architecture of the neural network as well as the numerous design choices are presented and explained. A simulation study proves the artificial neural network is successful at removing a simulated Parkinson's tremor from computer mouse movements even with minimal training data. Resulting tremor-free motion estimated by the artificial neural network is shown to be similar to normal tremor free computer mouse movements.

Keywords

Filter, Human Computer Interface, Neural Network, Tremor, Mouse Cursor.

DOI: 10.9781/ijimai.2018.10.003

I. Introduction

USING a computer mouse accurately and quickly is a vital part of modern day life that many people take for granted. Conditions that affect muscular control or cause hand tremors can hinder a person's ability to use a standard computer mouse, limiting their ability to work or participate in aspects of modern life. Many government processes and forms are now only available online, making interfacing with a computer a daily necessity. It is estimated that 7 million Americans, or roughly 2.2% of the American population, has some form of an essential tremor (ET) [1]. These 7 million Americans have a wide range of condition symptoms and severities, which makes it difficult to provide universal assistance solutions.

Some people with tremors have found that using a track-ball style mouse works better than a standard mouse, but this costs money and the person still needs to learn how to effectively use it. Other researchers believe that an adjustable user interface can solve the problem [2], but while this is a solution it is not allowing the user to resume normal use of their computer. The optimal solution will be to allow the person to continue to use a standard computer mouse that they are used to, and remove the tremor-induced motion in software.

Much work has been done to determine how different conditions can affect the use of a computer mouse [3][4][5][6]. This research has improved our statistical understanding of how these tremors affect computer mouse movements and provide insight into the range of tremor severities for different conditions. Such statistical understanding of how these tremors affect computer mouse inputs has led to incredible advancements in diagnosing these conditions by simply analyzing a user's mouse movements [7][8], some have even attempted to diagnose these conditions by analyzing the users physical movements [9]. Diagnostic developments like these have the potential to lower the cost and complexity while increasing the speed of testing patients.

Rapid and accurate diagnostic achievements such as these are a large step forward in the field, but to date very little has been done to actively remove the unwanted tremor motion. One of the main problems is that there is such a wide range of types of tremors, all with varying levels of intensity and frequencies making it hard to account for them all. For example, Parkinson's tremors are known to exist from 3-7Hz, while orthostatic tremors range from 13-18Hz [10]. Characteristics of the tremor will also vary person to person depending on severity of the condition.

Traditional filtering techniques such as moving average filters, infinite impulse response (IIR) and finite impulse response (FIR) filters can easily attempt to remove the unwanted frequencies, but these filter designs are not easily adjustable or customizable to a specific patient's tremor and usually produce sub-optimal results. Some software packages that claim to be able to remove tremor-induced motion from a computer mouse exist, but all have a set number of filters for the user to choose from [11]. While these software packages may work well, they will never be truly tailored to an individual user, and a person with a tremor can only hope that one of these predetermined filters will fit their tremor well enough.

A truly universal software solution would need to accurately model a user's tremor, recognize the patterns in the tremor to remove them, and function smoothly and continuously in time. Artificial neural networks have been shown to be effective at prediction, modeling, pattern matching tasks [12], and have been useful in time-series tasks [13][14][15], making them a good choice for removing tremor-induced motion from standard computer mouse inputs.

This paper describes the design of a multilayer artificial neural network that is capable of being trained by the backpropagation algorithm to remove the unwanted mouse cursor movements caused by essential tremors in an individual patient or a generalized group of patients. A simulation study is offered that demonstrates the capability of removing the unwanted mouse cursor tremor from an individual user

* Corresponding author.
E-mail address: SteveCRM@gmail.com

with a simulated Parkinson's tremor. Results of the simulation study demonstrate the power and potential uses of this filtering technique.

The following sections present and verify this new method of removing unwanted tremor induced motion from a computer mouse based on artificial neural networks that can be customized for an individual patient.

II. Problem Statement

This paper presents an artificial neural network and corresponding training method to remove tremor-induced noise from the motion of a standard computer mouse. Section III presents the architecture of the neural network. Section IV describes how the neural network will be trained, and provide methods for collecting suitable training data. Section V details a simulation study to prove the effectiveness of the presented neural network solution. Finally, Section VI presents conclusions and avenues for future research.

III. Design of Neural Network

Artificial neural networks are an effective method at modeling nonlinear mathematical functions. The neural network presented here will attempt to accurately model the nonlinear relationship between computer mouse input with and without a tremor.

The output of the neural network will be an estimate of tremor-free mouse cursor motion, and it will make this estimate with a recent history of cursor motion with tremor. The artificial neural network takes in recent time samples in parallel, meaning each input neuron will receive a recent historical mouse sample. How many historical samples used as inputs can be varied to achieve the best results, but it is recommended that it should be at least a second or more to capture enough information about tremors of a wide range of frequencies and the desired motion of the mouse.

Because the neural network will be analyzing the motion of the cursor it is important that the historical samples used as inputs be cursor velocities instead of cursor positions. Velocity is a better input choice because we care about where the cursor is going and how quickly, and as the cursor rapidly changes direction the input will be normalized around 0 pixels per second. If position were to be used we would need training data for all areas of the screen evenly, and would need to retrain the entire network if the user changed screen sizes or resolutions. This amount of training data is technically feasible, but not realistic for the goal of making a simple and easy to configure universal filter to remove tremors from the computer mouse movements.

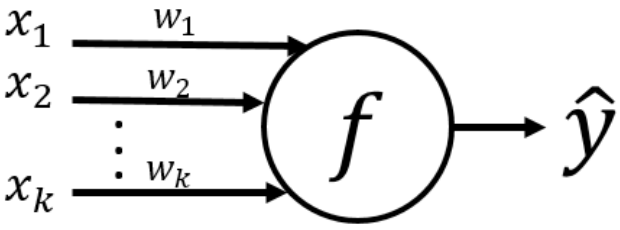


Fig. 1. Visual model of an artificial neuron.

The artificial neural network is made up of multiple neurons, usually in multiple layers. The output of an artificial neuron is defined by Eq. (1), and Fig. 1 provides a graphical.

$$\hat{y} = f(w^T x) \tag{1}$$

where f is the activation function, x is the input matrix, $\hat{y}$ is the output, and w is the matrix of weights associated with the inputs. The error of the neuron is given in Eq. (2).

$$e = y - \hat{y} \tag{2}$$

where y is the known training data.

Training the neural network means that all of the weights will have to be carefully chosen to approximate the appropriate nonlinear function to remove the tremor from the computer mouse input motion. The backpropagation algorithm is used to train artificial neural networks, using the gradient descent algorithm to determine and update the neural network weights. The weights in the artificial neural network are updated using Eq. (3).

$$w_k^{new} = w_k^{old} - n\frac{de}{dw} \tag{3}$$

where n is the step size, e is neuron error, and k is the neuron index.

Choice of activation function can greatly affect the performance of a neural network. If all neurons use a linear activation function then the entire network will only be able to approximate a linear function, no matter how large the network is. This means that it is vitally important that the neural network use nonlinear activation functions for the majority of the network. To ensure that the network is capable of approximating a nonlinear relationship the rest of the neurons in the network are given a nonlinear activation function, in this case the rectified linear function (ReLu) as shown in Eq. (4). The output of the neural network is an estimate of tremor-free cursor velocity, and because the velocity will theoretically have no upper or lower limit the single output neuron will use a linear activation function.

$$f(x) = \max(0, x) \tag{4}$$

If this neural network is to be designed to fit a single user then it is likely that there will be a realistic limit to the amount of training data available, a user will not sit for hours to provide training data. With small amounts of training data there is a very real possibility that the neural network can suffer from over-fitting where it will predict specific behaviors in the training data rather than the general trend in the training data. To combat over-fitting of the neural network and to ensure that the network output is not too sensitive to a small subset of neuron weights, a 20% dropout is used between each layer of the neural network.

The total number of neurons in the network will depend on the number of layers, and the number of neurons in each layer. There will be an input layer that will have as many neurons as there are historical input samples, and an output layer with a single neuron. All the layers in-between are referred to as hidden layers. The number of hidden layers and the number of neurons in each layer will affect how well the overall neural network can remove the tremor motion, but it can also affect the amount of computation time needed to make a prediction. If the neural network is to work in real time there are definite time constraints to creating a prediction. Future research and testing will be needed to optimize these values for a functional software prototype.

The neural network is estimating the velocity of the mouse cursor, meaning that to find the most recent cursor position we need to sum the velocity with the last known cursor position as shown in Eq. (5).

$$x[k] = x[k-1] + \hat{y}[k] \tag{5}$$

where x is the cursor position, k is the current sample, and $\hat{y}$ is the estimated cursor velocity. It is important to note that the cursor position is never updated with a true cursor position from the operating system, but only from an initial starting point and velocity estimates from the neural network. Similar to many dead reckoning algorithms, updating in this way means that any error in the neural network estimate will create a constantly growing error drift in cursor position. This drift

is likely to be small enough that the user will be able to naturally and unknowingly account for it with low frequency adjustments. This issue will need to be thoroughly tested in future research.

IV. Data Collection & Neural Network Training

The presented artificial neural network is to be trained with the backpropagation algorithm, which means that input data as well as classified output data is required to model the tremor filtering behavior. Input data will simply be a user's recent mouse velocities with tremor motion, and the output data will be the correct next cursor velocity if the user had no tremor.

The input data is simple to collect, but it is much more difficult to obtain acceptable output data for training. Traditional filtering methods could be applied to the input in an attempt to recreate a cursor velocity without a tremor but then the neural network would be approximating known methods, providing little additional value to the state of the art. The best obtainable data to represent tremor free mouse motion is the user's desired cursor velocity. Desired location is not a simple concept to measure, unless a desired location is provided for the user to track with their cursor.

To collect mouse movements with tremors as well as the desired movement, a software package was developed to record a single user's mouse cursor and tracking target location data over a short period of time. The software asked the user to track a target on the screen as best they can, trying to keep their cursor as close to the cross-hair as possible with their unique tremor.

The target moves around the screen in regular straight lines, only changing directions when it reaches the end of the screen where it appears to bounce off at a predictable angle. The ball also slowly changes velocity to ensure that data is collected over all ranges of normal cursor speeds. Changes in target position and velocity are meant to be extremely predictable to the user so that any cursor pointing error derives from the hand tremor, and not uncertainty in the target motion. To sufficiently train a neural network to remove tremor motion and give the user normal cursor motion we must collect data covering all situations that are considered normal cursor motion. Changing direction and speed of the target ensures that the neural network will be trained over all normal mouse cursor movement situations.

To assist the user in accurately tracking the target the software removes any distractions by taking up the whole screen with a single color. The target is made easy to visually track by making it bright red with a bold black crosshair in the center. To assist the user in predicting the motion of the target a green tail shows the targets most recent motion. With these aids we attempt to assure that any error in pointing is from the patient's tremor, and not any other external factor.

Cursor data is collected for as long as the user would like, though the first and last ten seconds of data are removed so that any cursor position error coming from the user purposefully leaving the target to hit the start or close buttons is ignored. Beyond this trimming of the recorded data the software in no way alters the cursor data, it simply records the cursor location and the target location.

The resulting data set is a time stamped list of the location of the two dimensional mouse cursor location as well as the desired target location. Similar data sets are collected by researchers attempting to diagnose specific tremor conditions, but in this case the data will be used to compensate for the existing tremor. A screenshot of this software collecting data can be seen in Fig. 2.

This method can be used on a single patient to personally customize their software, or could be used for large scale data collection programs to create more universal neural network filters. Ideally the neural network will be trained with as much data as is pragmatic, as training with too little data can lead to poor performance.

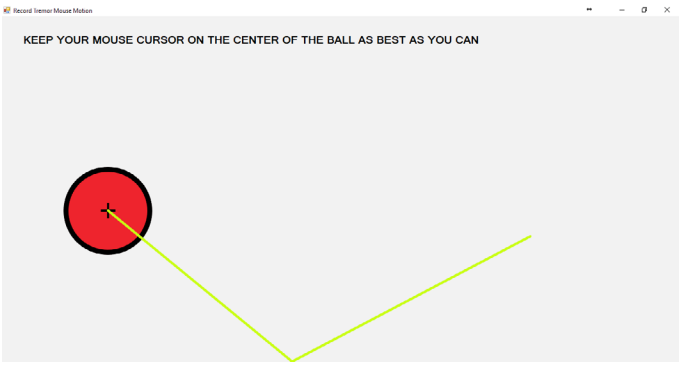


Fig. 2. Screenshot of software recording mouse data with moving target as desired location.

Collecting this data is not a computationally intensive process, though training a neural network can easily become an unreasonable task for a low-end home computer. The difference in computational abilities means that for a final product it would be beneficial to collect the training data on a user's machine, but offload the neural network training to a dedicated server off-site. For such a setup to work the training data would need to be collected at the user's computer, then uploaded to a server where the neural network would be trained, and finally the server would send back a trained neural network to the user when complete. Currently the software only records training data to a file, but can be turned into a service by sending the results over the internet to a waiting server with neural network training software.

V. Simulation Study

A simulation was created to demonstrate the ability of the artificial neural network to remove a simulated Parkinson's tremor from computer mouse movement. For simplification and presentation purposes this simulation study focuses on removing tremor only from the horizontal axis of the mouse cursor data, though it would not be difficult to adjust this study to accommodate the two axes of motion.

This study aims to remove a simulated a 6Hz Parkinson's tremor from a test patient's mouse cursor movements. Data used for this simulation was collected using the software described in section IV. The previously presented data collection software records cursor positions every 15.5ms, or a sample rate of 64.5Hz, and converts them to cursor velocities to be saved to a file. This sampling rate is sufficiently greater than the Nyquist sampling rate of the 6Hz Parkinson's tremor. Two sets of data must be collected, one to train the neural network and a second to validate the trained neural network.

The first data set will use the cursor velocities with tremor as well as the desired velocities to train the neural network. Once the network is trained it will use the second data set to validate its predictions of tremor free motion. It is important to use two different data sets for training and validation so that we can be sure that the neural network is learning the underlying dynamics of removing tremor motion and not simply memorizing and regurgitating the data we gave it during training like we would see in severely over-fit models.

Training data was collected for 32 seconds producing 2089 samples, and validation data was collected for 26 seconds producing 1661 samples.

A portion of the validation data set showing the cursor position error is shown in Fig. 3.

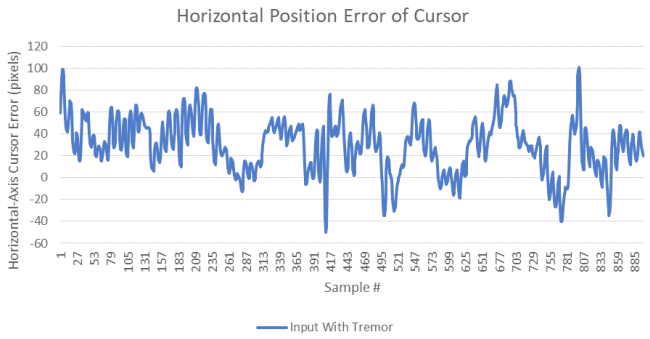


Fig. 3. Validation data set horizontal position error of cursor.

Cursor position error is centered near zero though there is obviously a high frequency component to the motion, unlike what would be seen in a user without a tremor. There are also long periods where the error is obviously purely positive or negative, meaning the user had difficulty tracking exactly on the target.

The artificial neural network in this simulation is comprised of 5 layers: 20 input nodes taking in the 20 most recent cursor velocity samples, 25 nodes in layer 2, 16 nodes in layer 3, 16 nodes in layer 4, and a single output node. A 20% dropout factor is added between each layer to prevent over-fitting. The output node uses a linear activation function, all other neurons use the rectified linear activation function (ReLu).

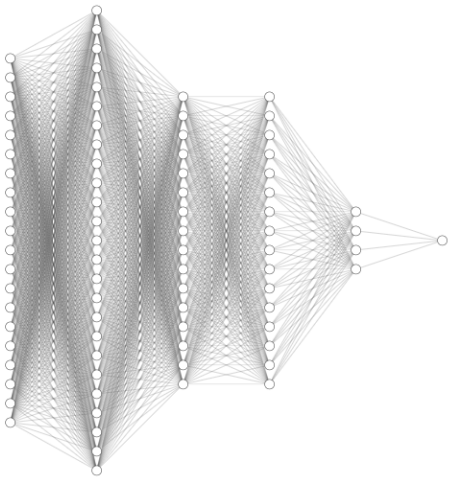


Fig. 4. Architecture of the simulated neural network.

Fig. 4 depicts a visual representation of the proposed neural network. Each circle represents a neuron, and each line represents an interconnection with an associated weight. The input layer is represented on the left, with the single output neuron on the right.

The artificial neural network is then trained with the training data set using python and keras software packages. Once the neural network was sufficiently trained, it was used to predict the tremor-free cursor velocities for the validation data set. For visualization purposes the velocities have been converted back into position errors measured in pixels. Fig. 5 shows the same horizontal cursor position error from the validation data set as in Fig. 3, but also includes the error of the validation data set after the neural network attempted to remove the tremor.

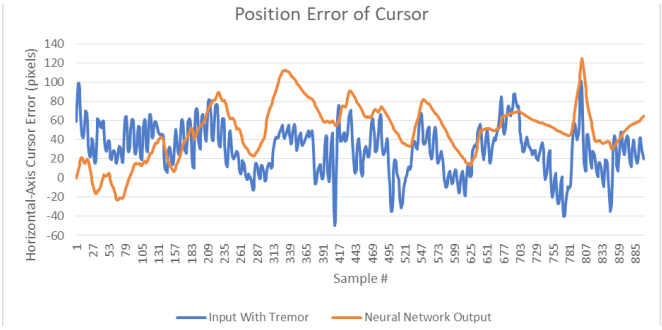


Fig. 5. Horizontal cursor position error before and after neural network improvement.

As shown in Fig. 5, the neural network successfully removed much of the higher frequency cursor motion, creating a much smoother looking position error graph. While the high frequency components of the error have been removed there does appear to be some offset error as predicted in section III. The offset error is relatively small and low frequency as expected, meaning the user could likely compensate for the effect without realizing.

While visually it appears that the higher frequency components of the cursor motion have been removed we still need to verify that the simulated 6Hz Parkinson's tremor has been sufficiently suppressed. To quantify the amount of tremor that was inhibited by the neural network a Fourier analysis is performed on the input with tremor as well as the neural network output without tremor. The Fourier analysis was done on using cursor position data rather than cursor velocity for visualization purposes. As shown in Fig. 6, the validation set has a tremor between 5Hz and 6Hz, consistent with tremors in patients with Parkinson's disease. The output of the trained neural network showed almost no motion at 6Hz, meaning the artificial neural network successfully filtered the tremor induced motion.

It is also interesting to note that the spectral intensity at 4 Hz and 8 Hz is roughly the same for each plot, meaning the neural network isn't simply acting as a low pass filter, but has learned to target the tremor specifically around 6 Hz.

Fig. 6 shows that the simulated Parkinson's tremor was removed but does not prove that the resulting mouse movement is similar to natural tremor-free computer mouse motions. To prove the neural network output is similar to natural mouse movements Fourier analysis is used to compare the output of the trained neural network to the desired cursor position. Fig. 7 demonstrates that the neural network produces similar frequency components to the tracking target motion, meaning the mouse should act like a regular mouse cursor to the user.

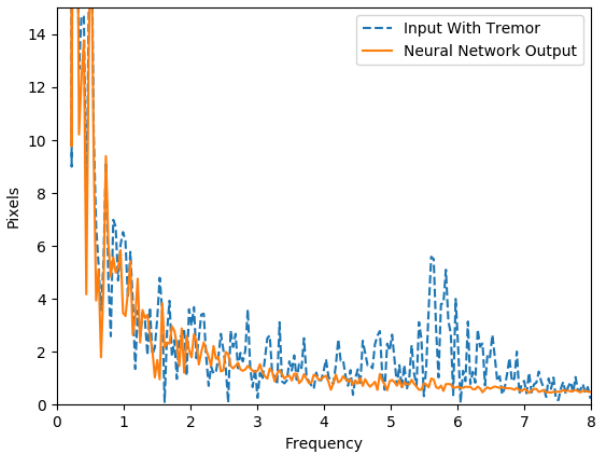


Fig. 6. Fourier analysis – removing tremor.

Key to note is that above 2 Hz the desired target motion and the neural network estimated motion are nearly identical. Traditional filtering methods would likely have acted as band-stop filters, suppressed the entire region where the tremor is present, but instead of simply removing spectral information the neural network actually returns the tremor motion back to a normal non-tremor state.

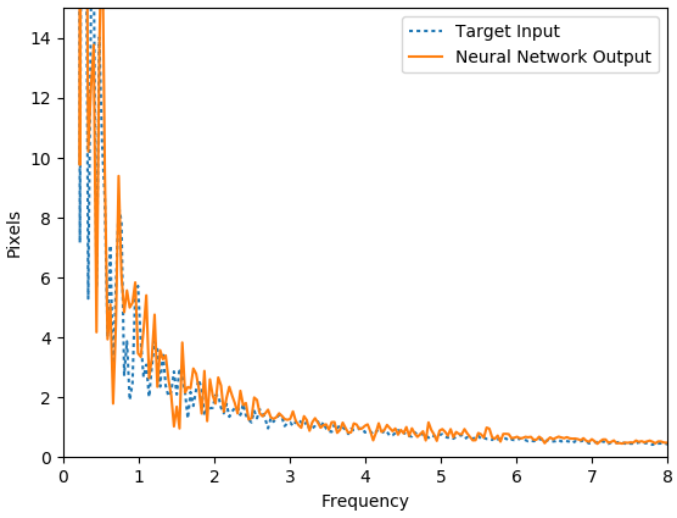


Fig. 7. Fourier analysis – performing like standard computer mouse.

The simulation study demonstrates that the proposed neural network design can effectively remove simulated Parkinson's tremor motion in one axis of a computer mouse cursor, and that the resulting tremor-free motion is similar to normal computer mouse movements.

VI. Conclusions & Future Work

The concept of using a neural network to remove tremors from computer mouse input has been shown to be extremely effective even with minimal training data from a single user. Not only did the neural network remove the tremor but the resulting mouse motion closely mimicked normal mouse movements, meaning a user will hopefully experience no discomfort or clumsiness while using a computer.

Additional research must be conducted to find an optimal size and structure of the neural network. A neural network larger than necessary will be a waste of the user's computer resources, and runs the risk of not being able to keep up with the speed at which cursor position samples are arriving. Alternative number of hidden layers, number of neurons per layer, and choice of activation function could also improve the functionality of the design.

In addition to the success of the neural network, the data collection software has also proven to be effective at collecting mouse movements with tremor while also supplying an accurate representation of the user's desired cursor motion. The software can be easily connected to the internet and scaled up for research into more universal filters, or for implementation into a consumer product.

The ability to design an easily customizable and situation specific filter without fully characterizing the noise opens this work up many possible applications beyond removing tremors from computer mice. New research avenues into neural network based noise cancellation can be applied to fields in signal processing, control theory, and many others.

Work on this topic will continue with the creation of a network capable of handling horizontal and vertical mouse movements, as well as creating a real time functioning prototype. With a functioning prototype experimentation on developing a more universal filter can be explored. A prototype will also allow patients with confirmed and medically diagnosed tremors to validate the concept and provide valuable feedback on functionality.

References

[1] Louis, Elan D., and Ruth Ottman. "How Many People in the USA Have Essential Tremor? Deriving a Population Estimate Based on Epidemiological Data." *Tremor and Other Hyperkinetic Movements*, 2014, Vol. 4, pp. 259.

[2] Aqueasha Martin-Hammond, Foad Hamidi, Tejas Bhalerao, Christian Ortega, Abdullah Ali, Catherine Hornback, Casey Means, and Amy Hurst, "Designing an Adaptive Web Navigation Interface for Users with Variable Pointing Performance". *Proceedings of the Internet of Accessible Things 2018*, pp. 31.

[3] Simeon Keates and Shari Trewin. "Effect of age and Parkinson's disease on cursor positioning using a mouse". *In Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility,* 2005, pp. 68-75.

[4] Alvaro D. Taveira, Sang D. Choi, "Review Study of Computer Input Devices and Older Users", *International Journal of Human–Computer Interaction*, 2009, Vol. 25, no. 5, pp. 455-474.

[5] R. C. Eberhart and Xiaohui Hu, "Human tremor analysis using particle swarm optimization," *Proceedings of the 1999 Congress on Evolutionary Computation,* 1999, Vol. 3, pp. 1927-1930.

[6] Cohen O, Pullman S, Jurewicz E, Watner D, Louis ED. "Rest Tremor in Patients With Essential Tremor" *Archives of Neurology*. 2003, Vol. 60, no. 3, pp. 405–410.

[7] N. M. Aly, J. R. Playfer, S. L. Smith, D. M. Halliday; "A novel computer-based technique for the assessment of tremor in Parkinson's disease", *Age and Ageing*, 2007, Vol. 36, no. 4, pp. 395–399.

[8] Hirschauer, T.J., Adeli, H. & Buford, J.A. "Computer-Aided Diagnosis of a Parkinson's Disease Using Enhanced Probabilistic Neural Network" *Journal of Medical Systems*, 2015, Vol. 39, no. 11, pp. 179.

[9] Enas Abdulhay, N. Arunkumar, Kumaravelu Narasimhan, Elamaran Vellaiappan, V. Venkatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease", *Future Generation Computer Systems*, 2018, Vol. 83, pp. 366-373.

[10] Seyed Yashar Bani hashem MEng, Nor Azan Mat Zin PhD, Noor Faezah Mohd Yatim PhD & Norlinah Mohamed Ibrahim MD, "Improving Mouse Controlling and Movement for People with Parkinson's Disease and Involuntary Tremor Using Adaptive Path Smoothing Technique via B-Spline", *Assistive Technology: The Official Journal of RESNA*, 2014, Vol. 26, no. 2, pp. 96-104,

[11] Lee, Hong Ji et al. "Tremor frequency characteristics in Parkinson's disease under resting-state and stress-state conditions", *Journal of the Neurological Sciences*, 2016, pp. 272 - 277.

[12] B. Widrow and R. *W*inter, "Neural nets for adaptive filtering and adaptive pattern recognition," *in Computer*, 1988, vol. 21, no. 3, pp. 25-39.

[13] Iebeling Kaastra, Milton Boyd "Designing a neural network for forecasting financial and economic time series", *Neurocomputing*, 1996, Vol. 10, no. 3, pp. 215-236.

[14] Georg Dorffner, "Neural Network for Time Series Processing", *Neural Network World*, 1996, Vol. 6, pp. 447-468.

[15] Abbasi, R., M. Esmaeilpour, "Selecting Statistical Characteristics of Brain Signals to Detect Epileptic Seizures using Discrete Wavelet Transform and Perceptron Neural Network", *International Journal of Interactive Multimedia and Artificial Intelligence*, 2017, Vol. 4, no. 5, pp. 33-38.

Stephen Mack

Stephen Mack earned his BS in Electrical Engineering from The College of New Jersey and his MS in Electrical Engineering from the University of Virginia. His research interests include control theory, RF/communications, space systems, and machine learning.

Exploratory Boosted Feature Selection and Neural Network Framework for Depression Classification

Vanishri Arun[1]*, Murali Krishna[2], Arunkumar B.V.[3], Padma S.K.[1], Shyam V.[4]

[1] Department of Information Science and Engineering, SJCE, JSS S&T University, Mysuru (India)
[2] Department of Psychiatry, CSI Holdsworth Memorial Hospital, Mysuru (India)
[3] Department of Anaesthesiology, Apollo BGS Hospital, Mysuru (India)
[5] Director and President, Forus Healthcare Pvt. Ltd., Bengaluru (India)

Received 11 July 2018 | Accepted 19 September 2018 | Published 5 October 2018

unir
LA UNIVERSIDAD EN INTERNET

Abstract

Depression is a burdensome psychiatric disease common in low and middle income countries causing disability, morbidity and mortality in late life. In this study, we demonstrate a novel approach for detection of depression using clinical data obtained from the on-going Mysore Studies of Natal effects on Ageing and Health (MYNAH), in South India where the members have undergone a comprehensive assessment for cognitive function, mental health and cardiometabolic disorders. The proposed model is developed using machine learning approach for classification of depression using Meta-Cognitive Neural Network (McNN) classifier with Projection-based learning (PBL) to address the self-regulating principles like how, what and when to learn. XGBoost is used for feature selection on the available data of assessments with improved confidence. To improve the efficiency of McNN-PBL classifier the best parameters are found using Particle Swarm Optimization (PSO) algorithm. The results indicate that the McNN-PBL classifier selects appropriate records to learn and remove repetitive records which improve the generalization performance. The study helps the clinician to identify the best parameters to analyze the patient.

Keywords

XGBoost, Meta-Cognitive Neural Network, Projection-based Learning, Particle Swarm Optimization, Depression, MYNAH Cohort.

DOI: 10.9781/ijimai.2018.10.001

I. Introduction

DEPRESSION is a non-communicable psychiatric disorder which is burdensome to people of all ages. According to psychiatrists depression is characterized by insomnia, tearfulness, anhedonia, suicidality, anorexia, self-depreciation, fatigue and lack of concentration. These characteristics will impede a person's capability to perform her or his daily activities normally. Due to lack of culturally adapted and validated assessments in low and middle income countries, assessment of the prevalence of depression becomes difficult. Psychiatrists diagnose depression using various parameters based on Diagnostic and Statistical Manual of Mental Disorders (DSM) [1]. Early diagnosis of depression helps in prevention of onset of depression related mental disorders like dementia, memory deterioration, Alzheimer's, etc., by treating at right place and time [2].

The risk of depression in elderly who have crossed the age of 60 years, which is called geriatric depression, increases due to factors like dependency on children for their living, physical disability, strokes, hypertension, diabetes, obesity, cancer, chronic pain, and also due to certain medicines. Depression affects an individual's personal and professional life that leads to decrease in productivity, increase in healthcare cost and exclusion from family and friends. Due to poor resource setting, lack of skilled people and increase in population depression is grossly undertreated. There is a very little data available or recorded on geriatric depression and very less research has been done on identification of depression using machine learning approaches. Some of the researchers have worked on small datasets using tools like Weka, SPSS, etc. The reason behind this area not being well explored is that the datasets are not properly maintained by the hospitals or the clinicians. Most of the datasets available are incomplete with either missing parameters or missing values. Therefore MYNAH cohort is used in our study which has good dimensions of 1201 features and 1321 patient records with phenotypic data. The values in the cohort are recorded after a comprehensive assessment of the patients for various mental and physical disorders.

Machine learning techniques help in detection of disorders faster with good accuracy and reduced misclassification rate. Knowledge acquisition for identification of depression from dataset is done using machine learning approaches. A novel model is presented in this paper for depression classification using XGBoost technique for feature selection and McNN-PBL for classification by finding the best parameters for McNN using PSO to improve the efficiency of the classifier. Using the XGBoost technique on data for feature selection makes the classifier learning and training faster, also reduces the misclassification error rate and improves the overall accuracy of the model. The algorithm is scalable and can handle large datasets. Therefore XGBoost technique was chosen for feature selection. Depending on each feature of the patient record in the cohort, McNN decides which record it should choose to learn or the record to be deleted so that redundancy is avoided addressing what-to-learn. When a patient record is input to the model, it is not always necessary for the model to learn immediately. It may be reserved for future learning

* Corresponding author.
E-mail address: vanishriarun@gmail.com

which addresses when-to-learn. The records may also be used for either addition of a neuron or updating of weights of the output which addresses how-to-learn component of meta-cognition. Initially, McNN starts with zero hidden neurons. It then adds sufficient neurons so that the decision surface is approximated. The principle adopted by the projection based learning algorithm is minimization of energy function. It finds the network output parameters which have a minimum energy function. The best parameters to improve the efficiency of McNN-PBL classifiers are found using PSO algorithm. McNN classifier discards repeated records from the training input data records. This reduces the memory requirement, also minimizes the computation time of the model and avoids over training. Another advantage of McNN is that it helps to reduce misclassification error [3].

In the past, researchers have worked on feature selection and disease classification using various approaches. To develop a scalable and efficient classification model, XGBoost has been incorporated with gradient boosting [4]. Boosting technique has been widely used across different domains for both classification and regression. Prediction of bioactive molecule facilitates the computer-aided drug discovery and XGBoost has shown a good performance on various datasets used [5]. XGboost has been used for the classification of DDoS attack which has shown significant performance compared to SVM and Random Forest [6]. Diabetes detection on a larger dataset using boosting approach has shown to be easily scalable [7]. Various Machine learning approaches are used to identify patients with depression. Depressed patients have been identified through classification of patients into subtypes based on the syndromes using the Beck Depression Inventory (BDI) item scores implementing categorization algorithm [8]. Depression diagnosis and feature reduction have been done using Support Vector Machine (SVM) with Voxel based morphometry and Filter method using ANOVA [9]. SVM which is a linear kernel with Principal Component Analysis (PCA) was used with image modality during depression related functional MRI tasks [10]. PCA was also used to minimize the number of attributes in the dataset and an ensemble classification framework has been used with Hierarchical Majority Voting (HMV) for disease classification and prediction [11]. Robust Spatial Kernel FCM (RSKFCM) segmentation method to detect Diabetic Retinopathy (DR) has been used for optic disk elimination in the retinal image and McNN approach has been used for DR classification [12]. Relevance Vector Regression with filtering out voxels from brain regions has been used to evaluate the BDI and HRSD scores. Non-linear Gaussian Kernels like Relevance Vector Machines and Support vector machines have also been used [13]. Effective Machine learning techniques are applied for detection of depression which has yielded an average accuracy of 80% and the size of the dataset used for the study is small which ranges between 18 and 62 patient records using mainly neuro-imaging. Image analysis uses a wide range of data for the assessment of depression. The previous studies on classification and prediction using machine learning approaches are done on various areas. But no study has yet been conducted on mental health cohort.

Hence we propose a machine learning model using a novel approach to classify depressed patients in the well populated cohort with good number of parameters using exploratory feature selection using boosting technique and McNN-PBL for classification.

The organization of the paper is as follows: Section II describes the Exploratory Feature selection using XGBoost technique. Section III presents the proposed methodology for depression detection using boosting technique for feature selection and McNN-PBL approach for depression classification. Section IV evaluates the performance of the proposed methodology and section V summarizes the conclusion and the scope for future study.

II. Exploratory Feature selection Using Boosting technique

A. Exploratory Feature Selection

Feature selection is a method used to choose a small set of parameters among a large parameter set. In this study, we have employed exploratory feature selection since the number of significant parameters required for identification of depression was unknown. The psychiatrist identified 45 out of 1201 parameters as significant parameters for detection of depression in the cohort based on his expertise. In exploratory feature analysis on SPSS statistical tool, Principal Component Analysis (PCA) method was used to reduce the complexity in the dataset and the Kaiser rule which is based on distribution theory of Eigen values was used as stopping criterion [14]. PCA involves a mathematical concept to transform a large number of correlated parameters into a small set of uncorrelated parameters called principal components [15]. These principal components serve as the significant features which act as predictors for depression detection. Eigen value is computed by examining the relationship between the parameters. The feature with Eigen value greater than 1 is chosen as the significant feature.

Using SPSS tool, 13 features were identified as significant features for detection of depression as shown in the Scree plot in Fig 1.

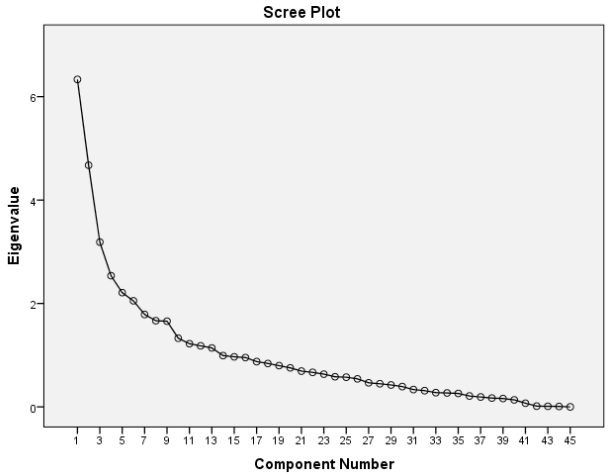


Fig. 1. Scree plot for Factor Analysis using PCA.

In our study, we used boosting technique to select the significant parameters by comparing each parameter applying decision rules to a decision tree model. The parameters are represented as internal nodes of the tree and the branching or the decision path is judged based on the parameters of the node. To get a single outcome at every leaf, a tree is created for the entire data. The model consists of a group of trees called an ensemble. In this ensemble, each tree is a decision tree called a weak learner or shallow tree based on the depth. XGBoost stands for Extreme Gradient Boosting and is an implementation of gradient boosted decision trees. XGBoost learns the data using the ensemble of boosted trees. It also handles the trade-off between the complexity of the model which is caused due to factors like number of trees, depth, etc. XGBoost technique for feature selection makes the learning and training faster. It also reduces the error rate during classification that in turn improves the overall accuracy of the model. Using boosting technique the significant factors chosen by the model were 10 out of 45 parameters. The 10 features chosen using XGBoost were given as input to the McNN-PBL model for depression classification.

B. Depression Detection Using the Features Selected from the Cohort Using XGBoost and McNN-PBL Classifier Model

The McNN is trained with the input and the desired output after which the training occurs. Weights at both the hidden and output layers are adjusted so that the actual output corresponds to the desired output. Once trained, the neural network takes a new patient record and gives either 0 or 1 as output where 0 denotes absence of depression and 1 denotes presence of depression. If the classifier is uncertain it will produce a value somewhere in between 0 and 1.

When a patient record is input to the model, the learning process in McNN addresses the self-regulating principles like what, when and how to learn in a manner similar to the cognition principle of a human brain. The learning process in McNN uses estimated class label, maximum hinge error and class wise significance. McNN chooses the patient record from the cohort to be used for learning to address what-to-learn principle depending on the parameter of the record. It also chooses the patient record to be deleted to avoid redundancy. The model does not have to start the learning process as soon as the patient record is presented to it. It may reserve the record for future learning addressing when-to-learn principle. To address how-to-learn component, the input patient records may be used for either addition of a neuron or updating of weights of the output. To start with, McNN will have zero hidden neurons. It then adds sufficient neurons to approximate the decision surface. Projection Based Algorithm adopts the principle of minimization of energy function in which it finds the network output parameters having a minimum energy function. To find the best parameters to improve the efficiency of McNN-PBL classifiers, we used PSO algorithm. Another advantage of McNN classifier is that it discards the same records from the training input data records which in turn reduces the memory requirement and computation time of the model. It avoids over training and helps in reduction of misclassification error rate.

III. Depression Detection Using Boosted Feature Selection and McNN-PBL Approach - Proposed Methodology

The detection of depression using MYNAH cohort is done in two phases. The first phase consists of feature selection using XGBoost technique. The second phase involves detection of Depression using McNN-PBL.

Fig. 2 represents the overall proposed model for the detection of depression using MYNAH cohort.

The model developed can be used as a classifier with the following functionality:

A. Data Acquisition and Assessment of Patients

1) The MYNAH Cohort

MYNAH is an abbreviation of MYsore studies of Natal effect on Ageing and Health, a cohort consisting of a total of 3,427 men and women born between 1934 and 1966 as singletons at Holdsworth Memorial Hospital, Mysore, Southern India. These people were traced by matching their birth records through a house-to house survey of the area of Mysore city surrounding the hospital between the years 1993 and 2001. A comprehensive examination of lung function, relationship between size at birth, adult cardio-metabolic disorders, etc., was conducted in which a total of 1069 people participated and this constituted the Mysore Birth Records Cohort. This cohort study was one of the first of its kind in low and middle income countries aiming to test developmental origins of health and disease (DOHaD) concepts with the predictions of associations between small size at birth and adult heart disease, resistance to insulin and reduced lung function and so on [16]. The surviving members in the cohort are aged over 60 years and the cohort serves as a unique resource for studies in epidemiology on old age.

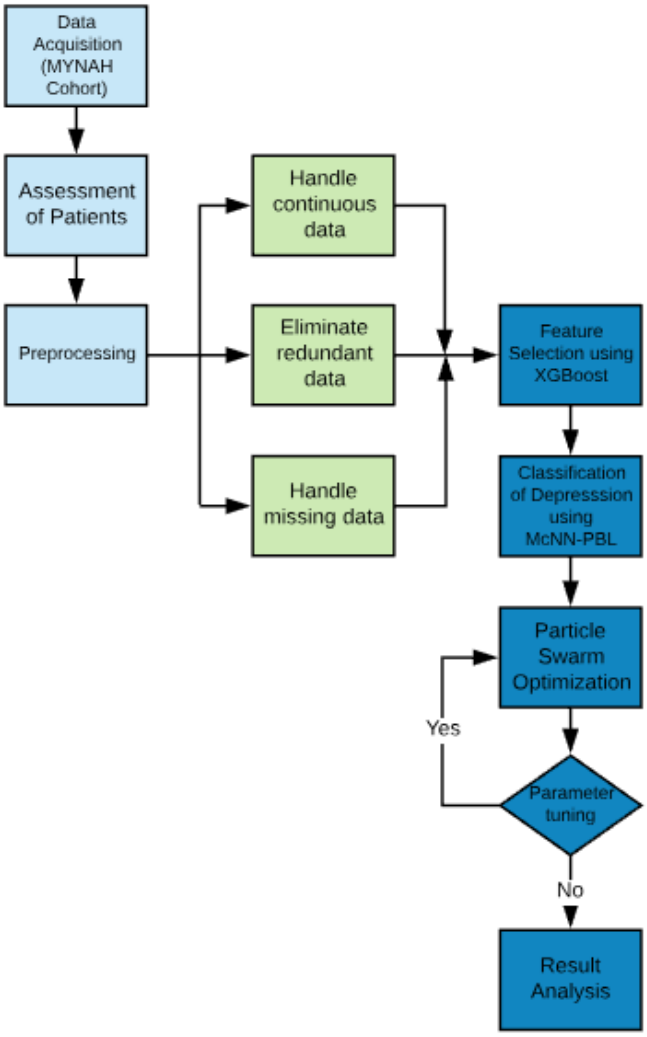


Fig. 2. Depression Classification Model.

2) Assessment

Examination was conducted on 721 surviving members in the cohort between the years 2015 and 2017 for mental and cardio-metabolic disorders. Based on the symptoms depression was assessed using Geriatric Mental State (GMS) Examination which is internationally used for assessment of geriatric mental health [17]. A computerized diagnostic algorithm called Automated Geriatric Examination for Computer Assisted Taxonomy (AGECAT) is used to group the symptoms to form patterns recognized by a clinician as illness or syndrome class [18]. International Classification of Diseases (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria are used to add the items together and to generate affective disorder diagnoses. In-patient, out-patient and community samples are used to demonstrate the reliability and validity of the GMS. Validity of GMS/AGECAT algorithm has been investigated in several studies [19]. A Socio-demographic questionnaire was developed to collect information on age group, gender, Body mass index (BMI) group, level of education, Standard of living index based on standard indicators, Job, Job category, and others.

B. Preprocessing

In preprocessing stage, three issues relative to the cohort were addressed. Firstly, dealing with missing data, secondly, handling

the redundant features and finally, mapping of continuous data to categorical data or encoding of categorical features. The cohort consists of 1201 parameters recorded during the assessment. Based on feature analysis done by the psychiatrist, 45 parameters were chosen for the study. In some of the observations the assessment for the required parameters was missing and some observations were redundant. Such records were dropped which resulted in 270 patient records out of the total 1321. Continuous data features were encoded into categorical data like age group (0-below 65 years, 1- above 65years), gender (0-male, 1-female), Body mass index (BMI) group (0-Under weight, 1-Healthy, 2-Over weight, 3-Obese), level of education (0-Illiterate, 1-Secondary, 3-College, 4-Diploma, 5-Graduate, 6- Post graduate), and so on.

To avoid dominance of some parameters it is essential to normalize input features for the objective function to perform well and for the algorithm to converge during optimization. Euclidean norm was used to rescale the data into the range [0,1] by dividing each feature vector by the Euclidean length of the vector as shown in Eq. (1).

$$x' = \frac{x}{\|x\|} \tag{1}$$

In Eq. (1), x' is the normalized value, x is the original feature value and $\|x\|$ is the Euclidean length of the vector.

Stratified Sampling method has been used in this study to handle the class imbalance by dividing the dataset into training and testing datasets in the ratio 75:25. The sampling function makes a split so that the proportion of values in the sample will be same as the proportion of values provided to parameter. It helps in dividing the training data into homogeneous groups and the ratio of samples from all the classes will be proportional. This will help in avoiding over fitting and bias towards the dominant class. Considering the stratified samples for training will help the algorithm to learn in an optimized way and it saves from further tuning or regularization.

C. Feature Selection Using XGBoost Technique

The cohort consists of a large number of features. Hence to find the best set of features, we employed feature selection based on the feature relevance to identify significant factors as predictors for detection of depression. At each iteration of boosting, the best feature is searched and then added to the ensemble. This forms a combination of selected feature with new features. Model is fit to the training data and the importance of each feature is measured using a weight. Weight is taken as the number of times each feature appears in a tree and using these weights, feature selection is done. The top 10 features are retained based on the accuracy of the classifier. We conducted a number of trails including different number of features. But 10 features were found to be optimum after which, addition of more number of features enhances the accuracy of the classifier negligibly. For instance when the number of features was chosen as 22 there was an increase of 0.26 % in the overall accuracy. If 22 features were chosen the resultant model would have been too complex. With 10 features, the accuracy of the classifier was 97.22% and with 22 features the accuracy of the classifier was 97.48% after which the accuracy became constant.

The feature accuracy graph considering different number of features is shown in Fig. 3. Out of the 45 features identified by the psychiatrist, 10 features were considered to be of importance based on feature selection process as shown in Fig. 4. The features used for the study are named automatically based on their index in the input array from f1 through f45.

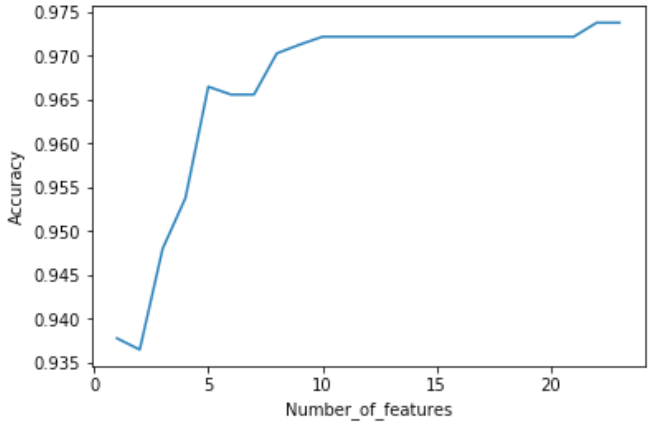


Fig. 3. Feature Accuracy graph.

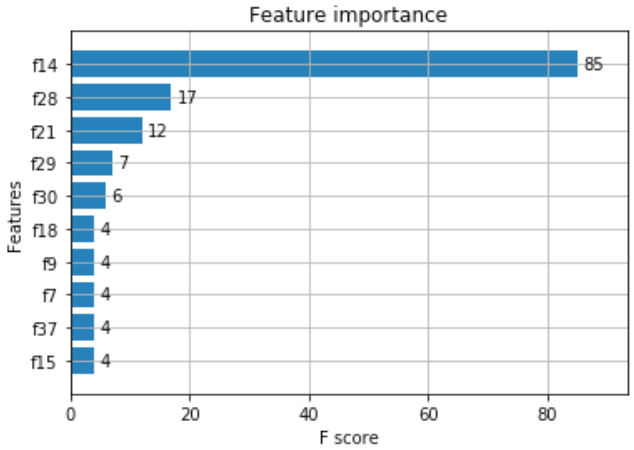


Fig. 4. Features of importance.

To identify the features of importance we manually mapped the indices to the feature names in the problem description and found that Eurotot has the highest importance followed by other features which are shown in Table I.

The importance of each feature considered for the study based on the psychiatrist's analysis was identified by computing the F-score using Eq. (2)

$$\text{F-score} = \frac{2*precision*Recall}{\text{precision}+\text{Recall}} \tag{2}$$

where precision and recall are based on confusion matrix in Table II.

Therefore the classifier model was built with 10 features given in Table I. The description of the features selected is given below:

- The Eurotot symptom scale is developed in Europe to compare the symptoms of geriatric depression across the continent. There are 12 eurotot items namely depressed mood, pessimism, wishing death, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness which represent the depression scores taken from GMS. Each item is scored 0 for absence of symptom and 1 for presence of symptom. Therefore an ordinal scale is generated with a maximum score of 12.
- avggrip indicates right and left hand grip strength measured as the amount of force a hand can squeeze around a dynamometer [20].
- HTN or hypertension is one of the strong predictors of Depression [21].
- Frifrailtytot is related to frailty which reduces a person's ability to endure environmental stress. Frifailtytot is the sum of frailty scores where 0 indicates unintentional weight loss, 1indicates exhaustion,

2 indicates muscle weakness, 3 indicates slowness while walking and 4 indicates low levels of activity [22].

- avgspifev is recorded through Spirometry Forced expiratory volume test and is the total amount of air exhaled in 1 second [23].
- cobese is the central obesity. [24].
- pjobgroup2 is associated with paid employment. The job categories used in this study are 0 indicating manager/administrator, 1 indicating professional, 2 indicating associate professional, 3 indicating clerical worker, 4 indicating shop keeper, 5 indicating skilled laborer, 6 indicating semi-skilled, 7 indicating laborer and 8 indicating agricultural worker.
- lencms is the length of the baby recorded at birth in cms. If the length of the baby at birth is less than the standard length, the foetal brain growth is shunted [25].
- m13ldlsi indicates the low density lipoprotein cholesterol measured in SI units. There will be an increased risk of depression if there are low cholesterol levels [26].
- bmi or the body mass index is calculated using the height and weight of a person which is a measure of body fat. There is a strong association between both underweight and obesity and depression [27].

TABLE I. FEATURES SELECTED USING XGBOOST TECHNIQUE

Feature Index	FEATURE
f14	Eurotot
f28	Avggrip
f21	HTN
f29	Frifrailtytot
f30	Avgspifev
f18	Cobese
f9	pjobgroup2
f7	Lencms
f37	m13ldlsi
f15	Bmi

D. Tuning Hyper Parameters

- Choosing depth indicates the height of tree that is used as an estimator and is computed using log loss function. It is run over multiple values and the best is chosen where loss is minimal as shown in Fig. 5.
- Choosing the number of estimators is an important task as we use an ensemble technique to make a collection decision for the model. We chose the number of estimators by computing the minimum loss as shown in Fig.6.

From the different trails we conducted, we set the number of estimators to be 50 with the depth of 3.

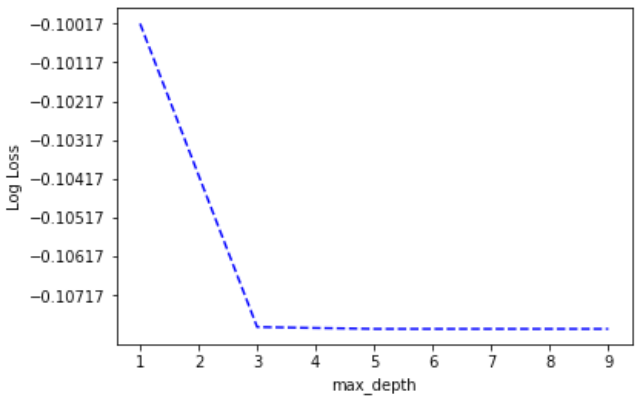


Fig. 5. Depth of the tree.

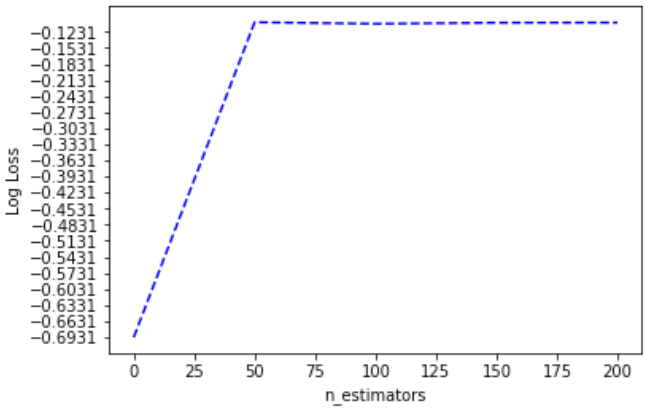


Fig. 6. Estimator setting.

The feature set obtained using boosting technique is given as input to the classification model built using McNN-PBL framework for detection of depression.

E. McNN-PBL Classifier

The two components of McNN are cognitive component and meta-cognitive component. The cognitive component consists of a three layered feed forward radial basis function network which takes each input x^t and produces an output $\hat{y}_j^t$. The meta-cognitive component consists of a dynamic model of the cognitive component. When a new training record x^t arrives, it takes the output $\hat{y}_j^t$ and the knowledge present in the new training record is estimated. Now based on this information, the meta-cognitive component selects an appropriate strategy to control the cognitive component in learning from the training record. By this way, the meta-cognitive component addresses the principles namely what-to-learn, when-to-learn, and how-to-learn properly.

1) Cognitive Component of McNN:

The three layered cognitive component of McNN consists of linear input, hidden and linear output layers in the network. The hidden layer of the cognitive component employs Gaussian activation function [28]. An assumption is made that the McNN builds K Gaussian neurons from $t-1$ training records without loss of generality. Let x^t be the given input, then $\hat{y}_j^t$ which is the predicted output of the j^{th} output neuron of McNN is as in Eq. (3).

$$\hat{y}_j^t = \sum_{k=1}^{K} w_{kj} h_k^t, j = 1,2,\ldots,n \tag{3}$$

In Eq. (3), w_{kj} represents the weight connecting the k^{th} hidden neuron to the j^{th} output neuron. h_k^t represents the response of the k^{th} hidden neuron to the input x^t as in Eq. (4).

$$h_k^t = \exp\left(-\frac{\|x^t-\mu_k^l\|^2}{(\delta_k^l)^2}\right) \tag{4}$$

In Eq. (4), $\mu_k^l \in Rm$ is the center and $\delta_k^l \in R$ is the width of the k^{th} hidden neuron. l represents the class of the hidden neuron to which it belongs.

For learning process, the cognitive component uses Projection Based Learning (PBL) algorithm.

2) Projection Based Learning (PBL) Algorithm:

The principle used by projection based learning algorithm is minimization of energy function. The algorithm finds the network output parameters with minimum energy function [29].The energy function that is considered is specified based on the error at McNN output neurons and shown in Eq. (5).

$$J_j = \sum_{j=1}^{n} \left(y_j^i - \hat{y}_j^i\right)^2 ; i = 1, \dots, N \quad (5)$$

As we assume above, the McNN is processing the t^{th} input, therefore, the overall energy function is defined as in Eq. (6).

$$J(W) = \frac{1}{2}\sum_{i=1}^{t} J_i = \frac{1}{2}\sum_{i=1}^{t}\sum_{j=1}^{n}\left(y_j^i - \hat{y}_j^i\right)^2 \quad (6)$$

Using Eq. (2) we substitute the predicted output ($\hat{y}_j^i$) and the energy function reduces to Eq. (7).

$$J(W) = \frac{1}{2}\sum_{i=1}^{t}\sum_{j=1}^{n}\left(y_j^i - \sum_{i=1}^{K} w_{kj} h_k^i\right)^2 \quad (7)$$

In Eq. 7 h_k^i indicates the response of the k^{th} hidden neuron for i^{th} training record. The optimal output weights ($W^* \in R^{KXn}$) are estimated in such a manner that the total energy reaches its minimum and is shown in Eq. (8).

$$W^* := \arg \min_{W \in R^{K \times n}} J(W) \quad (8)$$

Let $W \in R^{K \times n}$ If $J(W^*) \leq J(W) \forall W \in R^{K \times n}$ then W^* is the optimal output weight corresponding to the minimum of the energy function. W^* has minimum value when the first order partial derivative of J(W) with respect to the output weight is 0 as in Eq. (9).

$$\frac{\partial J(W)}{\partial w_{pj}} = 0, \; p = 1, \dots, K; \; j = 1, \dots, n \quad (9)$$

The first derivative is as shown in Eq. (10).

$$\frac{\partial J(W)}{\partial w_{pj}} = \sum_{i=1}^{t} h_p^i \left[y_j^t - \sum_{k=1}^{K} w_{kj} h_k^i\right] \quad p = 1, \dots, K; j = 1, \dots, n \quad (10)$$

Equating the derivative to 0, we get Eq. (11).

$$\sum_{k=1}^{K}\sum_{i=1}^{t} h_k^i h_p^i w_{kj} = \sum_{i=1}^{t} h_i^p y_j^i \quad (11)$$

Let matrix $A \in R^{K \times K}$ be as in Eq. (12).

$$a_{kp} = \sum_{i=1}^{t} h_k^i h_p^i, \; k = 1, \dots, K; \; p = 1, \dots, K \quad (12)$$

and the matrix $B \in R^{K \times n}$ be as in Eq. (13)

$$b_{pj} = \sum_{i=1}^{t} h_p^i y_j^i, \; j = 1, \dots, n; \; p = 1, \dots, K \quad (13)$$

Equation 10 can be rewritten as Eq. (14).

$$\sum_{k=1}^{K} a_{kp} w_{kj} = b_{pj}, \; p = 1, \dots, K; \; j = 1, \dots, n \quad (14)$$

or in a matrix form as Eq. (15).

$$A\,W = B \quad (15)$$

Therefore, the optimal W^* is as in Eq. (16).

$$W^* = A^{-1} B \quad (16)$$

With this optimal value of W^*, the energy function reaches its minimum value.

3) Meta-cognitive Component of McNN :

As measures of knowledge in the new training record, the meta-cognitive component uses estimated class label ($\hat{c}^t$), maximum hinge error (E^t) and spherical potential based class-wise significance to control the learning process of the cognitive component. The definitions of these measures are as below:

4) Estimated Class Label ($\hat{c}^t$):

The estimated class label ($\hat{c}^t$) is obtained using the predicted output ($\hat{y}^t$) as shown in Eq. (17).

$$\hat{c}^t = \arg \max_{j \in 1,2,\dots,n} \hat{y}_j^t; t = 1, \dots, N \quad (17)$$

5) Maximum Hinge Error (E^t):

The main objective of the classifier is to reduce the error between the predicted output ($\hat{y}^t$) and the actual output y^t such that the error is minimized. The classifier that uses hinge loss function estimates more accurate posterior probability compared to the classifier developed using mean square error function. Therefore, in McNN, the hinge loss error $\left(e^t = \left[e_1^t, \dots, e_j^t, \dots, e_n^t\right]^T\right) \in R^n$ is used and is defined as in Eq. (18).

$$e_j^t = \begin{cases} 0 \; if \; y_j^t \times \hat{y}_j^t > 1 \\ y_j^t - \hat{y}_j^t \; otherwise \end{cases} \; j = 1,2, \dots, n \quad (18)$$

The maximum absolute hinge error (E^t) is as in Eq. (19).

$$\left(E^t = \max_{j \in 1,2,\dots,n} \left|e_j^t\right|\right) \quad (19)$$

6) Class-wise Significance (Ψ_c):

The input feature (x^t) is mapped on to a hyper-dimensional spherical feature space S by using K Gaussian neurons, i.e., $x^t \rightarrow \phi$. Hence, all $\phi(x^t)$ lie on a hyper-dimensional sphere.

In McNN, the feature space S is described by center (μ) and width (σ) of the Gaussian neurons. Let the center of the K-dimensional feature space be $\phi_0 = \frac{1}{K}\sum_{k=1}^{K} \phi(\mu_k)$.

The potential of the new data x^t in original space is the knowledge present in it. It is the squared distance from the K-dimensional feature space to the center ϕ_0. The potential is as shown in Eq. (20).

$$\Psi = \left\|\phi(x^t) - \phi_0\right\|^2 \quad (20)$$

Eq. (20) can be expressed as in Eq. (21).

$$\Psi = \phi(x^t, x^t) - \frac{2}{K}\sum_{k=1}^{K} \phi\left(x^t, \mu_k^l\right) + \frac{1}{K^2}\sum_{k,r=1}^{K} \phi(\mu_k^l, \mu_r^l) \quad (21)$$

In the Gaussian function, the first term $\phi(x^t, x^t)$ and last term $\frac{1}{K^2}\sum_{k,r=1}^{K} \phi(\mu_k^l, \mu_r^l)$ in Eq. (21) are constants. These constants are discarded because potential is a measure of novelty. The potential can be reduced to Eq. (22).

$$\Psi = -\frac{2}{K}\sum_{k=1}^{K} \phi\left(x^t, \mu_k^l\right) \quad (22)$$

The class-wise distribution influences the performance of the classifier significantly. Therefore the measure of the spherical potential of the new training record x^t belonging to class c with respect to the neurons associated to same class is used. That is $l=c$.

The class-wise spherical potential or class-wise significance (Ψ^c) is defined using K^c which is the number of neurons associated with the class c as shown in Eq. (23).

$$\Psi_c = -\frac{2}{K^c}\sum_{k=1}^{K^c} \phi\left(x^t, \mu_k^c\right) \quad (23)$$

The knowledge contained in the records is directly indicated by the spherical potential. A smaller potential means that the record is similar to the existing knowledge in the cognitive component. A higher potential that is close to zero means that the record is novel.

The main objective of McNN is to approximate the underlying function that maps $x^t \epsilon R^m \rightarrow y^t \epsilon R^m$. Initially the neuron network starts with zero neuron. Addition of new neurons or update of existing

neurons is done during the processing of records [28]. The 4 strategies used by the McNN are:

1. Sample/Record deletion:

   In our study the dataset has to be classified as Depressed or Not Depressed. Therefore the model has to be trained to classify as Depressed and Not Depressed using some significant features when a new patient record is given for classification. Suppose the new record is the same as the trained one then that record is deleted.

2. Sample/Record reservation:

   Suppose the input record is very much similar to the trained record with slight change then that feature of the new record is reserved for further use.

3. Parameter updating:

   Suppose all the features of the input record are the same as trained record and a new feature is recognized then the parameter is updated in the training set.

4. Neuron Addition:

   As the new features are updated along with the new input records, the neuron addition or growth happens.

7) PBL Algorithm for McNN Classifier:

PBL algorithmic steps for McNN Classifier are as shown below:

1. Using Eq. (3) & (4), compute the cognitive component output ($\hat{y}^t$) for each new training record input (x^t).
2. Using Eq. (17),(19) & (23), for the new training record (x^t), the meta-cognitive component finds ($\hat{c}^t$), (E^t) and (Ψ_c) which are estimated class label, maximum hinge error and class-wise significance measures respectively.
3. One of the strategies like Deletion of record, Growth of Neuron, Updating of parameters and Reservation of record are selected by the meta-cognitive component.
4. The above selected strategy is executed by the cognitive component.
5. Until there are no more records in the training data set iterate steps 1 to 4.

F. Particle Swarm Optimization (PSO)

To maximize the performance efficiency, optimized parameter values must be set for each dataset. The user can find the optimized parameter after choosing data files where each of the 9 parameters lies in a range which can be set manually. Otherwise default values for the 9 parameters are set based on several experimentation. The default ranges set are as shown in Fig. 7.

The dataset consists of two files namely training file and testing file containing training data and testing data respectively. The training file is used to train the McNN classifier by creating and growing its neurons network using PBL and the testing file is used to test the classifier to assess the efficiency of the algorithm. The performance of the classifier depends on 9 parameters namely Skip threshold (β_b), Adding error threshold (β_a), Learning error threshold (β_u), Self-adaptive decay factor (δ), Overlap factor (K_2), Overlap factor for first neuron (K_1), Maximum number of reserved records, Spherical potential threshold (ϕ_S) and Center shifting factor (ζ). Hence for the classifier to perform better these parameters need to have best values fixed. We use a computational method called PSO in our study to fix best values for these parameters. The implementation of PSO consists of a class that provides methods to optimize parameters for a neuron network. It also keeps the default values and the ranges of the 9 parameters of a neuron network. For MYNAH data set, PSO is used to achieve the best parameters as shown in Fig. 8.

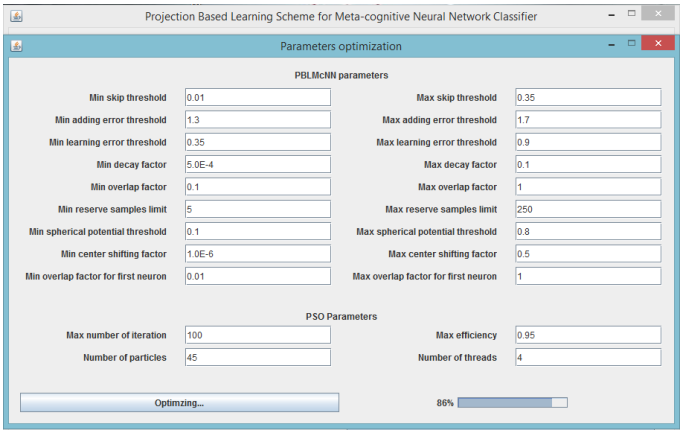


Fig. 7. Default range for 9 parameters.

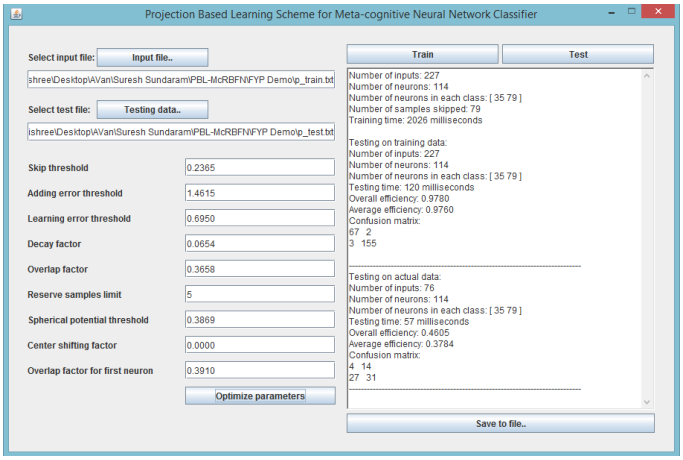


Fig. 8. Optimized Parameters for MYNAH cohort.

PSO algorithm contains a vector of the parameters. A parameter lies within a range, and it is initialized as in Eq. (24).

$$p = p_{min} + r \times (p_{max} - p_{min}) \tag{24}$$

Where p is the parameter initialized value, p_{min} is the minimum value of the parameter, p_{max} is the maximum value of the parameter and r is a random number which can be created by Math function.

Each particle of PSO algorithm also has a certain velocity, which is a vector of the 9 parameters. In McNN, the velocity is initialized as in Eq. (25).

$$v_i = r \times 0.1 \times (p_{i\,max} - p_{i\,min}) \tag{25}$$

Where v_i is the velocity of dimension i (i=$1, ..., 9$), r is a random number which is generated using the function Math.random(), $p_{i\,max}$ is the max value of i^{th} parameter (i=$1, ..., 9$) and $p_{i\,min}$ is the min value of i^{th} parameter (i=$1, ..., 9$).

The main objective function of PSO algorithm is to enhance the overall testing efficiency. The program creates N particles to optimize parameters. In each iteration, train() and test() methods are invoked using the parameter set for each particle. To evaluate the particle current position we use the overall testing efficiency of the model. To terminate the application we used two stopping criteria namely number of iterations and efficiency. If the number of iterations is greater than a certain value, we have set 100 as default, then the application terminates. The second stopping criterion is the efficiency. If the overall training efficiency is less than 0.95 or if the overall testing efficiency is greater than 0.95 then the application is terminated. The stopping criterion for the optimization process is the global best value. If the global best value

is greater than 1.95 then optimizing process terminates. Since PSO takes more time to optimize because it requires iterations with many particles, optimizeParameters() method was implemented to temporarily store the best set of parameters into a file after each iteration where the file contains the current global best value, the current number of iteration, the current data set, and the current set of parameters. Therefore best results can be kept in case unexpected errors occur.

We conducted experiments to fix best values for the 9 parameters of McNN-PBL classifier using other meta-heuristic methods like Genetic algorithm and Ant Colony optimization algorithm. With the implementation of these methods there was increase in misclassification error compared to PSO. Hence we used PSO to fix the best values to improve the performance of McNN-PBL classifier.

G. Result Analysis

In our study, based on the selection of estimators and selection of depth, gender-wise variation in prevalence of depression is reported. As expected depression was more common in women compared to men and among those who were not in paid employment and is shown in Fig. 9 and variable 2 in Fig. 11. Unintentional weight loss is found more in men with BMI < 18.5 kg/m2. Exhaustion and muscle weakness are higher in men because 62% are in paid job in our study when compared to women. Slowness while walking is prevalent in women due to higher body mass. Frifrailtytot graph is as shown in Fig. 10. Pattern of depressive symptoms varied by gender: The levels of Guilt, appetite, irritability, interest and impaired concentration were more common among women; the levels of insomnia or sleep, tearfulness and enjoyment were prominent symptoms among men. Suicidality or wishing death are more prevalent in men than women as shown in Fig. 11. According to Fig. 12 parameter avggrip is found higher in men since the BMI of men fall within the healthy range. In the Fig. 13, 64% of men and 79.62% of women are obese. The rate of obesity in women is high since their BMI is higher compared to men. HTN is found to be more prevalent in women since the BMI does not fall under healthy range as in Fig. 13 and Fig. 14.

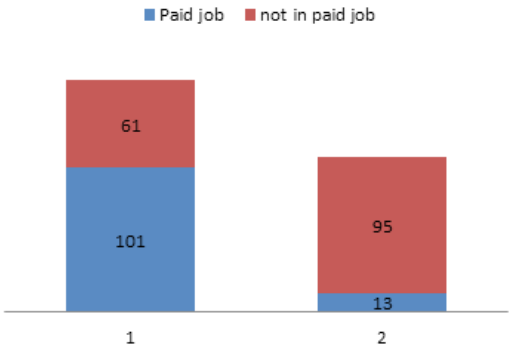


Fig. 9. Class wise distribution of Pjobcat2.

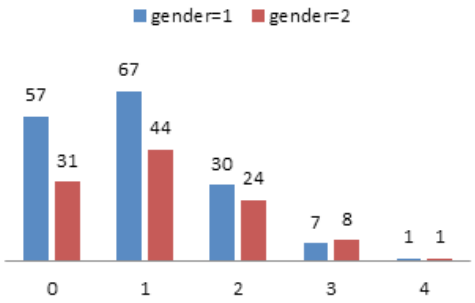


Fig. 10. Class wise distribution of Frifrailtytot.

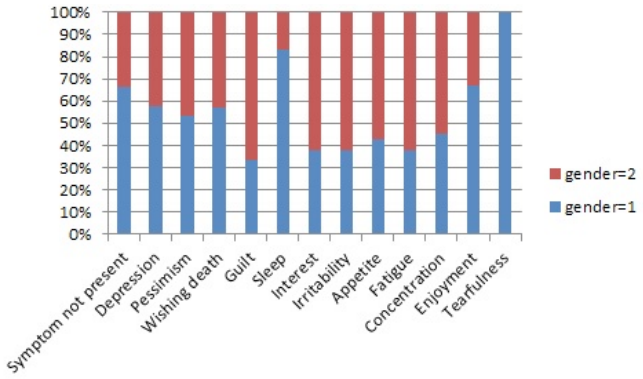


Fig. 11. Class wise distribution of Eurotot.

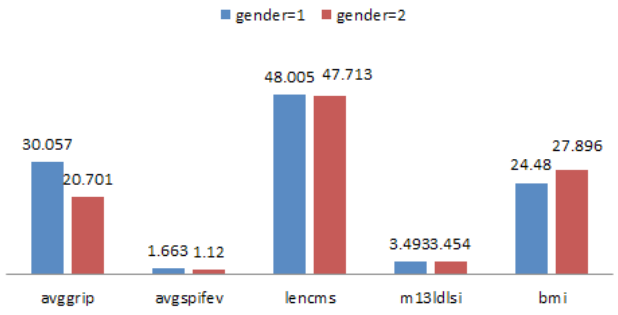


Fig. 12. Class wise distribution of other parameters.

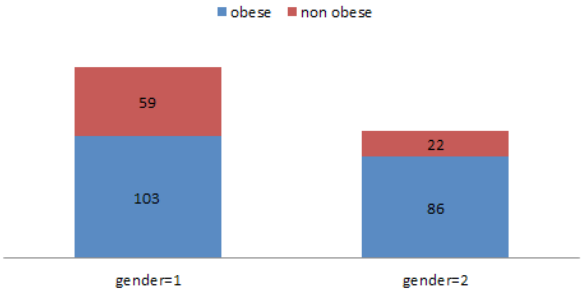


Fig. 13. Class wise distribution of Cobese.

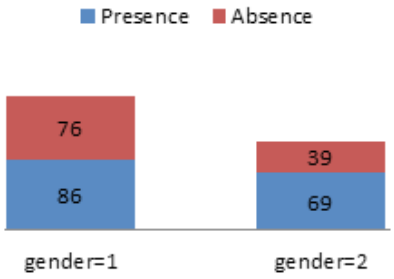


Fig. 14. Class wise distribution of HTN.

A 10 fold-Stratified cross validation evaluation technique was used to enforce the same distribution of classes in each fold of training, which handles the class imbalance problem. The average accuracy in this method is 97.80% and standard deviation of 2.98%.For each trial in the study, a ratio 75:25 of training and testing samples is randomly chosen.

IV. Performance Evaluation

The model was built using Java and the experiment on feature selection was conducted using Python running on Intel i3 processor. To evaluate the classifier performance, misclassification rate and correct classification rate play important roles. To find these we use sensitivity

and specificity values which are defined as in Table II. The parameters Eurotot, paid employment, Education and Obesity contribute to better classification that helps clinician to identify patients with depression. Adding these parameters to the model improved the efficiency of the classification.

TABLE II. CONFUSION MATRIX

Confusion Matrix Positive		**Target** Negative			
Model	Positive	a	b	Positive Predictive value	a/(a+b)
	Negative	c	d	Negative Predictive value	d/(c+d)
		Sensitivity	Specificity	**Accuracy** (a+d)/(a+b+c+d)	

In the confusion matrix a is True Positive, the number of patients correctly classified as Depressed, b is True Negative, the number of patients correctly classified as Not Depressed, c is False Positive, the number of Not Depressed patients classified as Depressed and d is False Negative, the number of Depressed patients classified as Not Depressed. In disease diagnosis, sensitivity plays a significant role. Classifying a patient who is Not Depressed as Depressed is better than classifying a Depressed patient as Not Depressed.

The performance of the McNN-PBL classifier with and without boosted feature selection and feature selection using PCA on MYNAH Cohort is shown in table III and Fig. 15. McNN-PBL classifier picks out the best sequence of training records and produces a sensitivity of 94.55% on the testing records, 95.40% when PCA was used for feature selection and 97.8% when XGBoost technique was used for feature selection which indicates that the classifier performance improved with reduced misclassification rate. McNN-PBL algorithm used 70 neurons of which 28 were for class-1(Depressed) and 42 for class-2 (Not Depressed) in both training and testing phases. The model used 599 milliseconds while training phase and 16 milliseconds while testing phase to produce training accuracy of 98.45 percent and testing accuracy of 97.80 percent.

TABLE III. PERFORMANCE OF MCNN-PBL CLASSIFIER ON MYNAH COHORT

Testing Phase	Without feature selection (%)	With feature selection using pca (%)	With feature selection using xgboost (%)
Efficiency	89.40	91.00	94.57
Specificity	54.11	90.45	94.10
Sensitivity	93.40	94.55	97.80
Execution Time (milli-seconds)	156ms	31ms	16 ms

After testing, the results were compared with the results of other algorithms including Self-adaptive Resource Allocation Network (SRAN), Extreme Learning Machines (ELM) and Support Vector Machines (SVM) and the results of the application of McNN-PBL with boosted feature selection using MYNAH cohort. All the results were compared based on properties like training efficiency, testing efficiency and number of neurons as shown in table IV.

From Table IV and Fig. 16, we can see that McNN-PBL classifier performed better than the best performing SRAN classifier on MYNAH cohort. McNN-PBL classifier also performed better than ELM and SVM classifiers. Also as shown in Table III, the McNN-PBL classifier with boosted feature selection performed much faster than the McNN-PBL classifier without feature selection and feature selection using PCA.

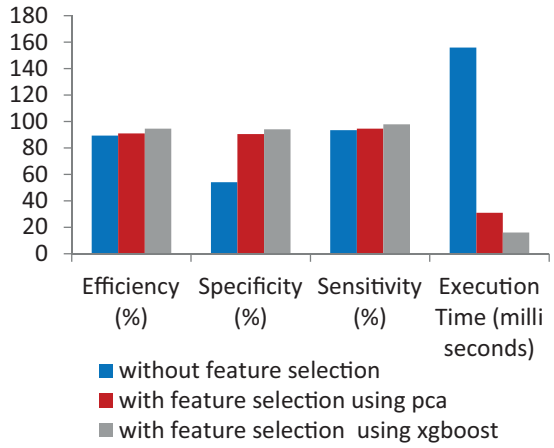


Fig. 15. Performance chart of the classifier.

TABLE IV. COMPARISON OF RESULTS USING MYNAH COHORT

Classifier	Training efficiency (%)	Testing efficiency (%)	Number of neurons	Training time in seconds
McNN-PBL	98.45	97.80	70	0.059
SRAN	98.00	92.29	69	0.088
ELM	100.00	90.67	120	0.150
SVM	100.00	90.62	110	0.140

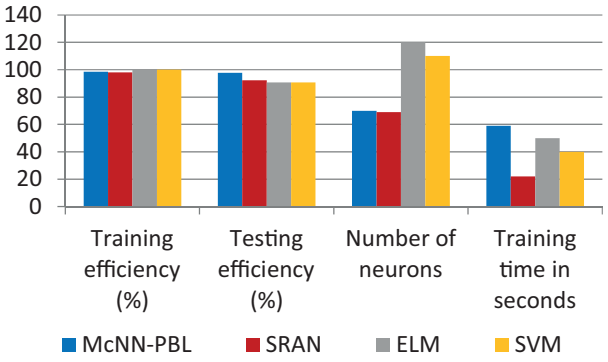


Fig. 16. Result comparison between different classifiers.

The limitation of the proposed model is that the process of optimization of parameters usually takes more time. Hence there is a need to store the optimized parameters. Therefore the optimized parameters are stored in training files so that the program only needs to run the parameter optimizing process once and reuse the optimized parameters in later executions for each training data file which slightly improves the execution time. To reduce this time complexity, other meta-heuristic methods have to be implemented to optimize the parameters.

V. CONCLUSION AND FUTURE ENHANCEMENT

A novel approach is presented in this paper for an accurate classification of depression using XGBoost technique, McNN-PBL classifier and PSO algorithm. For performance analysis, MYNAH cohort with 1321 patient records and 1201 parameters was used. Using the boosted feature selection technique, the McNN-PBL classifier

showed significant testing efficiency. The model specifically helps the clinician identify depressed patients resulting in improved treatment and prevention of progression of depression leading to self harm, schizophrenia, obsessive–compulsive disorder, etc. Using this machine learning approach relative to other approaches like SRAN, ELM and SVM, we can better identify the significance of parameters and can define whether the patient has depression or not with less memory requirement and computation time and reduced misclassification error. An experiment on Particles Swarm Optimization (PSO) algorithm was done to use it to find the best parameters for McNN-PBL classifier which significantly improved the efficiency of the classifier. To improve the application performance in terms of speed and usability, as a future work, experiments may be conducted using meta-heuristic optimization methods namely Grey-wolf optimizer, Moth-flame optimizer and others. Another problem for future work would be to identify different types of depression based on the symptoms and other psychiatric diseases using MYNAH cohort and classify them using McNN-PBL classification model. Psychiatric disorders can be predicted using other classifiers and a comparative study can be done. The model can also be trained to predict other cognitive disorders using MYNAH cohort which helps in improved mental health of the people. For the large dataset we can run the model using parallel threads for better performance.

Acknowledgment

This study was funded by the Wellcome DBT India Alliance and supported through a core programme grant to the MRC Lifecourse Epidemiology Unit, University of Southampton by the Medical Research Council, UK and Department for International Development, UK to Dr. Murali Krishna. We are greatly thankful to the research assistants at CSI Holdsworth Memorial Hospital, Mysore for data management. We sincerely thank Dr. Suresh Sundaram, Associate Professor, School of Computer Engineering, Nanyang Technological University, Singapore and Dr. B. S. Mahananda, Associate Professor, Department of Information Science and Engineering, SJCE, JSS S&T University, Mysuru, India for their valuable suggestions during the execution of this study.

References

[1] Vahia, Vihang N. “Diagnostic and Statistical Manual of Mental Disorders 5: A Quick Glance.” Indian Journal of Psychiatry, May 2018.

[2] B.S. Mahanand, S. Suresh, N. Sundararajan, M. Aswatha Kumar, “Identification of brain regions responsible for Alzheimer’s disease using a Self-adaptive Resource Allocation Network”, Neural Networks, Volume 32, August 2012.

[3] G.Sateesh Babu, S.Suresh, “Meta-cognitive Neural Network for classification problems in a sequential learning framework”, Neurocomputing, Volume 81, 1 April 2012.

[4] Tianqi Chen,Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”, KDD ‘16 The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining .San Francisco, CA, USA, August 13 - 17, 2016.

[5] Ismail Babajide Mustapha and Faisal Saeed, “Bioactive Molecule Prediction Using Extreme Gradient Boosting”, Molecules, 2016.

[6] Zhuo Chen,Fu Jiang, Yijun Cheng, Xin Gu, Weirong Liu, et.al., “XGBoost Classifier for DDoS Attack Detection and Analysis in SDN-based Cloud”, IEEE International Conference on Big Data and Smart Computing, 2018.

[7] Xun Wei, Fan Jiang, Feng Wei, Jiekui Zhang, et al., “An Ensemble Model for Diabetes Diagnosis in Large-scale and Imbalanced Dataset”, Proceedings of the Computing Frontiers Conference, Siena, Italy, May 2017.

[8] Haslam, Beck, “Categorization of major depression in an outpatient sample”, J. Nerv.Ment, 2012.

[9] Costafreda, S.G., Chu, C., Ashburner, J., Fu, C.H., “Prognostic and diagnostic potential of the structural neuroanatomy of depression”, PLoS One, 2009.

[10] Hahn, T., Marquand, A.F., Ehlis, A.C., Dresler, T., Kittel-Schneider, S., Jarczok, T.A.,Fallgatter, A.J., “Integrating neurobiological markers of depression”, Arch. Gen. Psychiatry, 2011.

[11] V Arun, Arunkumar B.V., Padma S.K., Shyam V, “Disease Classification and Prediction using Principal Component Analysis and Ensemble Classification Framework”, International Journal of Control theory and applications, ISSN:0974-5572, Vol 10, No. 14, 2017.

[12] Rubeena Banu; V Arun; N. Shankaraiah; Shyam V., “Meta-cognitive neural network method for classification of diabetic retinal images”, Second International Conference on Cognitive Computing and Information Processing (CCIP) 2016.

[13] Mwangi, B., Matthews, K., Steele, J.D., “Prediction of illness severity in patients with major depression using structural MR brain scans”, J. Magn. Reson, 2012.

[14] Braeken, J., & Van Assen, An empirical Kaiser criterion. Psychological Methods, http://dx.doi.org/10.1037/met0000074, 2017.

[15] Liton Chandra Paul, Abdulla Al Suman and Nahid Sultan, Methodological Analysis of Principal Component Analysis (PCA) Method, IJCEM International Journal of Computational Engineering & Management, Vol. 16 Issue 2, March 2013.

[16] Bhagyashree, S.R., Nagaraj, K., Prince, Muralikrishna., “Diagnosis of Dementia by Machine learning methods in Epidemiological studies: a pilot exploratory study from south India”, Soc Psychiatry Psychiatr Epidemiol, 2018.

[17] John RM Copeland, Martin Prince, K C M Wilson, “The Geriatric Mental State Examination in the 21st century”, International Journal of Geriatric Psychiatry, August 2002.

[18] John RM Copeland, M. R. E. Dewey, Scott Henderson, “The Geriatric Mental State (GMS) used in the community: Replication studies of the computerized diagnosis AGECAT”, Psychological Medicine, March 1988.

[19] Martin Prince, Daisy Acosta, Helen Chiu, “Effects of education and culture on the validity of the Geriatric Mental State and it AGECAT algorithm”, The British Journal of Psychiatry, December 2004.

[20] Massy-Westropp, N. M., Gill, T. K., Taylor, A. W., Bohannon, R. W., & Hill, C. L., “Hand Grip Strength: age and gender stratified normative data in a population-based study”, BMC Research Notes, 2011.

[21] Meurs M, Groenewold NA, Roest AM, et al., “The associations of depression and hypertension with brain volumes: Independent or interactive?” NeuroImage : Clinical, 2015.

[22] Lally F, Crome P. “Understanding frailty”. Postgraduate Medical, Journal. 2007.

[23] https://en.wikipedia.org/wiki/Spirometry.

[24] Agarwal A, Agarwal M, Garg K, Dalal PK, Trivedi JK, Srivastava JS. “Metabolic syndrome and central obesity in depression: A cross-sectional study”. Indian Journal of Psychiatry, 2016.

[25] Broekman BF1, Chan YH, Chong YS, Kwek K, et al., “The influence of anxiety and depressive symptoms during pregnancy on birth size”, 2014.

[26] https://www.greatplainslaboratory.com/articles-1/2015/11/13/the-implications-of-low-cholesterol-in-depression-and-suicide.

[27] Leonore M. de Wit, Annemieke van Straten, Marieke van Herten, Brenda WJH Penninx, and Pim Cuijpers, “Depression and body mass index, a u-shaped association”, BMC Public Health, Longitudinal assessment of neuropsychological functioning major depression. Aust. N. Z. J. Psychiatry, 2009.

[28] Ramasamy Savitha, Sundaram Suresh, Narasimhan Sundararajan,” Projection-Based Fast Learning Fully Complex-Valued Relaxation Neural Network”, IEEE Transactions On Neural Networks And Learning Systems, Vol. 24, No. 4, April 2013.

[29] G. Sateesh Babu, R. Savitha, S. Suresh “A projection based learning in Meta-cognitive Radial Basis Function Network for classification problems”, International Joint Conference on Neural Networks (IJCNN), 2012.

Vanishri Arun

Vanishri Arun received Bachelor of Engineering degree in Computer Science and Engineering in the year 1995 and Master of Technology in Software Engineering in the year 2009 from Sri Jayachamarajendra College of Engineering, Mysuru, India and currently pursuing Ph.D from Visvesvaraya Technological University, Belagavi, India in Biomedical Informatics under the guidance of Dr. Shyam V and Dr. S K Padma and medical experts Dr. Arunkumar B V and Dr. Murali Krishna. She is working in the capacity of Assistant Professor in the Department of Information Science and Engineering at JSS Science and Technology University, Mysuru, India. She has teaching experience of about 12 years at technical institutes and has industrial experience of about 8 years in software development. Her research interests include Bio-Medical Informatics, Machine Learning and Data Analytics.

Murali Krishna

Dr Murali Krishna is a Consultant in Old Age Psychiatry and Psychiatric Epidemiologist. He graduated from the Government Medical College, Mysuru, India and completed his specialist training in Psychiatry from University of Cambridge and Liverpool UK. He finished his higher specialist training in Old Age Psychiatry from University of Liverpool UK. He is a member of the Royal College of Psychiatrists UK. He has an advanced Masters Degree in Research from University of Sheffield. He is a honorary Senior Lecturer at Faculty of Health and Social Care, Edge hill University in England and a honorary Research Fellow from School of Social Sciences, University of Bangor, Wales. His current research work examining life course risk factors for late life neurocognitive disorders is supported by a prestigious award from Welcome Trust UK and DBT India.

Arunkumar B V

Dr. Arunkumar received MBBS in the year 1996 and Diploma in Anaesthesiology in 1999 from JSS Medical College, Mysuru, India, Diplomate of National Board in Anesthesia (DNB) from National Board of Examination, Ministry of Health, New Delhi in 2004 and EGMP on Business Administration from IIM-Bangalore in 2016. He is currently heading the Department of Anesthesiology, Apollo BGS Hospitals, Mysuru, India. His research interests are Medical Informatics, Data Analytics and Cognitive science.

S K Padma

Dr. Padma has pursued Ph.D. in Computer & Information Sciences. Currently a professor in the Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, India and has been guiding many research scholars in the areas of Biomedical Informatics, Natural Language Processing, Machine Learning, Data Analytics and others. She is a life member of Indian Society for Technical Education (ISTE), New Delhi, India, Institution of Engineers, India and Computer Society of India.

Shyam V.

Dr. Shyam holds a doctorate in Real Time Embedded Systems, specialising in Parallel Computer Architecture from Indian Institute of Science (gold Medal) & has been with CG Smith, Ericsson, Tata Consultancy Services and as Director of Technology at Philips. He was instrumental in devising one of the first hardware based antivirus solution in 90s. He currently holds the posts of President and CTO at Forus Health (P) Ltd, Director at MYMO Wireless Technology (P) Ltd at SID – IISc, Bangalore and Technical Director at Maastricht University Medical Center in Netherlands. His innovation 3nethra, all in one eye screening device, at Forus has won several laurels like DST Locked Martin Gold medal, Perimal Award, Sankalp Award, CNBC TV18 Award, Samsung Innovation Quotient Award, NASCOM Innovation Award and Anjani Mashelkar Inclusive Innovation Award for 2011. He has filed more than 20 international patents, published around 50 papers in international conferences and journals.

Diabetes Diagnosis by Case-Based Reasoning and Fuzzy Logic

Mohammed Benamina, Baghdad Atmani, Sofia Benbelkacem*

Laboratoire d'Informatique d'Oran (LIO), University of Oran 1 Ahmed Benbella (Algeria)

Received 9 octubre 2017 | Accepted 19 enero 2017 | Published 2 February 2018

Abstract

In the medical field, experts' knowledge is based on experience, theoretical knowledge and rules. Case-based reasoning is a problem-solving paradigm which is based on past experiences. For this purpose, a large number of decision support applications based on CBR have been developed. Cases retrieval is often considered as the most important step of case-based reasoning. In this article, we integrate fuzzy logic and data mining to improve the response time and the accuracy of the retrieval of similar cases. The proposed Fuzzy CBR is composed of two complementary parts; the part of classification by fuzzy decision tree realized by Fispro and the part of case-based reasoning realized by the platform JColibri. The use of fuzzy logic aims to reduce the complexity of calculating the degree of similarity that can exist between diabetic patients who require different monitoring plans. The results of the proposed approach are compared with earlier methods using accuracy as metrics. The experimental results indicate that the fuzzy decision tree is very effective in improving the accuracy for diabetes classification and hence improving the retrieval step of CBR reasoning.

Keywords

Case-Based Reasoning, Case Retrieval, Classification, Data Mining, Diabetes Application, Diabetes Diagnosis, Fuzzy Decision Tree, Fuzzy Rule Base, Rule Induction.

DOI: 10.9781/ijimai.2018.02.001

I. Introduction

Decision support is not intended to replace the decision-maker by proposing solutions, but rather to guide him to decisions that he will have to take under his responsibility. Decision support can be found in many application domains such as economics, mathematics, computing, medicine, etc. In medicine, the decision is considered to be the center of the medical procedure. The medical decision-making process involves making a diagnosis, proposing treatment, a surveillance plan, etc. Thus, many applications of decision support have been developed in this field. These applications are intended to support healthcare personnel in their decision-making. This involves the use of various decision-support tools, such as case-based reasoning (CBR), which can be seen as a management tool for decision-making.

The CBR is a powerful methodology for designing intelligent systems. It is based on the re-use of solutions of similar cases in order to solve new encountered problems, by capitalizing acquired knowledge from past experiences. In daily life, we are frequently confronted with problems already experienced and which probably have similar solutions.

During their practice, practitioners use not only their theoretical knowledge but also their acquired experience. To this end, case-based reasoning should be exploited for medical decision-making and diagnosis of diabetics. Diabetes has become the fourth leading cause of death in developed countries and there is substantial evidence that it is reaching epidemic proportions in many developing and newly industrialized countries.

In this paper we propose a fuzzy diagnosis aid system based on data mining and case-based reasoning. CBR systems based on fuzzy logic are successfully applied in various fields including medical diagnosis. Our approach is applied to assist experts in the diagnosis of diabetes. The objective is to help experts in the field in the choice of diabetic surveillance plan.

The aim of this paper is to demonstrate the value of a fuzzy inference system guided by data mining in CBR modeling. It deals with the originality of the integration of different techniques derived from artificial intelligence (AI) in the stages of CBR and its ability to jointly use expertise and learning from data in the same context, especially that artificial intelligence tools become a part of the diabetes health care [43].

Case retrieval is the first step in the CBR cycle that requires using similarity measures to find similar cases to the given problem. Generally, the similarity measures used such as k nearest neighbors (k-nn) consist in calculating a distance between each case of the base and the case to be solved. However, when the case base is too large, similarity calculation will be expensive in computational time. To remedy this problem, we opted for the fuzzy decision trees in the retrieval of similar cases.

Retrieval is often considered the most important phase of CBR. In this paper, we integrate fuzzy logic and data mining to improve this step. The proposed Fuzzy CBR is composed of two complementary parts; the modeling part fuzzy realized by Fispro and the reasoning part realized by the platform JColibri. The use of fuzzy logic aims to reduce the complexity of calculating the degree of similarity that can exist between individuals who require different monitoring plans.

This paper is organized as follows. A state of the art about work using AI techniques in CBR in the medical field is presented in Section II. Section III is devoted to our FDT4CR approach. Finally, a conclusion and some perspectives are given in section IV.

II. State of the Art

Case-based reasoning is a general decision-making paradigm used

in the medical field [1]. A lot of work on CBR has been done in this area of research.

Choudhury and Begum [2] presented a literature review about case-based reasoning systems used in the medical field over the past few decades. In this study the difficulties of implementing CBR in medicine have been discussed. Blanco et al. [44] presented a review study focused on papers published between 2008 and 2011 in the field of case-based reasoning applied to the health sector. The authors of this review study presented a new proposal to improve the adaptation step of CBR. They used association rules to reduce the number of cases and facilitate learning around adaptation rules. El-Sappagh and Elmogy [45] identified research papers on CBR published between 1999 and 2015. Integrated evaluation metrics have been used for the analysis. The results show that there is a need for more comprehensive improvements in medical CBR particularly in diabetes systems. Montani et al. [3] proposed a case-based decision support tool, designed to assist physicians in the diagnosis of type 1 diabetes. The authors have put in place a two-stage procedure; first, it finds the classes to which the new case might belong; then it lists the most similar cases, using the k-nn method. Similarly, Schmidt and Gierl [4] used case-based reasoning techniques to speed up the process to propose appropriate treatment recommendations for patients with complications due to infection. Doyle et al. [5] presented a decision-support system using case-based reasoning for the treatment of the bronchiolitis. The system provides guidance based on previous cases. Montani et al. [6] presented a case based-reasoning system for the management of patients with end-stage renal disease and undergoing hemodialysis. This system helps practitioners to develop treatment plans for dialysis. Marling et al. [7] presented a CBR decision-making approach for the management of diabetic patients. To prevent serious complications of the disease, patients must constantly monitor their blood sugar levels and keep it as close to normal as possible, but maintaining good blood sugar control is a difficult task for patients and their doctors. A prototype system for CBR decision support has been designed to help with this task. Azar and Bitar [46] proposed to compare three artificial intelligence-based algorithms, CBR, C4.5, and Genetic Algorithms, to treat the problem of Diabetes Mellitus. The three algorithms have been evaluated on a Diabetes data set extracted from the UCI Machine Learning Repository. The results demonstrated the effectiveness of all three techniques. De Paz et al. [8] presented a CBR decision-support system for the diagnosis of different types of cancer.

In our study, we focused on the retrieval of cases stage, the main step in the case-based reasoning process. This step is usually dealt with k-nn [9]. Retrieval by k-nn is costly in computing time and its learning base generally requires a large memory space. Data mining could be useful to improve case retrieval. Data mining uses several techniques of computer science, statistics and artificial intelligence in various fields, especially for decision support. In the following, we will cite some works which encouraged us to contribute in this field, especially in the medical field for the classification of diabetes. Quellec et al. [10] presented a CBR approach about the diagnosis of diabetic retinopathy. They proposed to proceed to the adaption of the decision tree method in the context of remembering. The authors propose to adapt decision trees, generally used for classification, to image indexing context and therefore extend CBR to cases containing images. Burke et al. [11] calculated the similarities between the cases using decision trees. The case base refers to a decision tree where cases are the graph attributes. Huang et al. [21] suggest a model for a prognostic and diagnostic system for chronic diseases that integrates data mining and case-based reasoning. They adopted data mining techniques to discover the implicit meaningful rules from the health examination data and used the extracted rules for the specific prognosis of chronic disease. In this work, case-based reasoning is used to support the diagnosis and treatment of chronic diseases. Kushwaha and Welekar [52] used a data mining technique to develop an approach of feature selection for image retrieval based on content. The experimental result shows that the selection of features using Genetic Algorithm reduces the time for retrieval. Houeland [22] developed a RDT (random decision tree) algorithm implemented in a CBR framework. RDT algorithm has been combined with a simple similarity measure. They used two measures of similarity. A local subset of cases is selected using the first similarity measure. Then, the subset is reduced with the second measure of similarity by using a decision tree approach. RDT lets to define the similarity between two cases where trees are fully developed binary trees. The results have been evaluated in the area of palliative care of cancer.

Remembering cases must face a certain degree of vagueness and uncertainty that are almost always encountered while dealing with complex applications of the real world. A recent survey [12] presents the role of fuzzy sets and fuzzy logics in the various stages of CBR. This study showed that the integration of fuzzy logic in CBR resulted in successful hybrid systems. Boyen and Wehenkel [23] describe a new algorithm capable of deducing fuzzy decision trees in domains where most of the input variables are numeric and the output information is referred to as a fuzzy set. It includes three complementary steps: growth to select relevant attributes and fuzzy thresholds; Pruning to determine the appropriate tree complexity; the redevelopment to adjust the parameters of the tree in a global way. Begum et al. [24] proposed a case-based decision support system to assist clinicians in the diagnosis of stress. Case-based reasoning is applied as the main methodology to facilitate the re-use of the experience and the explanation of the decision by retrieving similar previous profiles. Other fuzzy techniques are also used and incorporated into the case-based reasoning system to address imprecision, inherent uncertainty in clinicians' reasoning, and inaccuracy of characteristic values. The work of Barrientos and Sainz [25] supports the decision-making on resource planning of an emergency call center in order to achieve its obligatory quality of service. This is done by searching the activity data collected by an emergency call center. A linguistic prediction, a categorization and a description of the days based on the activity and the information of the call center make it possible to know the workload of each category of day. This was generated by a fuzzy version of an unsupervised decision tree, merging decision trees and clustering. Levashenko and Zaitseva [26] proposed a decision support system based on fuzzy logic for the diagnosis of oncology diseases. The decision-making procedure corresponds to the classification of the new case by analyzing a set of instances for which the classes are known. Solved cases are defined as fuzzy classification rules that are formed by different fuzzy decision trees. Three types of fuzzy decision trees are considered in the document: unordered, ordered and stable. The induction of these fuzzy decision trees is based on cumulative estimates of information. Adidela et al. [27] proposed a hybrid classification system to predict the onset of diabetes. The system adopts three phases. In the first phase, the data are grouped together using the EM algorithm. The second phase performs the classification of the individual clusters obtained using fuzzy ID3. From the second phase of the process, adaptation rules are obtained. These rules are essential to the prediction of diabetes. In the third phase, the test tuple is provided to the rules to predict the class tag. Tahmasebian et al. [47] tried to use the information documented in the summary of patient medical records to measure their similarity rate. First, the effective parameters were extracted from the structured discharge summary prepared for patients. Then, the weights of the parameters were extracted using data mining methods. Finally, the fuzzy system has been used to compare similarities between the current case and other cases for patients with chronic kidney disease.

The proposed similarity measure has been compared with other methods in CBR systems. As a result, employing fuzzy method to measure the similarity can lead to a higher flexibility compared to other methods. El-Sappagh and Elmogy [48] proposed a CBR fuzzy ontology (CBRDiabOnto) for the diagnosis of diabetes, specifically for DM diagnosis. The resulting ontology supports many kinds of reasoning, such as crisp, semantic data, and fuzzy reasoning. These different data types improved case representation and case retrieval. The evaluation of CBRDiabOnto shows that it is consistent, accurate and covers logic and terminologies of diabetes mellitus diagnosis.

This literature review confirms that CBR systems based on fuzzy logic are successfully applied in various fields including medical diagnosis. The aim of our contribution is to demonstrate the value of a fuzzy inference system guided by data mining in CBR modeling. The use of fuzzy logic aims to reduce the complexity of calculating the degree of similarity and hence the retrieval step of CBR reasoning.

III. Proposed Approach FDT4CR

The main purpose of the proposed approach is to improve the accuracy of Diabetes classification. We integrate fuzzy decision tree to improve cases retrieval step of CBR process. The main contributions of FDT4CR are given as follows:

- Fuzzy decision tree classifier is used for the generation of a crisp set of rules.
- Fuzzy modeling is used to deal with uncertainty of medical reasoning.
- The proposed approach is a new combination of different methods which performs classification of diabetic patients using CBR, fuzzy modeling and data mining techniques.
- Improvement of the retrieval step of JColibri CBR process.

The proposed Fuzzy Inference Mechanism model reaches high prediction and classification accuracy, which improved both specificity and sensitivity. Numerous research works were carried out on the Pima Indian Diabetes Dataset (PIDD) with the problem of developing a prediction model. Karegowda et al. [28], [29] discussed the accuracy of the classification of 65 classfiers. They developed a combined model cascaded decision tree and k-means, which has a classification accuracy of 93.33% for categorized data without missing data [30].

The architecture of the proposed approach is shown in Fig. 1. It is composed of two complementary parts; the fuzzy modeling part using Fispro and the Case based reasoning part using JColibri platform.

FisPro has been used for various modeling projects [15].

JColibri [17] is an object-oriented tool used to develop CBR applications. It is an open-source platform developed in Java that defines a clear architecture to design case-based reasoning systems. The user can customize methods and classes of the platform as required. The JColibri platform has been widely used to develop CBR applications in the medicine field. Sharma and Mehrotra [49] used the jColibri framework to develop a CBR application for diagnosis of chronic kidney disease while Kiragu and Waiganjo [50] developed a prototype for treatment and management of diabetes using JColibri.

A. Architecture of Fuzzy Case-Based Reasoning

There are several studies found in the literature that have used various techniques on the Pima Indian Diabetes dataset to train and test data. The architecture of the proposed Fuzzy CBR applied on diabetes is shown in Fig. 1. FDT4CR is the process by which the retrieval phase of a CBR is modeled not in the conventional form of mathematical equations, but in the form of fuzzy rules base.

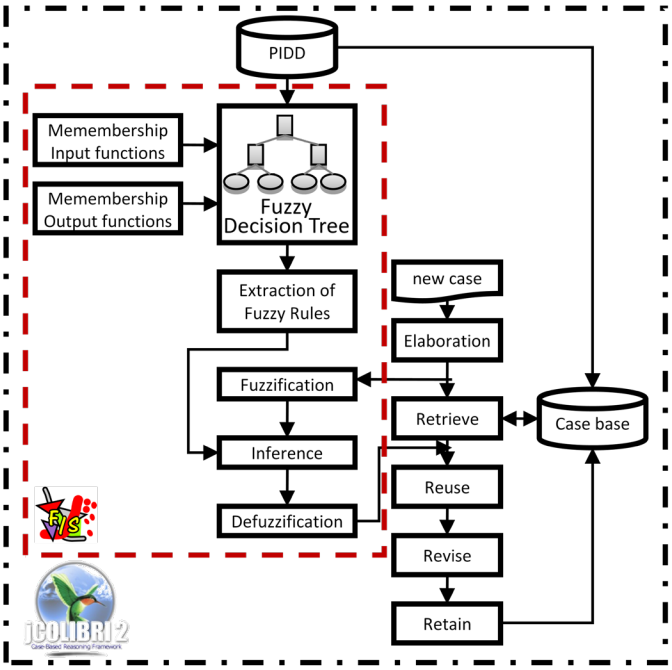


Fig. 1. The architecture of the proposed Fuzzy CBR.

B. Pima Indians Diabetes Database

The Pima Indian Diabetes Dataset (PIDD) has been taken from the UCI Machine Learning repository (http://archive.ics.uci.edu/ml/). The input variables are X_2: Plas (Plasma glucose concentration in 2-hours), X_5: Ins (2-hour serum insulin), X_6: Mass (Body mass index), X_7: Pedi (Diabetes pedigree function) and X_8: Age; the output variable is Y: DM (Diabetes Mellitus) (Table I). The data used to test the Fuzzy Inference Mechanism Framework has been taken with the age group from 25-30. The experiment compares the accuracy of the fuzzy classification task with results of studies involving the PIDD [19], [20].

TABLE I. PIDD Attributes

Exogenous Var	Abbreviation	Semantic
X_1	Preg	Number of times pregnant
X_2	Plas	Plasma glucose concentration in 2-hours
X_3	Dias	Diastolic blood pressure
X_4	Tric	Triceps skin fold thickness
X_5	Ins	2-hour serum insulin
X_6	Mass	Body mass index
X_7	Pedi	Diabetes pedigree function
X_8	Age	Age
Y	DM	Diabetes Mellitus where 1 is interpreted as tested positive for diabetes

C. Fuzzy Modeling Using Fispro

The Algorithm 1 explains the process of fuzzy inference mechanism. First, the fuzzification phase consists of gathering a crisp set of input data and converting it into a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. Then, an inference is executed according to a set of fuzzy rules. Finally, the defuzzification step makes it possible to transform the resulting fuzzy output into a crisp output using the membership functions.

Algorithm 1. Fuzzy logic algorithm

1. Initialization
   - Linguistic variables and terms are defined.
   - Construction of the membership functions.
   - Construction of the fuzzy decision tree and extraction of the fuzzy rule base.
2. Fuzzification
   - This step converts crisp input data to fuzzy values using the membership functions.
3. Fuzzy inference
   - Evaluate the rules in the fuzzy rule base.
   - Combine the results of each rule.
4. Defuzzification
   - This step converts the output data to non-fuzzy values.

1) Fuzzification

The conversion from crisp to fuzzy input is known as fuzzification [31]. Crisp inputs are converted to their fuzzy equivalent using a family of membership functions [53]. Furthermore, an interface is proposed to validate and tune parameters of constructed fuzzy numbers.

Each parameter is set with the minimum value, the mean, the standard deviation and the maximum value. The membership function $\mu(x)$ of the triangular fuzzy number that we have used is given by:

$$\mu(x) = \begin{cases} 0, x \le a \\ (x-a)/(b-a), a < x \le b \\ (c-x)/(c-b), b < x \le c \\ 0, x > c \end{cases} \quad (1)$$

The Memberships graphic for the fuzzy values with Fispro are given in Fig. 2 and Fig. 3.

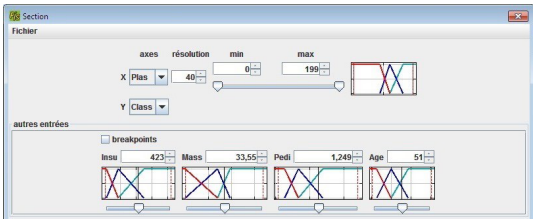


Fig. 2. Membership graphics for the fuzzy values of Insu, Mass, Pedi and Age.

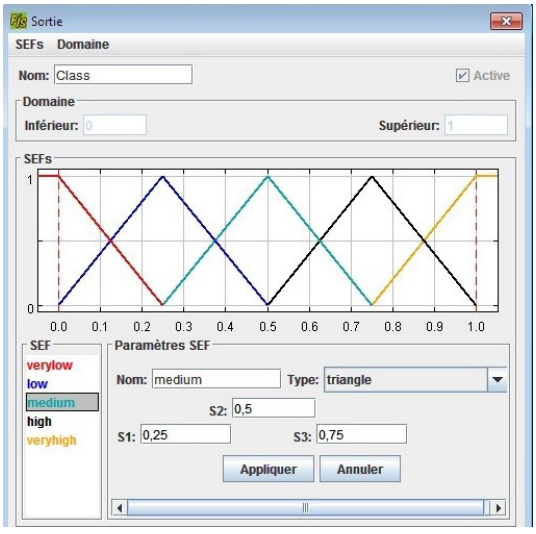


Fig. 3. Membership graphics of fuzzy values of Diabetes Mellitus.

Fuzzy numbers parameters are given in Table II. In this study other Fuzzy Triangular numbers are proposed inspired by the related work [32].

TABLE II. TRIANGULAR MEMBERSHIP FUNCTIONS PARAMETERS

Fuzzy variables	Fuzzy Numbers	Fuzzy Triangular numbers
Plas (X_2)	Low	[0, 88.335, 121.408]
	Medium	[88.335, 121.408, 166.335]
	High	[121.408, 166.335 ,199]
Ins (X_5)	Low	[0 ,17.276, 173.175]
	Medium	[17.276, 173.175, 497]
	High	[173.175,497, 846]
Mass (X_6)	Low	[0, 0, 27.792]
	Medium	[0, 27.792, 38.864]
	High	[27.792, 38.864, 67.1]
Pedi (X_7)	Low	[0.078, 0.272, 0.682]
	Medium	[0.272, 0.682, 1.386]
	High	[0.62, 1.386, 2.42]
Age (X_8)	Young	[21, 25.475, 40.537]
	Medium	[25.475, 40.537, 57.798]
	Old	[40.537, 57.798, 81]
DM (Y)	Very low	[0, 0, 0.25]
	Low	[0, 0.25, 0.5]
	Medium	[0.25, 0.5, 0.75]
	High	[0.5, 0.75, 1]
	Very high	[0.75, 1, 1]

2) Fuzzy Inference Engine

In our study, we make use of the FDT (Fuzzy Decision Tree) that is an extension of classical decision trees [33], [34]. The fuzzy decision tree proposed in FisPro (Fig. 4) is based on the algorithm [35]. The implementation of FisPro is based on a predefined fuzzy partition including *input* variables, which is left intact by the tree's growth algorithm.

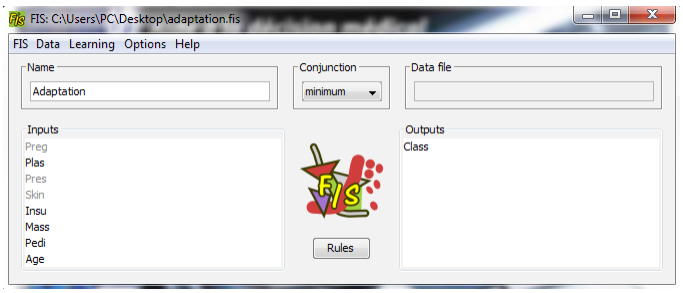


Fig. 4. Fispro interface for Decision Tree construction.

From a root node that includes all items of the data set, the fuzzy decision tree uses a recursive procedure to divide the nodes into M_j child nodes, where M_j is the fuzzy set number in the j^{th} input variable partition which is selected for the split. The algorithm proceeds to the selection of the variable which maximizes the gain based on a discriminated criterion for each node:

$$G_n^j = H_n - \sum_{i=1}^{M_j} w_i H_{n,i} \quad (2)$$

where H_n is the entropy of the node n, $H_{(n,i)}$ is the entropy of the node corresponding to the SEF (Fuzzy Sub Set) of the variable j (i=1,..,M_j) and w_i is the relative weight of SEF.

The entropy of a node n is defined as:

$$H_n = -\sum_k P_k^n * \log(P_k^n) \tag{3}$$

where P_k^n is the fuzzy proportion of examples of n node belonging to class K; it will be calculated according to the degrees of membership as follows:

$$P_K = \frac{\sum_{i=1}^{N} \mu_k(y_i) * \mu_k(x_i)}{\sum_{c=1}^{K} \sum_{i=1}^{N} \mu_k(y_i) * \mu_k(x_i)} \tag{4}$$

where K is the number of classes, N is the number of examples, $\mu_K(y_i)$ and $\mu_K(x_i)$ are the degrees of membership.

Entropy is the global disorder of the system. For example, for two classes in the case of a conventional tree, entropy is maximum if the probability of belonging to a class is 0.5, and minimum where the probability of belonging to a class is 1 (and thus 0 for the other class). The purpose of a decision tree, fuzzy or classical, is to successively divide the data space in order to reduce the maximum entropy.

FisPro is an open source tool used to create fuzzy inference systems (FIS). It is dedicated for reasoning purposes, specifically for simulating a physical or biological system [15]. It includes many algorithms (most of them implemented as C programs) to generate fuzzy partitions and rules directly from experimental data. It offers FIS visualization methods and data with a user-friendly interface based on java. We make use of the Fuzzy Decision Trees (FDT) [36] algorithm provided by FisPro.

After the fuzzy decision tree induction, we built a classifier based on rules. A fuzzy rule is given in the form of a simple IF-THEN rule with a condition and a conclusion. Fig. 5 lists a sample about fuzzy rules applied for the Diabetes classification system.

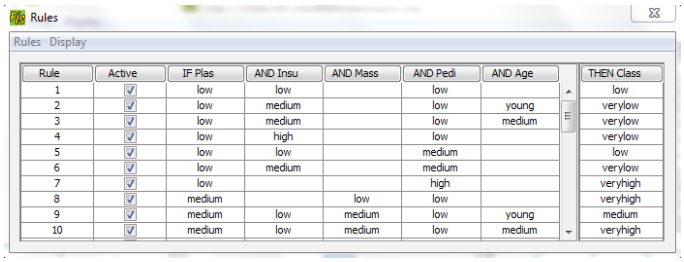


Fig. 5. Fuzzy rules for the Diabetes classification.

The fuzzy rules evaluations and combining results of individual rules are performed by fuzzy set operations [42]. The fuzzy sets operations are different from those on non-fuzzy sets. Let μA and μB the membership functions for the fuzzy sets A and B. Possible fuzzy operations for AND and OR operators on these sets are given in Table III. max and min are respectively the mostly used operations for OR and AND operators. For complement (NOT A) operation, (1- μA) is used for the fuzzy sets.

TABLE III. Fuzzy Set Operations

	OR (Union)		AND (Intersection)
MAX	Max $\{\mu_A(x), \mu_B(x)\}$	MIN	Min $\{\mu_A(x), \mu_B(x)\}$
ASUM	$\mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)$	PROD	$\mu_A(x)\mu_B(x)$
BSUM	Min $\{1, \mu_A(x) + \mu_B(x)\}$	BDIF	Max $\{0, \mu_A(x) + \mu_B(x)-1\}$

The result of each rule is evaluated. Then, the results should be combined to get a final result. This process is called inference. Results of individual rules can be combined in different ways. The maximum algorithm is generally used for accumulation.

3) Defuzzification

The aggregation result is converted into a crisp value for Diabetes Mellitus output. This is conducted by the defuzzification process [42]. A single number represents the fuzzy set outcome. Centroid method is used to convert the final combined fuzzy conclusion into a crisp value [20] as shown in Fig. 6.

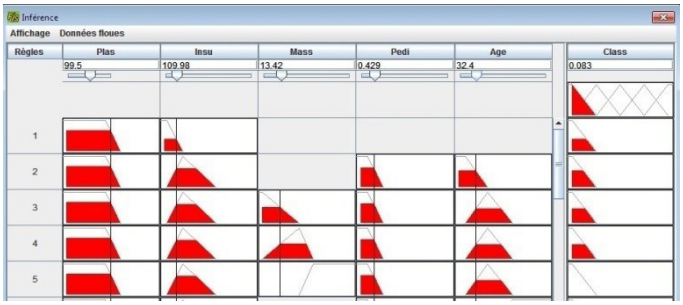


Fig. 6. Fispro interface for deffuzification.

4) Performance Evaluation

The proposed fuzzy inference mechanism has been implemented using Fispro. The data are taken from the Pima Indian Diabetes Dataset, the input variables are Plas, Ins, Mass, Pedi and Age; the output variable is DM (Diabetes Mellitus). To test the Fuzzy inference mechanism we took data including the age range 25-30 [19], [13]. The first experimental environment was constructed to evaluate the performance of the fuzzy classification task.

Well-defined MFs have been regulated and a fuzzy rule base has been produced. Then, the rules were extracted from a training data set of 768 instances including 500 no diabetics instances and 268 diabetics instances.

Specificity, sensitivity and accuracy are universal evaluation measures used to evaluate effectiveness of the data mining system, and particularly to compute how the assessment test is consistent and worthy. Sensitivity measure evaluates the diagnostic test correctly at detecting a positive disease. Specificity evaluates how the proportion of patients without disease can be properly ruled out. Accuracy can be concluded using specificity and sensitivity in the presence of prevalence. Accuracy is correctly defined by the diagnostic test by eliminating a given condition [18]:

$$Confusion\ Matrix = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix} \tag{5}$$

Where FP, FN, TP and TN are false positives, false negatives, true positives, true negatives respectively.

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{8}$$

Different results of a two-class prediction are listed in Table IV. Accuracy is the proportion of the total number of predictions that were correct.

TABLE IV. VARIOUS RESULTS OF A TWO-CLASS PREDICTION

Method	Accuracy (%)	Author
Our FDT framework for very young	91,67%	-
Fuzzy Expert System for Diagnosis of diabetes using Fuzzy Determination Mechanism	89,32%	Kalpana et al. (2011) [19][20]

D. Case Based Reasoning Using JColibri

Sometimes, we are confronted with similar problems that we try to resolve using our experience. CBR is a problem-solving approach based on the reuse by the analogy of past experience. We focus on the second step of the case-based reasoning: the retrieval phase. The aim of this step is to find cases that are similar to the given problem. In a previous work [14], a planning approach guided by case-based reasoning based on retrieval by decision tree has been proposed. In this paper, we propose a combination of methods to improve the retrieval step of this approach. To achieve this, we use fuzzy decision tree for the retrieval of similar cases.

The case-based reasoning consists in solving a new problem called target problem by using a whole of past solved problems and a source case is a past episode of solved problems. A case base consists of a set of source cases. Each source case has two parts: problem and solution. The problem part is described by a set of relevant characteristics called descriptors and the solution part is the plan to follow, it depends on the descriptors values. The process of case-based reasoning usually operates in five sequential phases: elaboration, retrieval, adaptation, revision and retain [16].

The elaboration step formalizes the problem using descriptors so as to build the target case. The cases retrieval step uses similarity measures to extract from the case base, the source cases which the problem part is similar to the target problem. The adaptation phase proposes one or more solutions to the target problem by adapting the proposed solutions. This step is often based on the use of knowledge in the field of the application. The objective of the revision step is to revise the proposed solutions by the previous phase according to certain rules and/or heuristics that depend on the field of application. This phase may be made by experts or in an automatic way. The last step, retain, is responsible for improving the experience of case-based reasoning system by enriching the case base with the new solved problems. Indeed, the resolved cases can be added to the case base and be used later to solve new problems.

1) Construction of the Case Base

We applied our approach to the Pima Indian Diabetes Database (PIDD), available on the official website of the UCI (machine learning repository). This database consists of a collection of medical diagnostic reports including 768 women, 268 who are diabetic and 500 non diabetic, over 21 years of age. Each case is described by 9 attributes. The solution part of cases represents the diagnosis of diabetes (0 non-diabetic patient, 1 otherwise). To establish the right diagnosis, the result of our fuzzy approach is a value between 0 and 1 (a percentage of having diabetes disease).

After discussing with experts in the field and consulting the site http://www.diabetes.co.uk/, we identified the type of diabetes and the best monitoring plan to be followed based on various factors including glucose, insulin, body mass, age.

In the jCOLIBRI tool, the database is given as an XML file. Thus, we need to convert the case base to an XML file. An excerpt of the case base in XML format is shown in Fig. 7.

```
create database travel;
use travel;
drop table travel;
create table travel(caseId INTEGER, Nbrenceinte INTEGER, Concentration INTEGER,
 Tension INTEGER, Epaisseur INTEGER, Insuline INTEGER, IndexMass DOUBLE,
FctPure DOUBLE, Age INTEGER, VarDiag INTEGER);
insert into travel values(1 , 6 , 148 , 72 , 35 , 0  , 33.6  , 0.627 ,  50  , 1 );
insert into travel values(2 , 1 , 85  , 66 , 29 , 0  , 26.6  , 0.351 ,  31  , 0 );
insert into travel values(3 , 8 , 183 , 64 , 0  , 0  , 23.3  , 0.672 ,  32  , 1 );
insert into travel values(4 , 1 , 89  , 66 , 23 , 94 , 28.1  , 0.167 ,  21  , 0 );
insert into travel values(5 , 0 , 137 , 40 , 35 , 168, 43.1  , 2.288 ,  33  , 1 );
```

Fig. 7. Extract from the case base in XML format.

To help the patient, we have proposed a monitoring plan and a printed notebook to take the blood sugar every day. The solution part refers to a plan of treatment Y that takes its values from the set of diabetes care plans C = {Plan1, Plan2, Plan3, Plan4}, where Plan1 = 'Ace-inhibitor therapy'; Plan2 = 'low fat diet'; Plan3 = 'determine exercise regime' and Plan4 = 'diabetes prevention program'.

2) Retrieval by Fuzzy Inference Mechanism

The studies cited above have led us to propose a new approach FDT4CR (Fuzzy Decision Tree For Cases Retrieval). The use of FDT for Retrieval optimizes the response time and avoids running whenever the k-nn algorithm. The objective of the fuzzy classification model consists in assigning a surveillance plan to the new case entry. Instead of using k-nn to find a plan, the retrieval by fuzzy decision tree is used in order to benefit from past experience. The classification model is responsible of the classification of the target case according to the descriptors values in order to find a solution.

The procedure of the fuzzy decision tree generation in FisPro needs both a Fuzzy Inference System configuration file and the PIDD data file. The prediction is based on one output only, even if the FIS has several outputs. This output is chosen by the user according to four possibilities:

1. A Fuzzy output, with the classification option.
2. A Fuzzy output, without the classification option.
3. A Crisp output, with the classification option.
4. A Crisp output, without the classification option

3) Performance Evaluation

The learning data are taken from the Pima Indian Diabetes Dataset and will serve to build the fuzzy decision tree. The second set of data including the age range 25-30, is used to test the validity of the classification procedure.

The experimental environment used JColibri platform to evaluate the performance of the proposed framework;

The proposed FDT approach analyzes the PIDD data and generates a corresponding plan based on Fuzzification, Fuzzy Inference Mechanism and Defuzzification for very young parameter [19].

The first experiment compared with the results of [19] is listed in Table V, indicating that the proposed approach automatically supports the analysis of the data.

TABLE V. RESULT OBTAINED FROM FISPRO

	Glucose (mg/dl)	INS (mu U/ml)	BMI (Kg/m2)	DPF	Age
Data	172	579	42.4	0.702	28
Our FDT4CR Crisp output with classification option	The Decision statement justifies that the possibility of suffering from diabetes for this person is medium (53,03%)				
Medical Practitioner	The person is diabetic				

To compare FDT4CR with other techniques, we applied k-nearest neighbours (k-NN), decision tree and our fuzzy decision tree for cases retrieval (FDT4CR) on the same case base. Table VI lists the different accuracy results. These results show that FDT4CR gets the highest accuracy for very young surveillance plans.

TABLE VI. Results of Experimentation

JColibri	Weka	Fispro
k-NN	Decision tree	Fuzzy DT
66%	73%	81%

IV. Conclusion and Perspectives

Multiple competing motivations allowed to define a Fuzzy model for case-based reasoning knowledge base systems. Effectively, we did not just want to experiment with a new approach of case indexing by FDT, but the aim is also to improve the modeling of uncertain and vague natural language concepts.

Planning guided by case-based reasoning is described through the following steps:

1. Construct the fuzzy Decision Tree by symbolic learning using Fispro platform and extract fuzzy rules;
2. Import fuzzy knowledge base using the JColibri platform.
3. Build the base of cases by JColibri;
4. Facing a new problem (target case), the CBR process using JColibri begins with the stage of retrieval. The retrieval is to find, among the source cases, the most similar cases to the target case. This step is treated using the FDT (Fispro). If the suggested solution does not satisfy the target case constraints, adaptation rules are used to adjust the solution. In the revision step, first the expert verifies the validity of the outcome. Then he alters or confirms the solution. Finally, in the retain step the system checks if the case base does not contain the new case before adding it [41].

k-nearest neighbors (k-NN) is typically used to calculate similarity in the retrieval step (cases indexing). We compared our Fuzzy Inference Mechanism with decision tree and k-nearest neighbours. We noticed that case indexing for the selection of a diabetes surveillance plan is considerably better with our Fuzzy Inference Mechanism (FDT4CR). Compared to retrieval by k-nn which is costly in computing time, retrieval by FDT provides one major advantage, it optimizes the response time. Furthermore, fuzzy reasoning combined with data mining for retrieval presents several advantages. First, it reduces the complexity of similarity calculation between individuals. Second, it presents an improved retrieval in the CBR process of JColibri. However, FDT4CR has some disadvantages related to the complexity of the domain of Diabetes Diagnosis. Building the cases-base from diabetic patient databases, the encoding of case base knowledge with standard medical files and the adaptation of vague data are examples of these challenges. The case structure used in FDT4CR is quite simple. The part problem of cases has been described by Body mass index (BMI), 2-hour serum insulin (INS), Plasma glucose concentration in 2-hours OGTT (Glucose), Age (Age) and Diabetes pedigree function (DPF). The solution part is defined by a monitoring plan corresponding to a crisp output, with the classification option.

In this paper, FDT4CR has been applied to improve the classification of diabetic patients. As future work, we suggest to expand FDT4CR approach to support the different phases of CBR life cycle. And to optimize response time and complexity we propose a Boolean modeling of induction and rule inference [37], [38], [39], [40].

As a result of the literature review [4], there is a specific need for more comprehensive improvements in clinical CBR. We plan to apply our FDT4CR approach in the emergency field following the proposed scheme [51].

References

[1] B. Campillo-Gimenez, W. Jouini, S. Bayat, and M. Cuggia, "Improving case-based reasoning systems by combining k-nearest neighbour algorithm with logistic regression in the prediction of patients registration on the renal transplant waiting list," *PLoS One*, vol. 8, no. 9, 2013.

[2] N. Choudhury and S. A. Begum, "A survey on case-based reasoning in medicine," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 7, no. 8, pp. 136-144, 2016.

[3] S. Montani, R. Bellazzi, L. Portinale, G. d'Annunzio, S. Fiocchi, and M. Stefanelli, "Diabetic patients management exploiting case-based reasoning techniques," *Computer Methods and Programs in Biomedicine*, vol 62, pp. 205–218, 2000.

[4] R Schmidt and L. Gierl, "Case-based Reasoning for Antibiotics Therapy Advice: An Investigation of Retrieval Algorithms and Prototypes," *Artificial Intelligence in Medicine*, vol. 23, no. 2, pp. 171-186, 2001.

[5] D. Doyle, P. Cunningham, and P. Walsh, "An evaluation of the usefulness of explanation in a CBR system for decision support in bronchiolitis treatment," *Computational Intelligence*, vol. 22, pp. 269-281, 2006.

[6] S. Montani, L. Portinale, G. Leonardi, R. Bellazzi, and R. Bellazzi, "Case-based retrieval to support the treatment of end stage renal failure patients," *Artificial Intelligence in Medicine*, vol.37, pp. 31-42, 2006.

[7] C. Marling, J. Shubrook, and F. Schwartz, "Case-based decision support for patients with type 1 diabetes on insulin pump therapy," *9th European Conference ECCBR 2008*, Springer-Verlag, Berlin, pp. 325-339, 2008.

[8] F. J. De Paz, S. Rodriguez, J. Bajo, and M. J. Corchado, "Case based reasoning as a decision support system for cancer diagnosis: A case study," *International Journal of Hybrid Intelligent Systems*, vol. 6,no. 2, pp. 97-110, 2009.

[9] R. Lopez de Mantaras, D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M.L. Maher, M.T. Cox, K. Forbus, M. Keane, A. Aamodt, and I. Watson, "Retrieval, reuse, revision and retention in case-based reasoning," *The Knowledge Engineering Review*, vol. 20, no. 3, pp. 215-240, 2006.

[10] G. Quellec, M. Lamard, L. Bekri, G. Cazuguel, B. Cochener, and C. Roux, "Multimedia medical case retrieval using decision trees," *International conference of the IEEE EMBS*, pp. 4536-4539, 2007.

[11] E.K. Burke, B.L. MacCarthy, S. Petrovic, and R. Qu, "Multiple-retrieval case based reasoning for course timetabling problems," *Journal of the Operational Research Society*, vol. 57, pp. 148-162, 2006.

[12] N. Choudhury and S. A. Begum, "The role of fuzzy logic in case-based reasoning: a survey," *Indian Journal of Computer Science and Engineering*, vol. 8, no. 3, 2017.

[13] D. K. Choubey, S. Paul, S. Kumar and S. Kumar, "Classification of Pima Indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection," *Communication and Computing Systems: Proceedings of the International Conference on Communication and Computing System*, pp. 451-455, 2017.

[14] S. Benbelkacem, B. Atmani, and A. Mansoul, "Planification guidée par raisonnement à base de cas et datamining : Remémoration des cas par arbre de décision,"*aIde à la Décision à tous les Etages Aide@EGC2012*, pp. 62-72, 2012.

[15] S. Guillaume and B. Charnomordic, "Learning interpretable fuzzy inference systems with FisPro," *Information Sciences*, vol. 181, no. 20, pp. 4409-4427, 2011.

[16] T. Hastie, R. Tibshirani, and J. Friefman, *The Elements of Statistical Learning*, Data Mining, Inference and Prediction, 2nd ed., Springer, New York, 2001.

[17] J. Bello-Tomas, P. Gonzalez-Calero, and B. Diaz-Agudo, "Jcolibri: An object-oriented framework for building CBR systems," *Advances in Case-Based Reasoning, 7th European Conference on Case-Based Reasoning*, pp. 32-46, 2004.

[18] P. K. N. Anooj. "Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules and decision tree rules," *Central European Journal of Computer Science*, vol. 1, no 4, pp. 482-498, 2011.

[19] M. Kalpana and A. S. Kumar, "Fuzzy expert system for diagnosis of

diabetes using fuzzy determination mechanism," *International Journal of Computer Science & Emerging Technologies*, vol. 2, no.6, pp. 354-361, 2011.

[20] M. Kalpana and A. S. Kumar, "Fuzzy expert system for diabetes using fuzzy verdict mechanism," *International Journal of Advanced Networking and Applications*, vol. 3, no 2, pp. 1128-1134, 2011.

[21] M. Huang, M. Chen, and S. Lee, "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis," *Expert Systems with Applications*, vol. 32, pp. 856-867, 2007.

[22] T. G. Houeland, "An efficient random decision tree algorithm for case-based reasoning systems," *24th International Florida Artificial Intelligence Research Society Conference*. May 2011.

[23] X. Boyen and L. Wehenkel, "Automatic induction of fuzzy decision trees and its application to power system security assessment," *Fuzzy Sets and Systems*, vol.102 , no.1, pp. 3-19, 1999.

[24] S. Begum, M. U. Ahmed, P. Funk, N. Xiong, and B. Von Schéele, "A case-based decision support system for individual stress diagnosis using fuzzy similarity matching," *Computational Intelligence*, vol. 25, pp. 180-195, 2009.

[25] F. Barrientos and G. Sainz, "Interpretable knowledge extraction from emergency call data based on fuzzy unsupervised decision tree," *Knowledge-based systems,* vol. 25, no.1, pp. 77-87, 2012.

[26] V. Levashenko and E. Zaitseva, "Fuzzy decision trees in medical decision making support system," *Proceedings of the IEEE Federated Confenrence on Computer Science & Information Systems,* pp.213-219, 2012.

[27] D. R. Adidela, D. G.Lavanya, S. G.Jaya, and A. R. Allam, "Application of fuzzy ID3 to predict diabetes," *International Journal of Advanced Computer and Mathematical Sciences*, vol. 3, no. 4, pp.541-545, 2012.

[28] A. G. Karegowda, V. Punya, M.A. Jayaram, and A.S. Manjunath, "Cascading K-means clustering and K-Nearest Neighbor classifier for categorization of diabetic patients," *International Journal of Engineering and Advanced Technology*, vol. 1, no. 3, pp. 147-151, 2012.

[29] A. G. Karegowda, V. Punya, M.A. Jayaram, and A.S. Manjunath, "Rule based classification for diabetic patients using cascaded K-means and decision tree C4.5," *International Journal of Computer Applications,* vol. 45, 2012.

[30] G. Kaur and A. Chhabra, "Improved J48 classification algorithm for the prediction of diabetes," *International Journal of Computer Applications*, vol. 98, no. 22, pp. 13-17, 2014.

[31] L. A. Zadeh. "Is there a need for fuzzy logic?" *Information Sciences*, vol. 178, no. 13, pp. 2751–2856, 2008.

[32] C. S. Lee and M. H. Wang, "A fuzzy expert system for diabetes decision support application," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol.41, no. 1, pp. 139-153, 2011.

[33] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees," *Wadsworth International Group*, Belmont CA, 1984.

[34] J.R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.

[35] R. Weber, "Fuzzy-id3: A class of methods for automatic knowledge acquisition," *2nd International conference on fuzzy logic and neural networks*, pp. 265–268, 1992.

[36] J. M. Alonso and L. Magdalena, "Generating understandable and accurate fuzzy rule-based systems in a java environment," *International Workshop on Fuzzy Logic and Applications*, Springer, Berlin, Heidelberg, pp. 212-219, 2011.

[37] B. Atmani and B. Beldjilali, "Knowledge discovery in database: Induction graph and cellular automaton," *Computing and Informatics*, vol. 26, no. 2, pp. 171-197, 2012.

[38] B. Atmani, S. Benbelkacem, and M. Benamina, "Planning by case-based reasoning based on fuzzy logic," *International Conference of Artificial Intelligence and Fuzzy Logic*, 2013.

[39] F. Barigou, B. Atmani, and B. Beldjilali, "Using a cellular automaton to extract medical information from clinical reports," *Journal of Information Processing Systems*, vol. 8, no 1, pp. 67-84, 2012.

[40] M. Benamina and B. Atmani, "Définition d'un modèle booléen de raisonnement flou adapté à la planification," *EGC*, pp. 553-554, 2012.

[41] S. Benbelkacem, B. Atmani, and M. Benamina, "Treatment tuberculosis retrieval using decision tree," *International Conference on Control Decision and Information Technologies*, 2013.

[42] S. Mokeddem and B. Atmani, "Assessment of Clinical Decision Support Systems for Predicting Coronary Heart Disease," *Fuzzy Systems: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, vol. 184, 2017.

[43] M. Rigla, G. García-Sáez, B. Pons, and M. E. Hernando, "Artificial intelligence methodologies and their application to diabetes," *Journal of Diabetes Science and Technology*, 2017.

[44] X. Blanco, S. Rodríguez, J.M. Corchado, and C. Zato, "Case-based reasoning applied to medical diagnosis and treatment," In: S. Omatu, J. Neves, J. Rodriguez, J. Paz Santana, and S. Gonzalez (Eds.) *Distributed Computing and Artificial Intelligence. Advances in Intelligent Systems and Computing*, vol. 217, pp. 137-146, Springer, 2013.

[45] S. El-Sappagh and M. M. Elmogy, "Medical case based reasoning frameworks: Current developments and future directions," *International Journal of Decision Support System Technology*, vol. 8, no 3, pp. 31-62, 2016.

[46] D. Azar and M. Bitar, "AI-based methods for predicting required insulin doses for diabetic patients," *International Journal of Artificial Intelligence*, vol. 13, no 1, pp. 8-24, 2015.

[47] S. Tahmasebian, M. Langarizadeh, M. Ghazisaeidi, and M. Mahdavi-Mazdeh, "Designing and implementation of fuzzy case-based reasoning system on android platform using electronic discharge summary of patients with chronic kidney diseases," *Acta Informatica Medica*, vol. 24, no 4, pp. 266-270, 2016.

[48] S. El-Sappagh and M. Elmogy, "A fuzzy ontology modeling for case base knowledge in diabetes mellitus domain," *Engineering Science and Technology, an International Journal*, vol. 20, no 3, pp. 1025-1040, 2017.

[49] S. Sharma and D. Mehrotra, "Building CBR based diagnosis system using jcolibri," *7th International Conference on Cloud Computing*, *Data Science & Engineering – Confluence*, pp. 634-638, 2017.

[50] M. K. Kiragu and P. W. Waiganjo, "Case based reasoning for treatment and management of diabetes," *International Journal of Computer Applications*, vol. 145, no 4, 2016.

[51] S. Benbelkacem, F. Kadri, S. Chaabane, and B. Atmani, "A data mining based approach to detect strain situations in hospital emergency department systems," *International Conference on Modeling, Optimization and Simulation*, Nancy, France, 2014.

[52] P. Kushwaha and R. R. Welekar, "Feature selection for image retrieval based on genetic algorithm," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no 2, pp. 16-21, 2016.

[53] A. Taibi and B. Atmani. "Combining Fuzzy AHP with GIS and Decision Rules for Industrial Site Selection." *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 4, no 6, pp. 60-69, 2017.

Mohammed Benamina

Mohammed Benamina is currently a PhD candidat at the University of Oran 1 Ahmed Ben Bella and affiliated researcher in Laboratoire d'Informatique d'Oran, Algeria. He received his Master degree in Computer Science in 2010 from University of Oran, Algeria. His research interests include data mining algorithms, expert systems and decision support systems.

Baghdad Atmani

Baghdad Atmani received his PhD in computer science from the University of Oran (Algeria) in 2007. He is currently a full professor in computer sciences. His interest field is artificial intelligence and machine learning. His research is based on knowledge representation, knowledge-based systems, CBR, data mining, expert systems, decision support systems and fuzzy logic. His research are guided and evaluated through various applications in the field of control systems, scheduling, production, maintenance, information retrieval, simulation, data integration and spatial data mining.

Sofia Benbelkacem

Sofia Benbelkacem is currently a PhD candidate at the University of Oran 1 Ahmed Ben Bella and affiliated researcher in Laboratoire d'Informatique d'Oran, Algeria. Her research interests include data mining, planning, case-based reasoning, medical decision support systems, machine learning.

QAM-DWT-SVD Based Watermarking Scheme for Medical Images

Habib Ayad* and Mohammed Khalil

Laboratory of Computer Sciences FST - Research Team: RTM, Hassan II University of Casablanca, B.P. 146 Mohammedia 20650 (Morocco)

Received 15 June 2017 | Accepted 30 November 2017 | Published 15 January 2018

Abstract

This paper presents a new semi-blind image watermarking system for medical applications. The new scheme utilizes Singular Value Decomposition (SVD) and Discrete Wavelet Transform (DWT) to embed a textual data into original medical images. In particular, text characters are encoded by a Quadrature Amplitude Modulation (QAM-16). In order to increase the security of the system and protect then the watermark from several attacks, the embedded data is submitted to Arnold Transform before inserting it into the host medical image. To evaluate the performances of the scheme, several medical images have been used in the experiments. Simulation results show that the proposed watermarking system ensures good imperceptibility and high robustness against several attacks.

Keywords

Watermarking, Medical Images, Information Security, Singular Value Decomposition, Discrete Wavelet Transforms, Quadrature Amplitude Modulation, Arnold Transform.

DOI: 10.9781/ijimai.2018.01.001

I. Introduction

With the widespread emergence of internet and computer applications, medical images can be shared between specialists and hospitals to determine suitable diagnostic procedures [6] and improve the understanding of a certain disease [9]. However, sharing medical images can lead data to be submitted to an act of tampering by unauthorized persons. As a result, a lot of worry has grown about the protection of authenticity, integrity and confidentiality of the content of medical images.

To avoid this kind of issues, image watermarking can be used as an effective and promising solution [4]. Image watermarking consists of hiding data into the original image without causing serious degradation of the perceptual quality [5]. In the inverse process, the watermark should be recovered from the watermarked image that can be disturbed by several attacks.

Image watermarking algorithms can be classified based on different views [19]. In terms of human perception, image watermarking can be grouped into visible and hidden methods. Visible watermarks such as logos are inserted into the corners of images for content or copyright protection. On the other side, hidden watermarks are imperceptible and are inserted on the unknown places in the host image. The similarity between the watermarked data and the original one should be high, in such a way that a simple user cannot make a difference. Image watermarking can also be categorized into fragile and robust, blind and non-blind.

In addition to above groupings, the digital image watermarking can be also classified into two groups according to the domain used for data embedding. The algorithms of the first group use the spatial domain for data embedding. In this case, the watermark is inserted by directly modifying the pixel values of the host image [12, 13]. In general, spatial methods are easy to implement but they are very fragile against attacks especially lossy compression. Moreover, the inserted data can be easily detected by computer programs since the watermark is embedded in the spatial domain of the image. The algorithms of the second group take advantage of transformation domains in which the watermark is embedded by modulating the coefficients in a transform domain such as discrete wavelet transform (DWT) [7], discrete cosine transform (DCT) [16], lifting wavelet transform (LWT) [8], integer wavelet transform (IWT) [2] and singular value decomposition (SVD) [3].

In general, transform domain methods are typically more robust to noise, attacks, common disturbances and compression compared to spatial transform algorithms. This is due to the fact that when image is inversely transformed, the watermark is distributed irregularly over the image. Furthermore, it is more difficult to detect the embedded data since the information contained in the watermark is distributed around the entire image. One of the limitations of transform methods is the capacity that is generally lower than that of spatial methods.

DWT based methods are among the most widely techniques used in image watermarking [19]. This is due to their good time-frequency features and directives that match well with the Human Visual System (HVS) [15]. Since the quality of medical images is very important for medical diagnosis, then the image quality must be preserved intact while the embedding capacity is increased [9].

The main goal of this paper is to propose an image watermarking scheme based on the discrete wavelet transform applied to a security context. We combine the DWT transform with SVD and QAM-16 to improve the performance of watermarking method. The main purpose of the proposed watermarking scheme is to increase the robustness without losing the imperceptibility of the embedded data.

* Corresponding author.
E-mail address: ayad.habib@gmail.com

The embedding process is carried out by inserting the watermark into the singular values of the DWT image. Specifically, the watermark is embedded by modifying the singular values of the DWT low frequency sub-band LL of the host image. The Arnold transform is used to increase the security of embedded data. At the extraction process, the operations are inversely done to extract the watermark from the watermarked image that can be disturbed by several distortions. The proposed scheme can be applied to several types of images especially medical ones that need higher quality for successful diagnosis.

The rest of the paper is organized as follows: in Section II, some useful and important preliminary ideas are discussed and then the proposed algorithm is introduced in Section III. Finally, simulation results are presented in Section IV followed by a discussion and conclusion in Section V.

II. Basic Concepts

In this work, Singular Value Decomposition (SVD), Discrete Wavelet Transform (DWT), Arnold Transform (AT) and Quadrature Amplitude Modulation (QAM) algorithms are used to design the proposed watermarking scheme. In the following subsections, a brief explanation of each algorithm is given. This section describes the overall basic concepts exploited in the proposed watermarking scheme.

- SVD is used to preserve significant amount of information of an image and makes the watermark more robust against attacks such as noise addition and scaling. The watermark can be then extracted effectively from the attacked watermarked image because of the special SVD properties.
- DWT transform is used to insert the watermark in imperceptible manner. The watermark bits are inserted in the significant coefficients sub-bands by considering the human visual system (HVS) characteristics.
- QAM technique is used to encode the character text before embedding it in the image.
- Arnold Transform is used to make the watermark more secure and protect the embedded data.

A. Singular Value Decomposition

Singular Value Decomposition (SVD) is an important technique of linear algebra that can be used to solve several mathematical problems. SVD is widely applied in many varieties of image processing applications such as image steganography, image watermarking, image compression and noise reduction [14].

From the perspective of linear algebra, a digital image can be viewed as a matrix composed of a number of non-negative scalars. The SVD of an image A with size $M \times N$ is represented mathematically as

$$A = USV^T \tag{1}$$

where U and V are the orthogonal matrices such that $UU^T = I_M$, $VV^T = I_N$ the columns of U are the orthonormal eigenvectors of AA^T, the columns of V are the orthonormal vectors of $A^T A$, and S is a diagonal matrix containing the square roots of the eigenvalues from U or V in descending order.

If r is the rank of the matrix A, then the elements of the diagonal matrix S satisfy the following relation:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq \lambda_{r+1} = \lambda_{r+1} = \cdots = \lambda_N = 0 \tag{2}$$

SVD has several interesting properties in image processing applications such as stability, proportionality, rotation and translation, etc. SVD can represent efficiently the intrinsic algebraic properties of an image. Indeed, the brightness of the image is specified by the singular values and corresponding pair of singular vectors reflect the geometry of the image.

The main goal of using SVD-based watermarking Techniques is to insert the data into the singular values by applying the SVD into whole or small blocks of the host image. Unlike the other watermarking methods, SVD can be utilized for non-square matrices because of its nonsymmetrical decomposition property. In general, SVD-based watermarking algorithms are robust against geometric attacks such as rotation, translation, noise addition and scaling. However, SVD still remains limited in comparison with transform domain methods. In order to increase the robustness, SVD can be combined with transform techniques such as DCT and DWT.

B. Discrete Wavelet Transform (DWT)

Discrete Wavelet Transform (DWT) is a multi-resolution mathematical tool that decomposes hierarchically an image and can be efficiently implemented using different digital filters. An image can be passed through high and low pass filters in order to be decomposed into several sub-bands with different resolutions. By applying DWT, the image is decomposed into four components namely LL, LH, HL and HH, corresponding to approximate, vertical, horizontal, and diagonal features respectively as illustrated in Fig. 1. The sub-band denoted by LL is approximately half of the original image. While LH and HL sub-bands contain the changes of edges or images along horizontal and vertical directions. Fig. 2 presents an example on 1-level DWT decomposition of Lena image that shows the four sub-bands LL, LH, HL an HH.

LL1	LH1
HL1	HH1

Fig. 1. The principle of 1-level DWT.

Fig. 2. 1-level DWT of Lena.

C. Arnold Transform (AT)

Arnold transform is a 2D chaotic map that is used to randomize a watermark matrix before embedding it into a cover image. Although there are many ways for scrambling, but in this paper, we will discuss only the Arnold transform [17] to increase the robustness and improve the security of the proposed watermarking scheme.

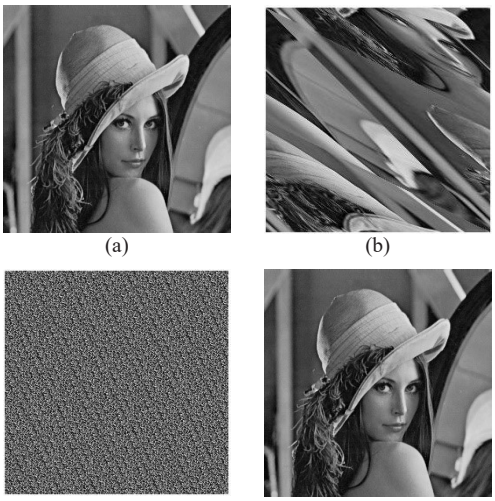
Arnold transform is an iterative process of moving the pixel position. Suppose that the original image is a $N \times N$ array and the coordinate of the pixel is $x, y \in \{0, 1, ..., N-1\}$. The generalized two dimension (2D)

Arnold transform is defined as:

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & k \\ l & kl+1 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ y_{n-1} \end{bmatrix} \bmod N \quad (3)$$

where x_n and y_n are the transformed coordinates corresponding to x_{n-1} and y_{n-1} after n iterations respectively, k and l are positive integers, and N represents the width or height of the square image processed.

Arnold transform is a periodic process, so the original position of $(x; y)$ coordinates gets back after T iterations. The factor T is called the transform period and depends on parameters k; l and n. These parameters will be used as secret keys in this paper. To recover the original image, periodicity is required. If the scrambling has performed n iterations; then the original image can be obtained by performing $T-n$ iterations. Fig. 3 illustrates an example of Arnold Transform into an image with different iterations.



(a) (b) (c) (d)

Fig. 3. Arnold transform with $k = l = 1$ (a) Lena 256 x 256. (b) AT with one iteration (c) AT with 10 iterations (d) AT with 192 iterations.

D. Quadrature Amplitude Modulation (QAM)

Quadrature Amplitude Modulation is a form of modulation that is a combination of phase modulation and amplitude modulation. A diversity of communication protocols implement quadrature amplitude modulation (QAM) such as digital video broadcast (DVB) and 802.11b wireless Ethernet (Wi-Fi).

For QAM-16, 4 bits are collected and mapped to one symbol from an alphabet with $2^4 = 16$ possibilities called constellations [11].

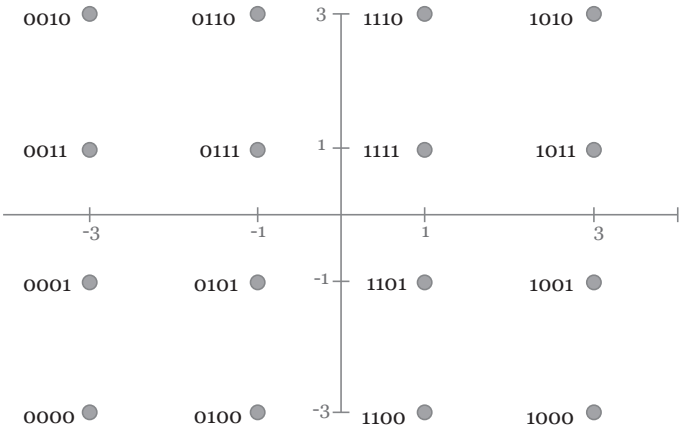


Fig. 4. Constellation diagram of QAM-16.

The symbols of QAM-16 alphabet are the complex numbers in the set $\{\pm 3 \pm 3j, \pm 3 \pm j, \pm 1 \pm 3j, \pm 1 \pm j\}$. The QAM-16 constellation is shown in Fig. 4. In this work, we use QAM-16 for encoding text characters to insert them into the host image.

Let z be a symbol in constellation QAM-16; $z = a + jb$ where $a, b \in \{-3, -1, 1, 3\}$ z can be also represented in the polar form as: $z = \rho e^{j\emptyset}$ where $\rho = \sqrt{a^2 + b^2}$ and $\emptyset = angle(z)$. Table I shows the correspondence between the binary codes and the complex symbols.

TABLE I. CORRESPONDECE BETWEEN BINARY CODES AND COMPLEX SYMBOLS

Binary code	Real part	Imaginary part	ρ	$\emptyset$ (degree)
0000	-3	-3	4.2426	-135
0001	-3	-1	3.1623	-161.5651
0010	-3	3	4.2426	135
0011	-3	1	3.1623	161.5651
0100	-1	-3	3.1623	-108.4349
0101	-1	-1	1.4142	-135
0110	-1	3	3.1623	108.4349
0111	-1	1	1.4142	135
1000	3	-3	4.2426	-45
1001	3	-1	3.1623	-18.4349
1010	3	3	4.2426	45
1011	3	1	3.1623	18.4349
1100	1	-3	3.1623	-71.5651
1101	1	-1	1.4142	-45
1110	1	3	3.1623	71.5651
1111	1	1	1.4142	45

Because the $\emptyset$ values will be used in a watermark matrix, the periodicity of sine and cosine functions can be used to change the negative values of $\emptyset$ by a positive values $(\emptyset + 360°)$ as shown in Table II.

TABLE II. CORRESPONDENCE BETWEEN BINARY CODES AND COMPLEX SYMBOLS WITH $\emptyset$>0

Binary code	Real part	Imaginary part	ρ	$\emptyset$ (degree)
0000	-3	-3	4.2426	225
0001	-3	-1	3.1623	198.4349
0010	-3	3	4.2426	135
0011	-3	1	3.1623	161.5651
0100	-1	-3	3.1623	251.5651
0101	-1	-1	1.4142	225
0110	-1	3	3.1623	108.4349
0111	-1	1	1.4142	135
1000	3	-3	4.2426	315
1001	3	-1	3.1623	341.5651
1010	3	3	4.2426	45
1011	3	1	3.1623	18.4349
1100	1	-3	3.1623	288.4349
1101	1	-1	1.4142	315
1110	1	3	3.1623	71.5651
1111	1	1	1.4142	45

TABLE III. CORRESPONDECE BETWEEN BYNARY CODES AND COMPLEX SYMBOLS SORTING ACCORDING TO ρ

Binary code	Real part	Imaginary part	ρ	Ø (degree)
0000	-3	-3	4.2426	225
0010	-3	3	4.2426	135
1000	3	-3	4.2426	315
1010	3	3	4.2426	45
0001	-3	-1	3.1623	198.4349
0011	-3	1	3.1623	161.5651
0100	-1	-3	3.1623	251.5651
0110	-1	3	3.1623	108.4349
1001	3	-1	3.1623	341.5651
1011	3	1	3.1623	18.4349
1100	1	-3	3.1623	288.4349
1110	1	3	3.1623	71.5651
0101	-1	-1	1.4142	225
0111	-1	1	1.4142	135
1101	1	-1	1.4142	315
1111	1	1	1.4142	45

Table III is obtained by sorting Table II according to the ρ column. It is remarkable that the angles {315, 225, 135, 45} corresponding to magnitudes 4.2426 or 1.4142 as the rest of the angles correspond to the magnitude 3.1623. In this work, we use this correspondence to extract QAM-16 symbols. The symbols extraction decision is made in two stages: firstly the extraction of Ø and secondly the extraction of ρ based on correspondence between ρ and Ø. For example, if the extraction of Ø belongs to {341.5651; 251.5651; 288.4349; 198.4349; 161.5651; 108.4349; 71.5651; 18.4349} ρ is automatically equal to 3.1623 and if the extraction of Ø belongs to {315; 225; 135; 45}. Finally, we make a decision of ρ on 4.2426 and 1.4142.

III. THE PROPOSED APPROACH

The overall system of our proposed approach is illustrated in Fig. 5. First, we convert the electronic patient record (ERP) text into a watermark matrix using the QAM-16, then Arnold Transform is applied to the watermark matrix and the watermark is scrambled. The parameters of Arnold Transform are used as a key to increase the security of the watermark. The scrambled watermark matrix is embedded then into the host image. To recover the secret data, the scrambled watermark matrix is extracted from the watermarked image using the extraction procedure. Finally, the inverse Arnold Transform and QAM-16 procedure are successively applied to retrieve the original ERP data. The proposed algorithm consists of six main steps:

1. Conversion ERP text into a watermark matrix.
2. Scrambling the watermark matrix by Arnold Transform.
3. Embedding process.
4. Extraction process.
5. Inverse Arnold Transform.
6. Conversion watermark matrix to the original ERP.

A. Watermark Matrix

In this work, the watermark that is embedded in the original medical image is a matrix that is generated from the EPR. The characters of the EPR text are grouped into a matrix of size $2^m \times 2^m$. For example, an EPR of 1024 characters is represented in a matrix of size $2^5 \times 2^5$. Then, the ASCII code for each character i is converted to 8-bit binary code.

The QAM-16 is applied to the first 4 bits and the last 4 bits to obtain two pairs $(\rho_1^i, \emptyset_1^i)$ and $(\rho_2^i, \emptyset_2^i)$ which are grouped into a matrix of size 2×2 as shown in Fig. 6.

To ensure that the watermark matrix elements are between 0 and 1 we use $\frac{\rho}{10}$ rather than the magnitude ρ and $\frac{\emptyset}{10}$ rather than the angle Ø. After this process, the resulted watermark matrix is obtained by replacing each character by a 2×2 matrix that is composed by ρ and Ø which gives a watermark matrix of size $2^{m+1} \times 2^{m+1}$.

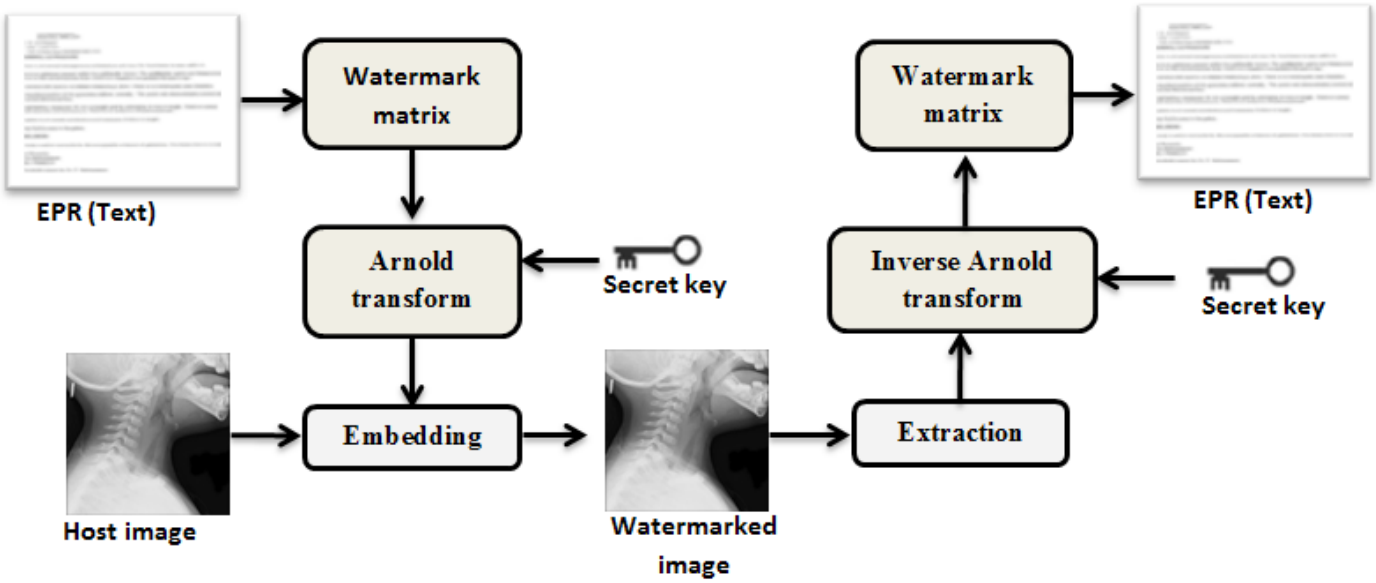


Fig. 5. The proposed watermarking scheme.

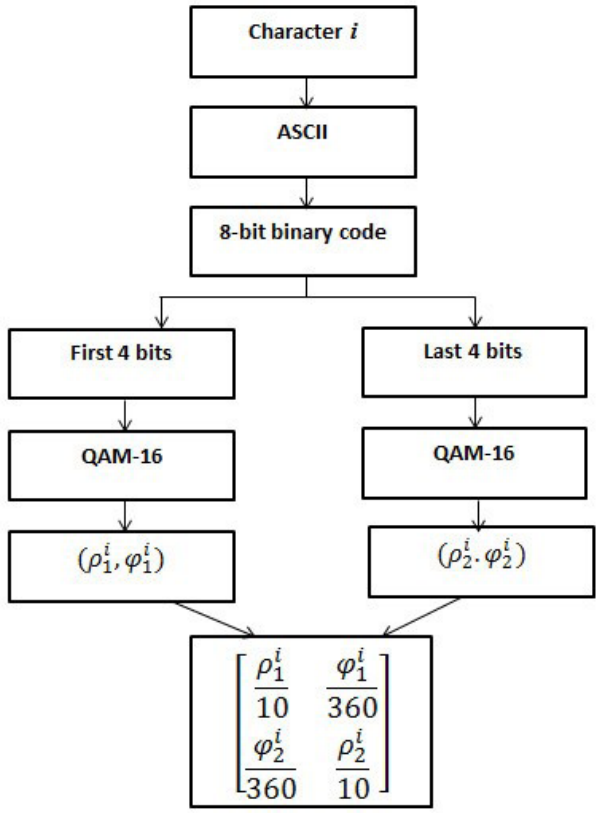


Fig.6. The characters conversion process.

Example

Let *tex*="Farabi2 hospital", the size of *tex* is 16 characters that can be grouped into $2^2 \times 2^2$ matrix by replacing each character to its ASCII code. The resulted matrix is:

$$\begin{bmatrix} 70 & 98 & 104 & 105 \\ 97 & 105 & 111 & 116 \\ 114 & 58 & 115 & 97 \\ 97 & 32 & 112 & 108 \end{bmatrix}$$

By applying the process as shown in Fig. 6, we convert each ASCII code to a 2×2 matrix:

$$F \underset{ASCII}{\rightarrow} 70 \underset{Binary}{\rightarrow} \overbrace{0100}^{QAM-16} \overbrace{0110}^{QAM-16}$$

$$\Rightarrow \begin{bmatrix} \frac{\rho_1^F}{3.1623} & \frac{\phi_1^F}{251.5651} \\ 10 & 360 \\ \frac{\phi_2^F}{108.4349} & \frac{\rho_2^F}{3.1623} \\ 360 & 10 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.3162 & 0.6988 \\ 0.3012 & 0.3162 \end{bmatrix}$$

Then, the resulted watermark $2^3 \times 2^3$ matrix is:

$$\begin{bmatrix} 0.3162 & 0.6988 & 0.3162 & 0.3012 & 0.3162 & 0.3012 & 0.3162 & 0.3012 \\ 0.3012 & 0.3162 & 0.3750 & 0.4243 & 0.8750 & 0.4243 & 0.9488 & 0.3162 \\ 0.3162 & 0.3012 & 0.3162 & 0.3012 & 0.3162 & 0.3012 & 0.1414 & 0.3750 \\ 0.5512 & 0.3162 & 0.9488 & 0.3162 & 0.1250 & 0.1414 & 0.6988 & 0.3162 \\ 0.1414 & 0.3750 & 0.3162 & 0.4488 & 0.1414 & 0.3750 & 0.3162 & 0.3012 \\ 0.3750 & 0.4243 & 0.3750 & 0.4243 & 0.4488 & 0.3162 & 0.5512 & 0.3162 \\ 0.3162 & 0.3012 & 0.4243 & 0.3750 & 0.1414 & 0.3750 & 0.3162 & 0.3012 \\ 0.5512 & 0.3162 & 0.6250 & 0.4243 & 0.6250 & 0.4243 & 0.8012 & 0.3162 \end{bmatrix}$$

B. Watermark Embedding Process

The proposed watermark embedding process is described as follows:

Input: Original Image I of size $2^J \times 2^J$, Watermark matrix of size $2^K \times 2^K$

Output: Watermarked Image I_W

1. DWT level $l = J - K$
2. Apply l-level DWT on the original image I to produce four sub-bands LL_l, LH_l, HL_l and HH_l,
3. Perform SVD operation for low-pass sub-band LL_l

$$LL_l = U_L S_L V_L^T$$

4. Modify S_L, the singular values of the sub-band LL_l, by adding a watermark matrix, with the scaling factor α.

$$S_2 = S_L + \alpha W$$

5. Compute SVD of S_2

$$S_2 = U_2 S_3 V_2^T$$

6. Using S_3 to compute a modified low-pass sub-band LL'

$$LL' = U_L S_3 V_L^T$$

7. Compute the watermarked image I_W by applying the inverse DWT on LL' .LH_l, HL_l and HH_l.

C. Watermark Extraction Process

In general, the extraction process can be completed by reversing the steps of the embedding process. In watermark extraction, an eventually distorted watermark W can be extracted from the eventually distorted watermarked image I_W by effectively reversing the above watermark embedding steps. The process of watermark extraction can be described as follows:

Input: Watermarked Image I_W

Output: Watermark matrix

1. Apply l-levels DWT on the watermarked image to produce four sub-bands LL_w, LH_w, HL_w and HH_w
2. Compute SVD of low-pass sub-band LL_w

$$LL_w = U_w S_w V_w^T$$

3. Compute $\hat{S}$ using left and right singular vectors U_2 and V_2 in step 5 in watermark-embedding algorithm

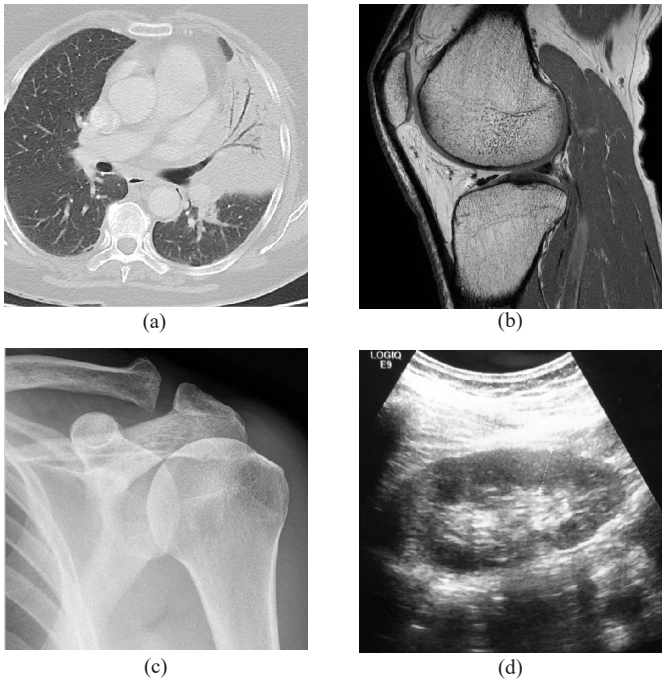
$$\hat{S} = U_2 S_w V_2^T$$

4. Extract the watermark matrix $\widehat{W}$

$$\widehat{W} = \frac{\hat{S} - S_L}{\alpha}$$

IV. Experimental Results

To evaluate the performances of our proposed scheme, we have applied the embedding algorithm to a database of 100 grey scale medical images of four modalities: X-ray, Ultrasound, MRI and CT. All test images are 512 × 512 pixels. An example of these medical images is illustrated in Fig. 7.



(a) (b) (c) (d)

Fig. 7. (a) CT, (b) MRI, (c) X-Ray and (d) Ultrasound.

A. Quality Measures

Peak Signal to-Noise Ratio (PSNR) is one of the most commonly used measures of imperceptibility between an original image I of size M × N and a watermarked image I_W, which is computed by:

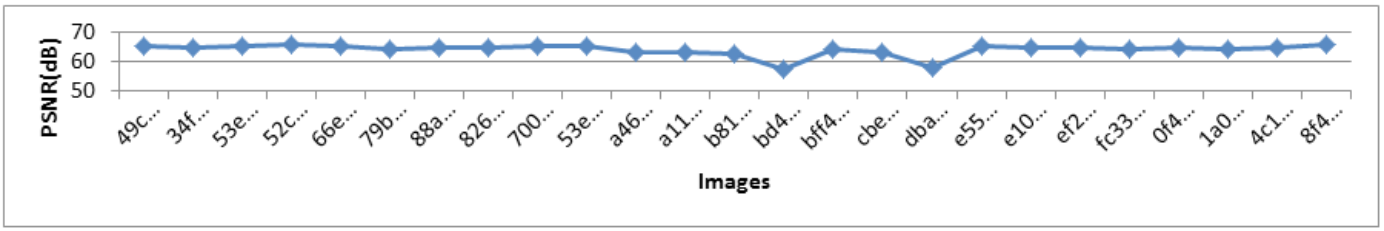
$$PSNR = 10\log\left(\frac{max(I(i,j)^2)}{MSE}\right) \tag{4}$$

$$MSE = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left(\frac{\left(I(i,j) - I_w(i,j)\right)^2}{NM}\right) \tag{5}$$

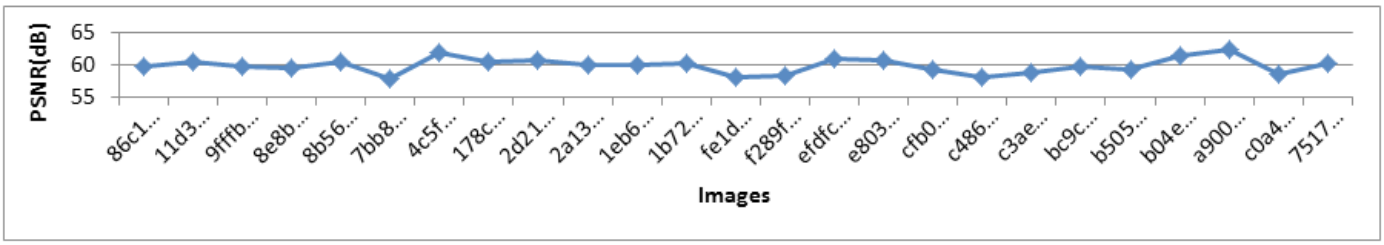
Structural Similarity Measure (SSIM) is another perceptual metric that quantifies the watermarked medical images quality. Image quality evaluation based on SSIM is based on the fact that the HVS is highly adapted to extract structural information from the viewing field. SSIM metric is ideal for testing of similarities in medical images because it focuses on local rather than global image similarity.

$$SSIM(A,B) = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)} \tag{6}$$

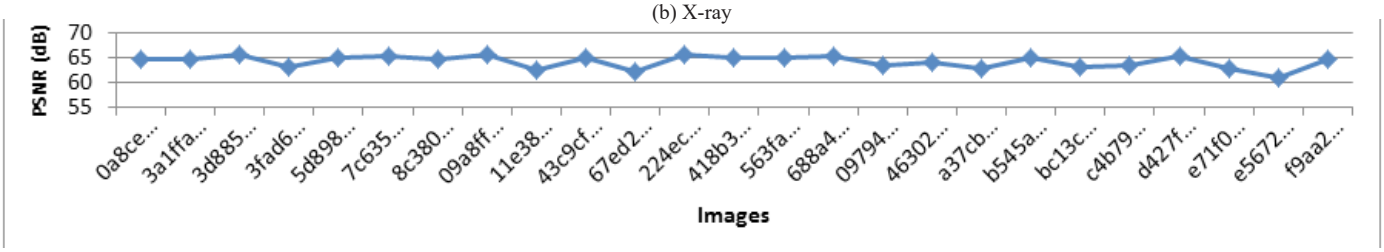
where μ_A and μ_B are respectively the averages of A and B. σ_A^2 and σ_B^2 are respectively the variances of A and B. $c_1(c_1 = k_1L)^2)$ and



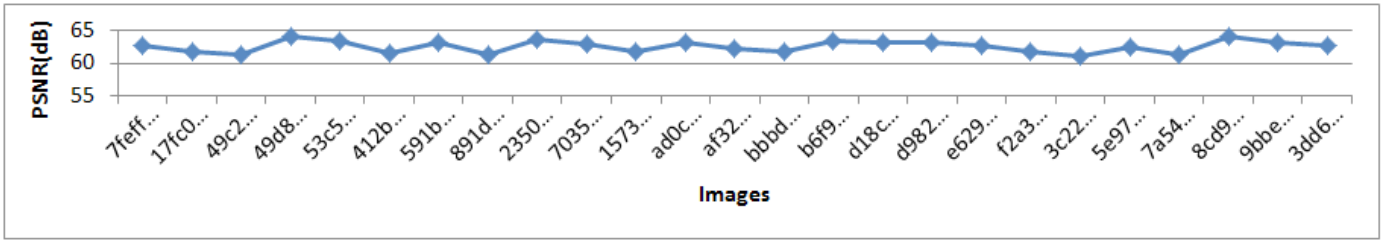
(a) CT



(b) X-ray



(c) MRI



(d) Ultrasound

Fig. 8. Curve of Peak Signal to Noise Ratio (PSNR) in dB for different medical images modalities (a) CT, (b) X-Ray, (c) MRI and (d) Ultrasound.

$c_2(c_2 = k_2L)^2$) are two variables to stabilize the division with a weak denominator. L is also the dynamic range of the pixel-values. k_1 and k_2 have default values as *0.01* and *0.03*, respectively [18].

In order to compare the similarities between the original ERP text and the extracted ERP text, we define the character error rate in percentage (CER) as follows:

$$CER = \frac{NEC}{TC} \tag{7}$$

where:

- NEC: number of erroneous characters
- TC: total number of characters in a ERP text

B. Imperceptibility Medical Images

In this subsection, we investigate the imperceptibility of the watermark. The PSNR is used to measure the similarity between the original image and the watermarked image. When the PSNR value is higher than 30 dB, it will be difficult to find the difference between the original image and the watermarked image on human's eyes [10].

Fig. 8 (a)-(d) show the PSNR for different medical images modalities by embedding an ERP text of 2048 characters (2kb). The PSNR values reached for these 100 images are between 57.5306 dB and 66.0223 dB, which demonstrate that the proposed method achieves good imperceptibility.

C. Similarity for Different Embedded Data

Fig. 9 represents the PSNR values for different number of embedded characters from 128 (2^7) characters to 16384 (2^{14}) characters on four medical images CT, X-Ray, MRI and Ultrasound. We can notice that the value of PSNR decreases when number of embedded characters increases. On the other side, the PSNR values are greater than 45 dB, which mean that embedded data is undetectable according to the human visual perception.

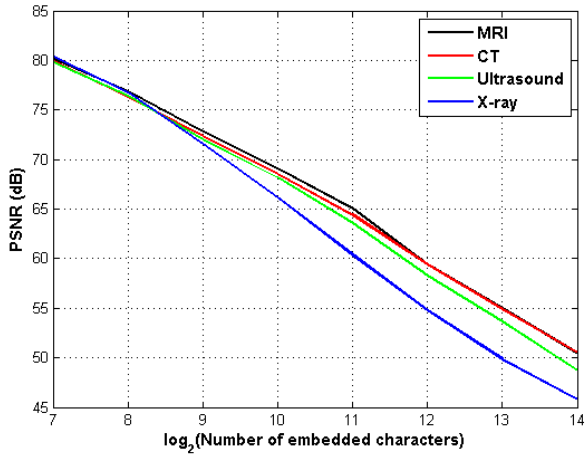


Fig. 9. PSNR values for different images and embedded data.

In Fig. 10, we evaluate the invisibility of embedded data with consideration to the properties of the human eye using Structural Similarity Metric Index (SSIM). For all images, SSIM is close to 1 (SSIM >0.999983) for embedded data less than 1024 characters (2^{10}). SSIM decreases when embedded data increases but remains greater than 0.998069 for maximal data embedding.

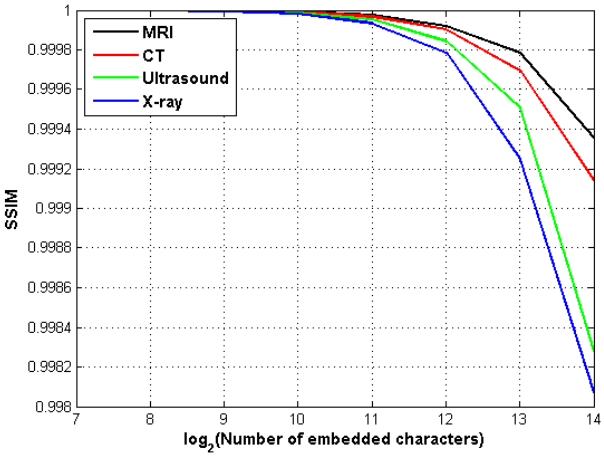


Fig. 10. SSIM values for different images and embedded data.

D. Comparison to Existing Scheme

To prove the effectiveness of the proposed scheme, our method is compared with another semi-blind scheme [1]. The watermarked image is attacked by applying salt & pepper noise and Gaussian noise in order to investigate the robustness.

In Table IV, we analyzed the variation of character error rate (CER) against varying density of Salt and Pepper for some images. For the proposed technique, we observe that the value of Character Error Rate (CER) is equal to zero for all density of Salt and Pepper noise, which means that the extraction of ERP text is done without any error and indicating the highly robust nature of our technique against Salt and Pepper noise. However, in the case of the Sleit's method, we observe that the Character Error Rate (CER) value increases as the density of noise increases, and that causes a deterioration of detection performance.

TABLE IV. VARIATION OF CER ON DIFFERENT VALUES OF SALT & PEPPERS ATTACK

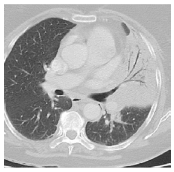
		Salt & Peppers noise density		
I_W	**Methods**	10^{-6}	10^{-5}	10^{-4}
	**Proposed**	0	0	0
	Sleit	0	0	4.6294
	**Proposed**	0	0	0
	Sleit	0	0	3.2109
	**Proposed**	0	0	0
	Sleit	2.8645	4.7582	32.5897
	**Proposed**	0	0	0
	Sleit	25.8245	32.5638	63.8453

TABLE V. Variation of CER on Different Values of Gaussian Attack

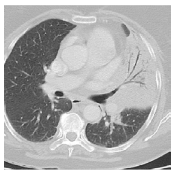
I_w	Methods	Gaussian noise variance 10^{-6}	10^{-5}	10^{-4}
	**Proposed**	0	0	0
	Sleit	0	0	5.7835
	**Proposed**	0	0	0
	Sleit	0	0	6.2578
	**Proposed**	0	0	0
	Sleit	6.1275	10.4852	27.8906
	**Proposed**	0	0	0
	Sleit	56.7582	61.2878	73.9245

Table V shows the robustness against Gaussian noise with different variances. We can observe that the CER value of Sleit's method is significantly higher than our method. It is also clearly that the CER value of the proposed method is equal to zero for all Gaussian noise values, which ensures an extraction without any error.

To further validate the robustness of the proposed scheme we compare it with the scheme presented in [1]. The considered disturbances are compression, low-pass filter (median filter) and speckle noise. From Table VI, it is obvious that our proposed system outperforms the algorithm introduced in [1] for all disturbances and for all test images. We notice also the compression deteriorates highly the watermark by using Sleit scheme [1] especially for MRI and CT images. Table VI indicates also that the robustness of our method against compression is much higher than the method in [1] and it guarantees the quality of the images with that mentioned behavior.

V. Conclusion

This paper presents a new watermarking scheme for medical images. The proposed scheme is based on a combination of DWT and SVD to embed the watermark in a transparent manner and extracted it with high fidelity. QAM-16 was also used to encode text characters and insert them into the host image. Overall, the proposed scheme demonstrates a good trade-off between of imperceptibility, robustness, and capacity as compared to state of the art methods. Our experimental results show the effectiveness of combination of wavelet algorithm with SVD technique as compared to non-hybrid SVD or DWT methods in terms of PSNR and SSIM.

In the future work, we will aim to overcome the limitation of the proposed semi-blind watermarking by extending it to the blind context. In particular, we will focus in reversible image watermarking by evaluating the performance with much more image types including not only medical image but also texture and biometric images.

References

[1] A. Sleit, R. Etoom, S. Abusharkh and Y. Khero. 2012. An enhanced semi-blind DWT-SVD-based watermarking technique for digital images. The Imaging Science Journal 60, 1 (2012).

[2] Muhammad Arsalan, Sana Ambreen Malik, and Asifullah Khan. 2012. Intelligent Reversible Watermarking in Integer Wavelet Domain for Medical Images. J. Syst. So‡w. 85, 4 (April 2012), 883–894. DOI:http://dx.doi.org/10.1016/j.jss.2011.11.005

[3] Veysel Aslantas. 2009. An optimal robust digital image watermarking based on SVD using differential evolution algorithm. Optics communications 282 (2009), 769–777.

[4] W. Bender, D. Gruhl, N. Morimoto, and Aiguo Lu. 1996. Techniques for Data Hiding. IBM Syst. J. 35, 3-4 (Sept. 1996), 313–336. DOI:http://dx.doi.org/10.1147/ sj.353.0313

[5] Ingemar J. Cox, Joe Kilian, Frank Œomson Leighton, and Talal Shamoon. 1997. Secure spread spectrum watermarking for multimedia. IEEE Trans. Image Processing 6, 12 (1997), 1673–1687. DOI:http://dx.doi.org/10.1109/83.650120

[6] Sudeb Das and Malay Kumar Kundu. 2012. Effective Management of Medical Information (Trough A Novel Blind Watermarking Technique. J. Medical Systems 36, 5 (2012), 3339–3351. DOI:http://dx.doi.org/10.1007/s10916-012-9827-1

[7] Pegah Fakhari, Ehsan Vahedi, and Caro Lucas. 2011. Protecting patient privacy from unauthorized release of medical images using a bio-inspired wavelet-based watermarking approach. Digital Signal Processing 21, 3 (2011), 433–446. DOI:http://dx.doi.org/10.1016/j.dsp.2011.01.014

[8] K. Ghaderi, F. Akhlaghian, and P. Moradi. 2013. A new robust semi-blind digital image watermarking approach based on LWT-SVD and fractal images. 21st Iranian Conference on Electrical Engineering (ICEE) (2013), 1–5.

[9] Aggeliki Giakoumaki, Sotiris Pavlopoulos, and Dimitris Koutsouris. 2006. Multiple Image Watermarking Applied to Health Information Management. IEEE Trans. Information Technology in Biomedicine 10, 4 (2006), 722–732. DOI: http://dx.doi.org/10.1109/TITB.2006.875655

[10] J. Han and X. Zhao. 2015. An Adaptive Gray Scale Watermarking Method in Wavelet Domain. International Journal of Security and Its Applications 9, 10 (2015), 103–114.

[11] L Hanzo, S Xin Ng, WT Webb, and T Keller. 2004. .Quadrature amplitude modulation: From basics to adaptive trellis-coded, turbo-equalised and spacetime coded OFDM, CDMA and MC-CDMA systems. (2004).

TABLE VI. CER Results Under Various Attacks

Attack	Proposed scheme				**Sleit scheme**			
	X-ray	Ultrasound	MRI	CT	X-ray	Ultrasound	MRI	CT
Compression (Q=80%)	1.1254	1.6113	9.6845	8.2651	7.4852	11.2635	35.9475	31.4763
3×3 **Median filter**	0.4882	1.2695	1.1718	3.5644	5.8492	9.2383	8.3962	12.3549
Speckle noise	0.3662	0.4882	0.5859	0.7324	4.7482	15.2375	22.6534	8.4742

[12] P. S. Huang, C. S. Chiang, C. P. Chang, and T. M. Tu. 2005. Robust spatial watermarking technique for colour images via direct saturation adjustment. IEEE Proceedings - Vision, Image and Signal Processing 152, 5 (2005), 561–574.

[13] C. H. Lin, D. Y. Chan, H. Su, and W. S. Hsieh. 2006. Histogram-oriented watermarking algorithm: colour image watermarking scheme robust against geometric attacks and signal processing. IEE Proceedings - Vision, Image and Signal Processing 153, 4 (2006), 483–492.

[14] M. Moonen and B. De Moor. 1995. SVD and Signal Processing, III. Elsevier Science(1995).

[15] A. Adhipathi Reddy and Biswanath N. Chatterji. 2005. A new wavelet based logo-watermarking scheme. Pattern Recognition Letters 26, 7 (2005), 1019–1027. DOI:http://dx.doi.org/10.1016/j.patrec.2004.09.047

[16] ASLANTAS Veysel, OZER Saban, and OZTURK Serkan. 2009. Improving the performance of DCT-based fragile watermarking using intelligent optimization algorithms. Optics communications 282, 14 (2009), 2806–2817.

[17] L. Wu, W. Deng, J. Zhang, and D. He. 2009. Arnold transformation algorithm and anti-Arnold transformation algorithm. Proc. of 1st International Conference on Information Science and Engineering (2009), 1164–1167.

[18] R. Karakış, İ. Güler, İ. Çapraz and E. Bilir, A novel fuzzy logic-based image steganography method to ensure medical data security, Computers in Biology and Medicine, 67,(2015), 172–183.

[19] Mousavi SM, Naghsh A, Abu-Bakar SA. Watermarking techniques used in medical images: a survey. J Digit Imaging, 27(6), (2014), 714-29.

Habib Ayad

Habib Ayad is a Professor-Researcher in Computer Sciences at Hassan II University of Casablanca, Morocco; he received his PhD in Computer science from Cadi Ayyad University, Marrakech-Morocco in 2013. His research interests include digital image watermarking and content-based image retrieval.

Mohammed Khalil

Mohammed Khalil received his PhD in Computer Science from Department of Informatics, Faculty of Sciences and Techniques, Mohammedia, Morocco in 2014. He has published four journal papers, 12 communications in international conferences (including EUSIPCO, ICT, ICISP and ISCCSP) and five communications in national conferences. His research interests include digital image and audio watermarking.

Issues of Visual Search Methods in Digital Repositories

Paulo Gaona-García[1]*, Carlos Montenegro-Marín[1], Elvis Gaona-García[1], Adriana Gómez-Acosta[2], Yusef Hassan-Montero[3]

[1] Universidad Distrital Francisco José de Caldas, Bogotá (Colombia)
[2] Fundación Universitaria San Mateo, Bogotá (Colombia)
[3] Universidad Internacional de la Rioja, Madrid (Spain)

Received 22 March 2018 | Accepted 23 October 2018 | Published 26 October 2018

Abstract

Repositories are important infrastructures which allow the dissemination of large collections of digital resources hosted in museums, libraries, academic institutions or specialized documentation centers. However, there are nowadays several limitations associated with irrelevant search results based on a knowledge area. Some studies have highlighted the major role of information visualization strategies based on Simple Knowledge Organization Systems (SKOS) so as to mitigate such difficulties. The main goal of this article is to present recommendations using information visualization based on SKOS for the development of navigational search interfaces in digital repositories focused on learning process. We use card sorting as methodology in order to obtain qualitative results in our study. As preliminary results we found that taxonomies in visual search engines improve the access to large collections of digital resources based on SKOS, but it depends on the design of taxonomy concepts defined in digital repositories. Finally, it is recommended that the creators of repositories focus their efforts on define levels of relationship and partnership between digital resources using knowledge representation structures like thesauri or ontologies; work with usable visualization interfaces like tree, radial or icicle; and link relevant metadata fields with the navigation structure.

Keywords

Search Interfaces, Metadata, Digital Repositories, Taxonomies, Information Visualization.

DOI: 10.9781/ijimai.2018.10.005

I. Introduction

CURRENTLY, the treatment of information poses major challenges not only because of the amount of information which needs to be handled but also because of the variety of formats in which it is presented. In this sense, recent studies conducted on the design and implementation of digital repositories indicate that a large majority include knowledge classification schemes [1], based on instruments such as ontologies and thesauri. The inclusion of these instruments allows experts to perform management and organization activities of thousands of digital resources. However, for their design and implementation with the aim of using them in a learning process, there are several considerations that should be taken into account and which are associated with: i) redundant information [2], ii) terminological use [3], iii) use of knowledge representation schemes [4, 5], and iv) lack of understanding shown by users to know how to use the representation scheme [6, 7].

To some extent, the difficulties faced by some knowledge representation schemes are factors which may limit their inclusion in the development of search interfaces in digital repositories. However, information visualization plays an important role in facilitating the use of such classification instruments [8]. Currently, information visualization is considered as one of the fastest growing strategies given in various knowledge areas in recent years. This is reflected in the communicability it offers to understand complex information that usually occurs without a classification and without any specific format. Therefore, on the front of design and development of digital repositories, they are presented as alternatives for gaining access to a collection of digital resources. Such initiatives have favoured the use of these tools and at the same time have become one of the most widely accepted alternatives to perform search processes on digital resources. However, the high consumption of resources which are necessary for their analysis, design and implementation to carry out their application in these types of fields make them to be undervalued.

The following article aims to present a number of factors and considerations for the development of search methods based on open digital repositories SKOS. Currently, within processes of search and navigation resources, the use of SKOS is very limited. However, there are several factors which can be carried out so as to give access to learning objects on the basis of knowledge areas. In order to do this, the following article is divided as follows: Section II provides a brief description of some of the most relevant visualization techniques. Subsequently, section III identifies the benefits, implications and limitations encountered in the use of visualization information. Section IV presents the study methodology. Section V shows an analysis of variables and evaluation criteria associated with visualization techniques, knowledge representation schemes and digital repositories to facilitate both the classification and visualization of specific themes. Finally, results of a case study are presented, applying these solutions to highlight critical success factors that these strategies offer to facilitate access to a collection of digital resources in repositories.

* Corresponding author.
E-mail address: pagaonag@udistrital.edu.co

II. Theoretical Framework and Related Studies

In this section, the importance of information in search processes and access to digital resources is stressed. To do this, the importance of visualization of information on the development of search interfaces is first presented, followed by studies related to assessment criteria and associated with the area of usability studies. Finally, the first works on visualization information related to the field of digital repositories are mentioned.

A. Works Associated with Digital Repositories

In order to facilitate management and administration processes, learning objects are grouped and stored in digital repositories. Within these repositories, two types can be identified: i) Repositories containing learning objects and their metadata [9] -learning objects and their descriptors are within the same system [10]-, and ii) Repositories containing only metadata [11] -they contain only descriptors and the access to the object is made through a reference to its physical location found in another system or object repository.

Considering the aforementioned, a search for learning objects in any of these repositories could provide a long list of results. Therefore, if indicators to evaluate the quality of recorded information are not defined, the search process of learning objects may be an activity which requires a great deal of wasted time and effort by the user [12].

The use of digital repositories is centred on performing storage-related educational materials in order to optimize their administration, management and search processes. However, the large volume of digital resources has generated a number of limitations, specifically those related to the use of repositories for gaining access to relevant instructional materials [8]. Such deficiencies have favoured the development and use of different alternatives associated with the implementation of enriched languages and knowledge representation schemes to execute classification activities, categorization and content management. The implementation of these strategies has generated the combination of a solid technological structure which is linked to a series of strategies of semantic enrichment from the use of knowledge representation schemes. To some extent, these solutions facilitate administration and management activities created by developers and repository creators [13]. However, for a conventional user (student and / or teachers), the use of such educational repositories is not an easy task, since they lack access strategies and mechanisms of conventional searches (textual, Boolean) working from the use of their interfaces [14]. This is a key factor which may make the learning process difficult, and therefore may generate a progressive abandonment of such tools.

B. Works Associated with Search Interfaces

In addition to language and knowledge representation schemes, repositories also provide alternative access to perform a search process by using visual interfaces, although some of these do not have them to facilitate such processes. Moreover, previous research has found that some search interfaces do not conform to the users' needs altogether [15]. Oftentimes, results displayed are not relevant according to search criteria defined by users [16]. Navigation problems have been identified when users want to check previously viewed records [17]. Authors like [18] found a series of problems related to interface design of an institutional repository in Korea. The study showed that subjects of interest were not sufficiently visible because navigation menus were too small and dark.

Other studies have revealed limitations to combine navigation strategies and search methods [19], in which interfaces do not allow to display (at first glance) the deployment of available materials in a repository from a specific knowledge area [20, 21]. Such limitation is an influential factor for a user to continue using these tools, since it is difficult to determine whether it is worth continuing the process of exploring materials in the repository or whether it is better to rely on other external search strategies.

C. Works Associated with Search Processes

With regard to search activities, some results of related studies on the basis of usability criteria are highlighted. Authors like [18] found a number of limitations associated with the use of search interfaces and deployment strategies of digital resources. The authors propose a list of suggestions and recommendations related to the improvement in the definition of criteria and the distribution of results on the screen. Finally, some of the results obtained from recent studies associated with the use of interfaces in academic repositories are mentioned. For example, in [22] authors identified deficiencies in the user interfaces related to utility, learning and knowledge of the classification scheme. In [8] authors claim that poor definition of a knowledge representation scheme, specifically associated with taxonomies and keyphrases [23], is one of the factors which hinders the process of finding digital resources using interfaces based on visualization techniques. In addition to this, there are also some other factors which impede navigation on knowledge representation structures for locating digital resources. In terms of efficiency, results of these studies show that, under certain conditions, search interfaces lead to an under-utilisation of information, in cases where results of search processes of digital resources associated with other subjects or knowledge areas are not relevant. Such results may often be related to a poor definition of metadata, thus losing all semantic ability to make enriched searches based on themes or some specific knowledge area [22, 24].

III. Methodology and Applied Model for Analysis

Considering these results, it is therefore necessary to design a work scheme outlining a series of activities linked to the identification of evaluation criteria for the analysis of search interfaces based on visualization techniques. Fig. 1 summarizes the work path which led to the development of this study.

In order to do this, the evaluation was performed throw a questionnaire, and made by five developers in the use of visualization tools. The purpose of the questionnaire was to identify qualitative aspects for the development of visual search interfaces based on 3 aspects: i) technical aspects of each library, ii) aspects of data integration of each library and finally, iii) visual aspects of each library. For the technical aspects, elements related to the characteristics of each library were considered such as ease of use, flexibility, scalability, performance, among others. Regarding data integration, its operation was considered from a predefined dataset for its deployment in each library, learning curve for data integration, navigation, hierarchical structure. Finally, some visual aspects associated with aesthetics, ease of navigation, among others, were considered.

Fig. 1. Work scheme source: proprietary development.

Such phases were considered with the aim to define a navigation map for the selection of criteria and factors which may have a significant influence on the development of visualization strategies applied to search interfaces in digital repositories. Considering this, each activity is presented below, as well as the assessment results shown in section IV and the definition of implications in section V respectively. In consequence the methodology is described as: 1) Analysis of knowledge representation schemes: At this stage, instruments associated with the most common knowledge representation schemes applied to digital repositories are identified. In addition to this, evaluation criteria for subsequent selection are defined. 2) Analysis of visualization techniques: For this phase, it is essential to define visualization techniques that facilitate exploration based on a knowledge representation scheme within a digital repository. Similarly, a study to determine the associated criteria for the selection of visualization techniques is conducted. 3) Analysis of libraries and graphical components: At this stage, the analysis of libraries and graphical components is carried out in order to select those that are most effective and which can lead towards the development of search interfaces based on visualization techniques, and finally 4) Analysis of search criteria on interfaces: A selection of factors associated with the design of search interfaces from the point of view of usability is made. Besides this, the principles related to the area of Human-Computer Interaction (HCI) are established.

IV. Results of Evaluation Criteria

This section presents an analysis of the factors which have a direct influence on the development of search interfaces based on visualization techniques. A number of analyses based on a study [25, 26] are made in order to identify selection criteria.

A. Analysis of Knowledge Representation Scheme

In the field of digital repositories, a knowledge representation scheme is defined as the different ways in which data can be structured and represented. This is done with the aim to facilitate classification processes, organization and association of concepts, based on a domain knowledge or a previously defined subject area.

Each knowledge representation technique requires a notation, which determines aspects of the subject area, levels of relationships, links and forms of association, among others. A knowledge representation scheme should be consistent and realistic to represent a subject area and its relationships in order to be effective in a learning environment [27]. This may either increase or decrease the user knowledge, among other things. Authors like [28], classifies these knowledge organization strategies in five different principles: 1) Elimination of ambiguities, related to the way a representation scheme facilitates the location of a term or concept without duplicating it. 2) Synonym control, referring to the monitoring performed by the structure to distinguish and / or associate multiple terms conceptually related. 3) Hierarchical relationships, which are related to the structure ability to determine the relationship level of two terms according to their meaning. 4) Associative relationships, which mean the structure ability to indicate the relationship of terms coming together to represent concepts associated with ideas. 5) Presentation of properties related to the knowledge representation instrument and its ability to associate levels of semantic representation from the use of metadata.

Considering the differences and features of the knowledge representation schemes analysed by [29], as well as the approaches raised by [28] on knowledge representation schemes, Table I shows some of the most representative features provided by each of these knowledge representation schemes and which are focused on the development of digital repositories.

TABLE I. Features Of Knowledge Representation Instruments

Features	Taxonomies	Thesaurus	Ontologies
Methods of use	Navigation	Textual and navigational	Textual and navigational
Forms of navigation	Sequential	Sequential – relational	Nonsequential - relational
Methods of access	Graphic	Textual and graphic	Textual and graphic
Forms of search	Hierarchic	Hierarchic and relational	Semantic
Data Representation	Structured	hierarchic, alphabetical, índex	Taxonomy: Tables with concepts, synonyms, descriptors, proceedings, relationships, attributes, values, axioms
Types of relationships	Hierarchic	Hierarchic, association and equivalence	Hierarchic and any type of relationship
Complexity of use	Low	Medium	High
Level of interoperability	Structural	Semi-structural	Semantic

According to previous studies made by [30, 8], in the design of digital repositories, the most common instruments used to represent knowledge are: taxonomies, ontologies, thesauri, graphs, mind maps, among others.

Fig. 2 shows a study of the first three instruments, as they are the most common knowledge representation schemes for the design of digital repositories. The vertical axis shows the factors which have been defined for executing the verification process in three different instruments of knowledge representation. The first five factors are related to management criteria and information classification, selected from studies made by [24]. The following three factors are associated with usability criteria, whereas the last two make emphasis on search criteria and ease of implementation within a digital repository.

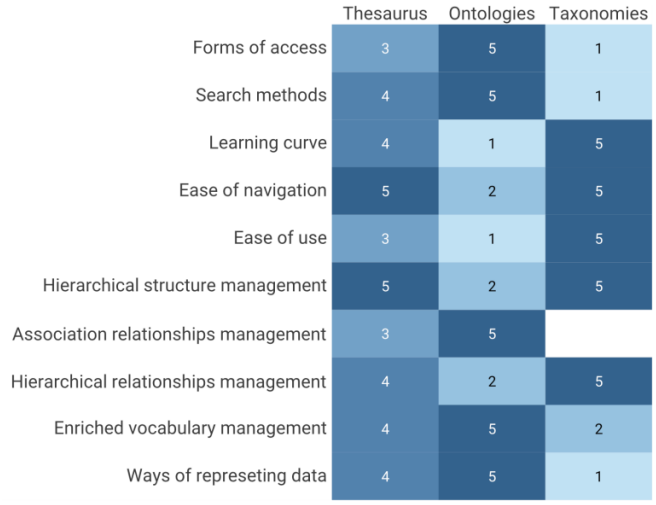


Fig. 2. Evaluation criteria for selection of knowledge representation scheme.

B. Analysis of Visualization Techniques

Fig. 3 shows summarised results of an evaluation administered for the selection of visualization techniques to facilitate the development of interface processes of visual search on digital repositories. The evaluation process executed at each interface was performed according to usability results obtained during the studies and the experience gained in the implementation and development of two projects: Organic.Edunet [31], and VOA3R/AGRIS [32].

From the assessment made in Fig. 3 by developers using the questionnaire defined, and according to evaluation criteria identified, the following interfaces can be identified as the most appropriate for implementation:

1. Specific features of each visualization technique: The interfaces which had better valuation according to aesthetic properties and classification methods of each of the hierarchies defined in a knowledge representation scheme were tree and radial.

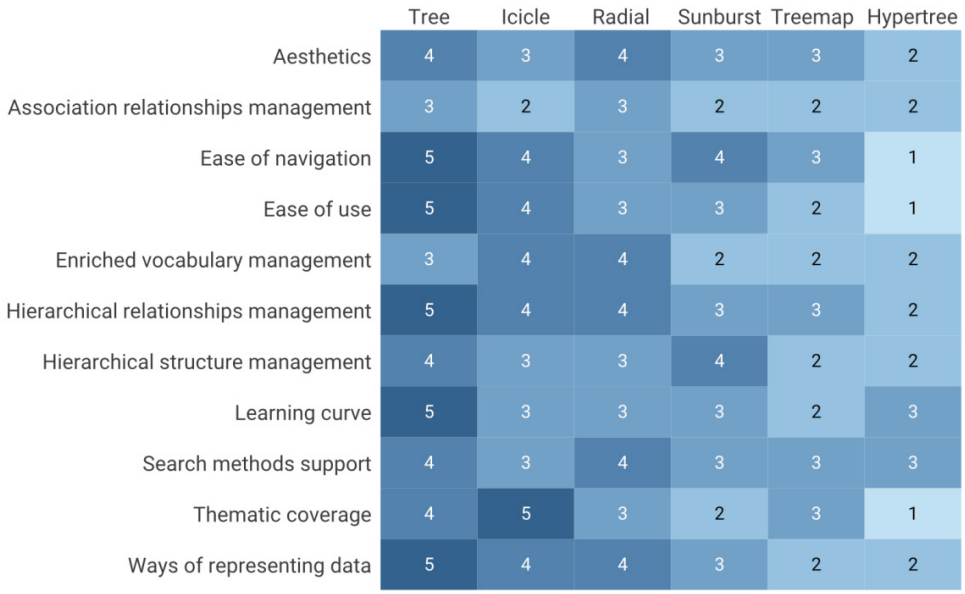


Fig. 3. Evaluation criteria for the selection of visualization techniques.

2. Development Features: In this category, the tree interfaces icicle and radial-search can be found as the most appropriate for the deployment of hierarchical structures, management of enriched vocabulary as well as hierarchy and association relationships. Similarly, they can be found for the integration of search methods specifically oriented to thematic coverage from a subject or area of knowledge.
3. Use features: For this category, the interfaces tree, icicle and radial-search are found as the most suitable to facilitate navigation and implementation processes from a previously established knowledge representation scheme.

C. Analysis of Libraries and Graphic Components

The development of applications based on information visualization requires components which can facilitate their use, adaptation and support throughout a project lifetime. There is a long list of projects on the Internet that can actually smooth implementation of visualization techniques from a number of function libraries and APIs, Open-Source-type. Some of them consist of computational functions which use diagramming strategies such as the (SVG) Scalable Vector (based on forms). This type of diagramming offers superb capabilities to support graphics applications on most Internet browsers and the development of interactive applications without the need for specialized technologies.

Other similar technologies offer support capabilities to display code on any browser compatible with HTML5 features ("w3c.github.io,") combining canvas-type labels (based on pixels). This type of technology improves -to some extent- the user experience as it omits the installation of plugins or add-ons, since most browsers provide graphics natively, making hardware graphics acceleration easier [33]. Thus, for any vector graphics based on these technologies and depending on the activity performed, the developer should make greater efforts. Fig. 4 presents a valuation related to their performance, for the purpose of determining selection criteria according to the needs required for implementation. For example, the most promising graphic component to made data integration with repositories in visual search interfaces based on the experience of Oraganic.Edunet project [31] was D3js.

D. Analysis of Search Criteria

Finally, it is important the definition of search criteria based on metadata defined by digital resources to perform the implementation of visualization techniques. Search criteria based on metadata have been considered in several studies in the field of digital libraries [25]. For example, in the Europeana digital library [36], some search criteria are defined such as: subject, language, content provider, digital resource format (image, video, audio, etc.). In the MERLOT digital repository [37] the following are defined: knowledge area, title, author, type of digital resource, date of update and evaluation of the digital resource. Such strategies facilitate search processes of digital resources [38, 39, 40, 41].

A digital resource itself lacks features that may facilitate its access, use and subsequent reuse. Therefore, for a digital resource to be considered a learning object, it must have at least some of the following characteristics: i) educational and instructional design for

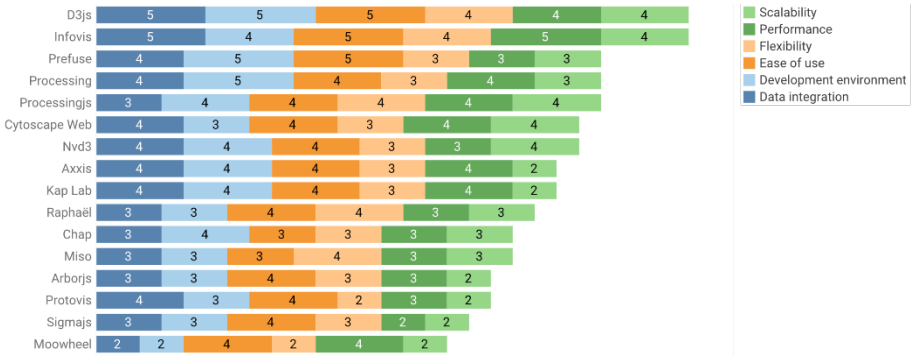


Fig. 4. Evaluation criteria for selection of visualization library.

its creation [42]; ii) educational components for its use [43]; iii) a set of descriptions called metadata which help to achieve some degree of interoperability, accessibility and reusability between different systems [44, 45].

Metadata is data whose purpose is to provide a description of a learning object. In education, it facilitates the construction of a digital resource and determines the most suitable elements for specific learning needs. In the area of learning objects, the use of metadata is widely accepted as a means to increase its quality and reusability [46]. Therefore, metadata is used to describe learning objects in order to simplify information search and retrieval processes. In addition to this, metadata also becomes essential for obtaining appropriate search results and meeting the needs of users [22].

V. Implications for Developing Interfaces Based on Visual Search Methods

This section presents the challenges that information visualization faces for its implementation taking place in interfaces and as search methods on digital repositories which facilitate access to learning objects. These will be addressed by areas of interest in order to identify elements that allow reaching decisions when using this type of strategy.

A. Issues with Regard to Information Visualization in Digital Repositories

The effectiveness that information visualization may have in the field of digital repositories is something unquestionable. However, there are also a number of challenges concerning the application of solutions and / or the implementation of alternatives. That said, a list of challenges facing the implementation of such strategies on digital repositories is presented below. Firstly, development time and adaptation to visualization techniques taking into account the digital resources features within the repository: This implies doing the planning under certain frequency –and depending on the versions given in the development process- for the approach of usability testing which can facilitate interface growth and adaptation according to the most suitable access criteria for development.

Another challenge is learning curve acquired by the development team to perform the implementation of visualization techniques: It is important that the development group carrying out such solutions have basic knowledge for the management of structures and components that are part of the Vector Graphics Scalable (SVG) library functions, or Canvas APIs graphics (based on pixels) supported by HTML5 ("w3c.github.io,"), as the case may be. These are technologies which support the use of radial, hyperbolic, tree, in categories and other structures in general to manage on-screen graphics resources.

B. Implications Associated with Knowledge Representation Schemes

A knowledge representation scheme may have many implications within a classification process and management of digital resources. In addition to this and in the same manner, it can have a significant influence in search processes which are defined through the use of visualization techniques within a digital repository. In order to reduce this gap, it is important to identify the most frequent problems when using such representation strategies. The most relevant difficulties associated with the use are presented below: 1) Interface limitation to visualize and deploy the knowledge representation scheme based on thesauri [47]; 2) Confusion in the use of thesauri for search processes based on preferred and non-preferred terms [5]; 3) Terminological problems in search processes [3] is another aspect that is often attributed to the difficulties faced by both creators and users of a digital repository defined from a thesaurus; 4) Difficulties in having access to digital resources from the use of ontologies with multiple terminological definitions [6, 7]; 5) Errors associated with the meaning of taxonomy within an ontology [2, 4]; 6) Errors related to the definition of an ontology, associated with inconsistency, and information redundancy [2].

All the aforementioned problems highlight the serious limitations of knowledge representation schemes, thesauri and ontologies, when providing access to digital resources, especially for end users who want to reuse such resources.

C. Limitations Associated with Interface Design

It is clear that information visualization cannot encompass a comprehensive and effective solution within a search process. However, its inclusion could facilitate access by implementing different techniques and search strategies. In this regard, the deployment of digital resources should be considered a part of the solutions that a creator of digital repositories needs to consider in order to facilitate basic elements of usability. To this end, it is important to make emphasis on the following principles from the point of view of the interface: a) The number of digital resources deployed in each consultation process must be shown clearly and in a visible area; b) It must have a paging mechanism to facilitate navigation as well as an instrument to set the number of actual pages that each query displays; c) The number of digital resources deployed per page must be clearly displayed considering the method of paging implemented; d) Each digital resource deployed by a list page should contain relevant information associated with the title, description, content provider, the digital resource connection with similar resources in the same area of knowledge and the best preview of digital resource (image, video); f) It is important to add the deployment of results by implementing mechanisms for evaluating digital resources, according to the profile that best suits the characteristics of the digital repository through evaluation: experts, registered users, professionals, students, etc.; g) It is necessary to classify the access to digital resources according to the characteristics of use of the digital resource (secondary education, higher education, ongoing education, vocational training, and university), the users' profiles (teachers, students and professionals), the formats of digital resource (image, video, rtf document, presentation, slides, etc.), and the digital resource language (English, Spanish, etc.).

Finally, two recent studies, which indicate how visualization techniques can be useful and effective for the access to collections of digital resources, are compared [8]. In addition to this, problems of usability associated with the use of search interfaces from factors related to the usefulness, ease of learning and knowledge of users on two academic repositories are presented (Organic.Edunet and VOA3R) [19]. These papers explore the fundamental reasons for problems related to usability at the light of the results of several usability studies carried out in the context of the European projects "Organic.Lingua" (www.organic-lingua.eu) and "VOA3R" (www.voa3r.eu), where two different digital repositories are being developed.

D. Limitations Associated with the Quality of Metadata

The creation of a metadata is an activity that is performed manually (one person) or automatically (a computer). This has generated a number of inconsistencies and limitations when managing them due to complex upgrade processes which are defined through hierarchical and semantic structures (knowledge representation scheme), from instruments such as: 1) The use of controlled vocabularies (e.g. Thesauruses), 2) The application of a domain of specific use (e.g. Ontologies), and 3) The application of a knowledge domain for free use (DBpedia).

Despite the fact that most repositories include knowledge classification systems (for example, ontologies, thesauri) to help manage the content by experts and creators of a digital repository, the

lack of effective mechanisms restricts users to locate digital resources according to their search criteria. The success of the location of digital resources depends heavily on the quality with which the metadata is designed [48]. This factor is essential to obtain relevant search results from the use of search engines defined in a repository [49, 50]. On the other hand, the quality of metadata is important for improving indexing strategies of learning objects stored in them [16, 40, 51]. Therefore, the omission of metadata negatively affects search results based on search criteria defined -for example- in a specific knowledge area.

The quality of the content is an indicator that allows evaluating digital resources. On this work scenario, there are several studies that refer to the design of methods to evaluate the quality of digital resources [52, 53, 54] as well as the quality of existing content in digital resource collections [51, 52]. Based on these methods, a series of assessment criteria are defined to carry out quality assessments of metadata on a collection of digital resources.

VI. Conclusions and Considerations

In general, and according to the analysis of this study, it is considered that the development of graphical interfaces associated with knowledge representation schemes indicate that repository creators should focus their efforts on: i) define levels of relationship and partnership between digital resources by using knowledge representation structures that support taxonomies (thesauri or ontologies); ii) work with user-friendly interfaces: (tree, radial or icicle) that support the use of navigation routes to identify the level of classification within the taxonomic structure, depending on the levels of hierarchy or depth provided by a repository; iii) link relevant metadata fields (fields for classifying resources) with the navigation structure to facilitate scanning processes by categories or knowledge areas as this offers users the possibility to have a more significant and effective alternative to a collection of digital resources; iv) have visual integration into the taxonomic structure and in the number of digital resources available in each category, in a proportional way. Since the thematic coverage is a factor that significantly improves access to a collection of digital resources, this becomes an important circumstance as long as the displayed results are visually appealing for the participant.

Undoubtedly, digital repositories must work hand in hand with strategies to facilitate the interoperability and reuse of digital resources based on semantic enrichment determined in their metadata. However, greater efforts should be exerted to facilitate access to large collections of digital resources by raising strategies for the management and maintenance of digital resources based on good design and development practices. Some of the most representative recommendations associated with the area of digital repositories are presented below: 1) Metadata quality management: Understood as those policy decisions which should be taken into consideration for the management and administration of the quality metadata so as to enhance the conditions of precision in its definition in order to improve relevant search results of digital resources. 2) Linking metadata fields associated by theme: It is important to define strategies by implementing metadata to facilitate the search for digital resources from a subject or knowledge area. This activity can be performed by LOM metadata defined as “classification” and “relation”. This metadata provides a classification by areas or themes, something which could facilitate search results of digital resources according to a specific area of knowledge in order to simplify access to a collection of digital resources. 3) Knowledge representation scheme: Which means selecting a knowledge representation scheme with at least the following features: a) Taxonomic hierarchy: A basic definition of taxonomic hierarchy in order to facilitate the classification and use of the visualization technique. b) Hierarchical relationship and partnership: It is important to establish a partnership relationship or hierarchy within the structures of the established taxonomic structure to facilitate linking related terms (generic or specific) to the search process. c) Enriched vocabulary management: Providing expression capacity within the determined relationships from complementary terms that could provide search alternatives based on synonyms, non-preferred terms, etc. d) Defining search criteria: Depending on strategies of use given in the repository, it is essential to determine search criteria by defining metadata based on themes or knowledge areas to facilitate the exploration and interaction of users on a collection of digital resources. Similarly, it is crucial to integrate defined criteria given by the educational environment, such as profile users, language and types of digital resources, etc.

References

[1] Gasevic, D.,Djuric, D., and Devedzic, V. “Model Driven Engineering and Ontology Development” (2nd ed.). book, Springer Publishing Company, Incorporated (2009).

[2] Gómez-Pérez, A. Facultad D. I. Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases. “Banff Knowledge Acquisition for Knowledge-Based Systems, KAW’99”, 26, 6.1.1--6.1.18, (1999). Retrieved from http://sern.ucalgary.ca/KSI/KAW/KAW99

[3] Aitta, M., Kaleva, S., and Kortelainen, T. Heuristic evaluation applied to library web services. New Library World, 109(1/2), 25–45, (2008). JOUR. http://doi.org/10.1108/03074800810845985

[4] Noshairwan, W., Qadir, M. A. and Fahad, M. “Sufficient Knowledge Omission Error and Redundant Disjoint Relation in Ontology”. In K. M. Wegrzyn-Wolska and P. S. Szczepaniak (Eds.), Advances in Intelligent Web Mastering: Proceedings of the 5th Atlantic Web Intelligence Conference -- AWIC’2007, Fontainbleau, France, June 25 -- 27, 2007 (pp. 260–265). inbook, Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-72575-6_42

[5] Shiri, A., and Revie, C. “Usability and user perceptions of a thesaurus-enhanced search interface”. Journal of Documentation, 61(5), (2005), 640–656. JOUR. http://doi.org/10.1108/00220410510625840

[6] Gašević, D., and Hatala, M. “Ontology mappings to improve learning resource search”. British Journal of Educational Technology, 37(3), (2006), 375–389. article. http://doi.org/10.1111/j.1467-8535.2006.00611.x

[7] J. Z. Li, D. Gasevic, J. C. Nesbit and G. Richards, “Ontology Mappings Enable Interoperation of Knowledge Domain Taxonomies”. 2nd LORNET Int. Annu. Conf. (2005).

[8] Gaona-García, P. A., Martín-Moncunill, D., Sánchez-Alonso, S., and García, A. F. “A usability study of taxonomy visualisation user interfaces in digital repositories”. Online Information Review, 38(2), (2014), 284–304. JOUR. http://doi.org/10.1108/OIR-03-2013-0051

[9] Ternier, S., Verbert, K., Parra, G., Vandeputte, B., Klerkx, J., Duval, E., & Ochoa, X. “The ariadne infrastructure for managing and storing metadata”. IEEE Internet Computing, 13(4). (2009).

[10] Sicilia, Miguel-Angel, et al. “Learning object analytics for collections, repositories & federations.” Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM, (2013).

[11] Sicilia, Miguel-Angel. “Metadata research: Making digital resources useful again?.” Handbook of metadata, semantics and ontologies. (2014). 1-8.

[12] Kumar, V., Nesbit, J., and Han, K. “Rating learning object quality with distributed Bayesian belief networks: the why and the how”. In Fifth IEEE International Conference on Advanced Learning Technologies (ICALT’05) (2005), (pp. 685–687). In proceedings. http://doi.org/10.1109/ICALT.2005.230074b06a286ea5152f1782b02f45b8f84e4edd15a @ w3c.github.io. (n.d.). Retrieved from http://w3c.github.io/html/#the-canvas-element

[13] Thelwall, Mike, and Kayvan Kousha. “Figshare: a universal repository for academic resource sharing?.” Online Information Review 40.3 (2016): 333-346.

[14] Tsakonas, G., and Papatheodorou, C. “Exploring usefulness and usability in the evaluation of open access digital libraries”. Information Processing & Management, 44(3), (2008), 1234–1250. JOUR. http://doi.org/http://dx.doi.org/10.1016/j.ipm.2007.07.008

[15] Carol Tenopir, D. Council on Library and Information Resources Use and

Users of Electronic Library Resources: "An Overview and Analysis of Recent Research Studies", (2003).

[16] Nash, S. S. "Learning objects, learning object repositories, and learning theory: Preliminary best practices for online courses". Interdiscip. J. Knowl. Learn. Objects, vol. 1, (2005), pp. 217–228.

[17] Jeng, J. "Usability Assessment of Academic Digital Libraries: Effectiveness, Efficiency, Satisfaction, and Learnability". Libri, 55(2–3), (2005), 96–121. http://doi.org/10.1515/LIBR.2005.96

[18] Hee Kim, H., and Ho Kim, Y. "Usability study of digital institutional repositories". The Electronic Library, 26(6), (2008), 863–881. http://doi.org/10.1108/02640470810921637

[19] Hartson, H. R., Shivakumar, P., and Pérez-Quiñones, M. A. Usability inspection of digital libraries: a case study. International Journal on Digital Libraries, 4(2), (2004), 108–123. JOUR. http://doi.org/10.1007/s00799-003-0074-4

[20] Hitchcock, S., Woukeu, A., Brody, T., Carr, L., Hall, W., and Harnad, S. "Evaluating Citebase, an open access Web-based citation-ranked search and impact discovery service". (n.d.).

[21] Tsakonas, G., and Papatheodorou, C. "Exploring usefulness and usability in the evaluation of open access digital libraries". Information Processing & Management, 44(3), (2008), 1234–1250. JOUR. http://doi.org/http://dx.doi.org/10.1016/j.ipm.2007.07.008

[22] Cechinel, C., Sánchez-Alonso, S., and Sicilia, M. Á. (2009). "Empirical Analysis of Errors on Human-Generated Learning Objects Metadata. In F. Sartori, M. Á. Sicilia, and N. Manouselis" (Eds.), Metadata and Semantic Research: Third International Conference, MTSR 2009, Milan, Italy, October 1-2, 2009. Proceedings (pp. 60–70). inbook, Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-04590-5_6

[23] Martin-Moncunill, David, et al. "Evaluating the practical applicability of thesaurus-based keyphrase extraction in the agricultural domain: Insights from the VOA3R project." Knowledge Organization 42.2 (2017): 76-89.

[24] Sánchez-Alonso, S., and Sicilia, M.-A. "Using an AGROVOC-based ontology for the description of learning resources on organic agriculture". In M.-A. Sicilia and M. D. Lytras (Eds.), Metadata and Semantics (2009), (pp. 481–492). inbook, Boston, MA: Springer US. http://doi.org/10.1007/978-0-387-77745-0_47

[25] Gaona-García, P. A., Stoitsis, G., Sánchez-Alonso, S., and Biniari, K. "An exploratory study of user perception in visual search interfaces based on SKOS". Knowledge Organization, 43(4), (2016), 217–238. JOUR. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84981167175&partnerID=40&md5=7144f9aa3530ef59c4aa377546c7ab6b

[26] Gaona-García, P. A., Martin-Moncunill D., Montenegro-Marin, C. "Trends and challenges of visual search interfaces in digital libraries and repositories". The Electronic Library. 35.1 (2017): 69-98.

[27] Chrysafiadi, K., and Virvou, M. "A knowledge representation approach using fuzzy cognitive maps for better navigation support in an adaptive learning system". SpringerPlus, 2(1), 81. article. http://doi.org/10.1186/2193-1801-2-81

[28] Zeng, M. L., and Chan, L. M. Trends and issues in establishing interoperability among knowledge organization systems. J. Am. Soc. Inf. Sci. Technol., 55(5), (2004), 377–395. http://doi.org/10.1002/asi.10387

[29] Soler Monreal, C., and Gil Leiva, I. "Possibilities and limitations of thesauri in comparison with other systems of knowledge organization: folksonomies, taxonomies and ontologies". Revista Interamericana de Bibliotecología, 33(2), (n.d.), 361–377. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-09762010000200004&lng=en&nrm=iso&tlng=es

[30] Gaona-García, P. A., Sánchez-Alonso, S., and Montenegro Marín, C. E. "Visualization of information: A proposal to improve the search and access to digital resources in repositories". Ingenieria E Investigacion, 34(1), (2014), 83–89. JOUR. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84902156291&partnerID=40&md5=6f9360c1c3b115235be0da38e56b481c

[31] Manouselis, Nikos, Kostas Kastrantas, Salvador Sanchez-Alonso, Jesus Cáceres, Hannes Ebner and Matthais Palmer. 2009. "Architecture of the Organic.Edunet Web Portal." International Journal of Web Portals 1, no. 1:71-91

[32] Šimek, Pavel, Jiri Vaněk, Vladimir Očenásek, Michal Stočes and Tereza Vogeltanzová. 2012. "Using Metadata Description for Agriculture and Aquaculture Papers." Agris On-Line Papers in Economics and Informatics 4:79-90.

[33] Boulos, M. N. K., Warren, J., Gong, J., and Yue, P. "Web GIS in practice VIII: HTML5 and the canvas element for interactive online mapping". International Journal of Health Geographics, 9(1), (2010), 14. article. http://doi.org/10.1186/1476-072X-9-14

[34] Ruecker, S., Radzikowska, M., and Sinclair, S. "Visual Interface Design for Digital Cultural Heritage". book, Brookfield, VT, USA: Ashgate Publishing Company. (2011).

[35] Doerr, Martin, et al. "The europeana data model (edm)." World Library and Information Congress: 76th IFLA general conference and assembly. 2010.

[36] Schell, George P., and Max Burns. "Merlot: a repository of e-learning objects for higher education." E-Service 1.2 (2002): 53-64.

[37] Chirita, P.-A., Gavriloaie, R., Ghita, S., Nejdl, W., and Paiu, R. "Activity Based Metadata for Semantic Desktop Search". (n.d.).

[38] Lytras, M. D., and Sicilia, M. Where is the Value in Metadata? Int. J. Metadata Semant. Ontologies, 2(4), (2007), 235–241. article. http://doi.org/10.1504/IJMSO.2007.019442

[39] Ochoa, X., and Duval, E. "Quality Metrics for Learning Object Metadata". (n.d.).

[40] Park, J.-R. "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art". Cataloging & Classification Quarterly, 47(3–4), (2009), 213–228. JOUR. http://doi.org/10.1080/01639370902737240

[41] Wiley, David A. The instructional use of learning objects. Vol. 1. Bloomington, IN: Agency for instructional technology, 2002.

[42] Polsani, P. "Use and Abuse of Reusable Learning Objects". Journal of Digital Information, 3(4). (2006). article. Retrieved from https://journals.tdl.org/jodi/index.php/jodi/article/view/89

[43] Friesen, N. "Open Educational Resources: New Possibilities for Change and Sustainability". The International Review of Research in Open and Distributed Learning, 10(5). (2009). article. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/664

[44] Sanz, J., Dodero, J. M., and Sánchez-Alonso, S. "A Preliminary Analysis of Software Engineering Metrics-based Criteria for the Evaluation of Learning Objects Reusability". International Journal of Emerging Technologies in Learning (iJET), 4(0), (2009), 30–34. article. Retrieved from http://online-journals.org/index.php/i-jet/article/view/794

[45] Sicilia, M.-A., and Garcia, E. "On the Concepts of Usability and Reusability of Learning Objects". The International Review of Research in Open and Distributed Learning, 4(2). (2003). article. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/155

[46] Shiri, A. A., Revie, C., and Chowdhury, G. "Thesaurus-enhanced search interfaces". Journal of Information Science, 28(2), (2002), 111–122. http://doi.org/10.1177/016555150202800203

[47] la Prieta, F., and Gil, A. B. "A Multi-agent System that Searches for Learning Objects in Heterogeneous Repositories". In Y. Demazeau, F. Dignum, J. M. Corchado, J. Bajo, R. Corchuelo, E. Corchado, … A. Campbell (Eds.), Trends in Practical Applications of Agents and Multiagent Systems: 8th International Conference on Practical Applications of Agents and Multiagent Systems (2010) (pp. 355–362). inbook, Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-12433-4_42

[48] Beltrán-Alfonso, R., Torres-Tautiva, A., Gaona-García, P.A.,Montenegro-Marin, C.E. "Exploring the Relevance of Search Engines: An Overview of Google as a Case Study". International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, n. 4 (2017). Retrieved from http://www.ijimai.org/journal/node/1519

[49] Muñoz-Arteaga, J., Calvillo-Moreno, E. A., Ochoa-Zezzatti, C. A., Santaolaya-Salgado, R., and Álvarez-Rodríguez, F. "Use of Agents to Realize a Federated Searching of Learning Objects. In Y. Demazeau, F. Dignum, J. M. Corchado, J. Bajo, R. Corchuelo, E. Corchado", … A. Campbell (Eds.), Trends in Practical Applications of Agents and Multiagent Systems: 8th International Conference on Practical Applications of Agents and Multiagent Systems (2010) (pp. 1–8). inbook, Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-12433-4_1

[50] Stuckenschmidt, H., and Van Harmelen, F. "Generating and managing metadata for Web-based information systems". (2004). http://doi.org/10.1016/j.knosys.2004.04.004

[51] Chuanjun, Suo. "On the Evaluation of the Quality of Digital Collections [J]." The Journal of The Library Science In China 4 (2004): 45-48.

[52] Downes, Stephen. "Models for sustainable open educational resources." Interdisciplinary Journal of E-Learning and Learning

Objects 3.1 (2007): 29-44.
[53] Gonçalves, Marcos André, et al. ""What is a good digital library?"–A quality model for digital libraries." Information processing & management 43.5 (2007): 1416-1437.
[54] Sanz-Rodriguez, Javier, Juan Manuel Dodero, and Salvador Sanchez-Alonso. "Metrics-based evaluation of learning object reusability." Software Quality Journal 19.1 (2011): 121-140.

Paulo Alonso Gaona-García

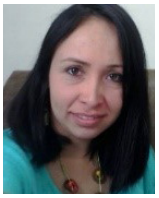
Paulo Alonso Gaona García is titular professor at Engineering Faculty of Universidad Distrital Francisco José de Caldas, Bogotá – Colombia, and active member of GIIRA research group. He earned a Ph.D. in Information and Knowledge Engineering from the University of Alcalá in 2014. He is Master in Information Science and Communication from Universidad Distrital Francisco José de Caldas (Bogotá Colombia). His research interest includes data science, network and communications, e-learning, and visual analytics.

Carlos Enrique Montenegro Marin

Carlos Enrique Montenegro Marin is a Ph.D.in systems and computer services for internet from University of Oviedo, Asturias, Spain (2012). He has a Diploma of advanced studies 2008 of the Pontifical University of Salamanca. He is MSc. Science in Information and Communication Systems from the Universidad Distrital Francisco José de Caldas. He is a System engineer. His research interests include Object-Oriented technology, Language Processors, Modeling Software with, DSL and MDA.

Elvis Eduardo Gaona García

Elvis Eduardo Gaona García is a Ph.D. in Engineering at Distrital University, Bogotá D.C. Colombia. He earned an M.Sc. in Information Science and Communications at Distrital University, Colombia. He is associated professor at Engineering Faculty of Distrital University. His research interests are network and communications, microgrids.



Adriana Gomez Acosta

Adriana Gomez Acosta, was born in Bogotá - Colombia. She is systems engineer, graduated from the San Mateo University Foundation. She has a master in Design and Management of technological projects of the University of La Rioja. Comes from a family traditionally dedicated to teaching; and precisely its line of professional training is oriented to university teaching and research in software development. She currently works as a research professor in the areas of Big Data and Business Intelligence, as well as advising projects of the Engineering Faculty of the San Mateo University Foundation.

Yusef Hassan Montero

Yusef Hassan Montero has a PhD in Information Science from the University of Granada. Currently he works at UNIR (Universidad Internacional de la Rioja), where he directs the master's degree program in User Experience and teaches information visualization at the master's degree in Big Data and Visual Analytics. In addition, he collaborates with Scimago Lab in the design and development of scientific information interfaces. His research interests include Interaction Design, Human-Computer Interaction and Data Visualization.

Building Phrase Polarity Lexicons for Sentiment Analysis

Rahim Dehkharghani*

Faculty of Engineering, University of Bonab, Bonab, I.R. (Iran)

Received 11 May 2018 | Accepted 17 October 2018 | Published 19 October 2018

Abstract

Many approaches to sentiment analysis benefit from polarity lexicons. Most polarity lexicons include a list of polar (positive/negative) words, and sentiment analysis systems attempt to capture the occurrence of those words in text using polarity lexicons. Although there exist some polarity lexicons in many natural languages, most languages suffer from the lack of phrase polarity lexicons. Phrases play an important role in sentiment analysis because the polarity of a phrase cannot always be estimated based on the polarity of its parts. In this work, a hybrid approach is proposed for building phrase polarity lexicons which is experimented on Turkish as a low-resource language. The obtained classification accuracies in extracting and classifying phrases as positive, negative, or neutral, approve the effectiveness of the proposed methodology.

Keywords

Sentiment Analysis, Polarity Lexicons, Polarity Classification, Phrases.

DOI: 10.9781/ijimai.2018.10.004

I. Introduction

Due to ever-increasing amount of online information especially in social media, manual processing of data to extract valuable information is impractical. The task of extracting information from text might attempt to extract the polarity of text--which is called sentiment analysis or polarity classification. This task has been very popular in recent decades but still it is far from the ideal.

Many approaches to sentiment analysis require polarity lexicons to assign a polarity tag (positive, negative or neutral) to a segment of text. There exist a good deal of workA on polarity lexicon generation which is grouped into two categories by Liu [1]: dictionary based methods and corpus based methods. Dictionary based methods start with a small seed word list and expand it upon synonymy and antonymy relations by using dictionaries such as WordNet [2]. In corpus based methods, semantic relations between terms in a corpus are employed to generate polar terms. These relations include pointwise mutual information [3] considering the co-occurrence of words in a window (e.g., a sentence), conjoined adjectives (by ``and'' or ``but'') [4], and delta tf-idf [5].

In this paper, a novel approach has been suggested for generating and classifying phrases as positive, negative, or neutral. The proposed approach is illustrated as a flowchart in Fig. 1. At first, raw phrases are collected; then, classification features are extracted; and finally, different classification tasks are accomplished to classify phrases as positive, negative, or neutral (objective). The contribution of this work is proposing a novel approach for generating phrase polarity lexicons and building the first phrase lexicon for the Turkish language. Note that the proposed approach is language independent, and it has been applied on Turkish as a case study. An alternative method for building such lexicons would be manually annotating the whole lexicon which has been employed in [6].

* Corresponding author.

E-mail address: rdehkharghani@bonabu.ac.ir

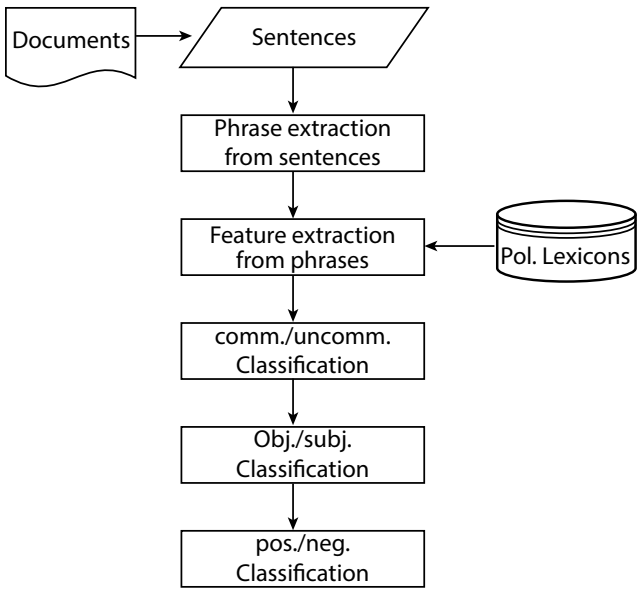


Fig. 1. The proposed methodology as a flowchart.

Among natural languages, most researchers have focused on English, while many other languages such as Turkish suffer from a lack of polarity resources. We have already generated two polarity lexicons for Turkish--Polar word Set (PWS) and SentiTurkNet (STN) in previous work [7]; however, because these lexicons have a limited coverage, a new polarity lexicon is generated in this work. Phrases require special attention in sentiment analysis because in most cases, the overall polarity of a phrase differs from the polarity of its parts. For example the phrase "...daha fazla olmalıydı" [...it should be more (better) than this] has a negative polarity but the constituting words are neutral. It seems that it is impossible to estimate negative polarity of this phrase based on the polarity of its parts.

In the remainder of this paper, Section II reports some previous works on sentiment analysis. The detailed explanation of the proposed approach is provided in Section III which is followed by experimental evaluation in Section IV. Applying the proposed method on other languages is discussed in Section V. Discussion on results are presented in Section VI, and Conclusions and future works are provided in Section VII.

II. Literature Review

Sentiment analysis can be done in different granularity levels: document, sentence, phrase, concept, and word levels. In [8], the authors investigated document level sentiment analysis using machine learning techniques.

In sentence level, Meena and Prabhakar [9] addressed the effect of conjunctions, and semantic relations between sentences.

In phrase level sentiment analysis, two works have been accomplished by Wilson and colleagues: [10] and [11]. In 2005, the authors proposed an approach which first classifies an expression as subjective or objective and then estimates its polarity in the case of subjectivity.

This method estimates the contextual polarity of an expression by using a large number of subjectivity clues and the prior polarity of appeared words in the expression. This work mostly relies on statistical methods. The obtained accuracies in classifying expressions as objective/subjective and also positive/negative range from 61% to 75%. The authors extended their work in 2009. The focus of this work is to figure out which features are more important in automatically distinguishing between prior and contextual polarity. Multi-perspective Question Answering (MPQA) is used as the opinion lexicon in this work.

Again in phrase level, Agrawal et al. [12] proposed a method to predict contextual polarity of subjective phrases in a sentence. The authors present new classification features which could achieve higher accuracies in ternary (positive/negative/neutral) classification of phrases over two baselines--majority class baseline as well as a more difficult baseline consisting of lexical n-grams.

Yi et al. [13] analyzed grammatical sentence structures and phrases for sentiment analysis purposes. The authors present Sentiment Analyzer which extracts sentiment towards a subject from online text documents. Instead of classifying the sentiment of an entire document about a subject, the designed system detects all references to the given subject, and determines the sentiment in each of the references.

In [14], the authors proposed an approach for extracting sentiments associated with positive or negative polarity for specific subjects in a document, instead of classifying the whole document as positive or negative. In this work, the goal is to identify semantic relationships between sentiment expressions and subject terms. Finally Kiritchenko and Saif [6] investigate phrases with opposite polarity such as *happy accident*. Phrases in this work are extracted from a large set of tweets using some patterns and they have been manually annotated by positive/negative tags.

In concept-level, Tsai et al. [16] presented a two-step methodology which combines iterative regression and random walk with in-link normalization to build a concept-level sentiment lexicon. In [16], the authors presented a methodology for enriching SenticNet [17]--a polarity lexicon in English-- concepts with affective information by assigning an emotion label to those concepts.

There exist also a good deal of research on building polarity lexicons. Liu [1] categorizes these methods into two groups: dictionary based approaches and corpus based approaches.

Dictionary based approaches start with a small seed set (e.g., 20 terms) and expand the list by using the existing relations such as synonymy and antonymy among terms in dictionaries. In [18], Hu and Liu used this method to generate a list of polar English terms and then manually cleaned up the generated list to remove errors. A similar approach was proposed in [19], which assigns also a sentiment score to each word by using a probabilistic method. Esuli and Sebastiani [20] proposed a methodology to assign three polarity scores (positive, negative, and neutral) to each synset in English WordNet. This approach was modified in [7] to build a polarity lexicon for Turkish based on the Turkish WordNet [21].

In corpus based approaches, having a seed word list with known polarities, new polar words are extracted based on the existing semantic relations in the corpus. One of the early ideas was proposed in [4]. The authors used conjunctions in a corpus to find new polar adjectives. They show that conjoined adjectives by "and" usually have the same polarity while they usually have the opposite polarity when conjoined by ``but''. Some extra relations such as "Either-or" and "Neither-nor" were also used for this purpose. Kanayama and Nasukawa [22] followed this approach and improved it by adding this idea: consecutive sentences usually have the same polarity.

There are also some effort on sentiment analysis of Turkish text. Yıldırım et al. [23] accomplished a sentiment analysis task on Turkish tweets in the telecommunication domain. The authors applied a multi-class ternary (positive/negative/neutral) classification by support vector machines on tweets using features such as inverse document frequency, unigrams and adjectives. They also benefit from NLP techniques such as normalization, stemming, and negation handling. Vural et al. [24] presented a system for unsupervised sentiment analysis in Turkish text documents using SentiStrength [25] by translating its polarity lexicon to Turkish. SentiStrength is a sentiment analysis tool for English which assigns a positive and a negative score to a segment of text. Kaya et al. [26] investigated sentiment analysis of Turkish political news in online media. The authors used four different classifiers--Naive Bayes, Maximum Entropy, SVM, and the character-based n-gram language models-- with a variety of text features: frequency of word unigrams, bigrams, root words, adjectives and effective (polar) words. They conclude that Maximum Entropy and the n-gram language models are more effective than the SVM and Naive Bayes classifiers in classifying Turkish political news. Boynukalın [27] has worked on emotion analysis of Turkish texts by using machine learning methods. The author has investigated four types of emotions: joy, sadness, fear, and anger on a dataset that she built for this purpose.

III. Phrase Polarity Lexicon Generation

A hybrid approach has been used for building a phrase polarity lexicon. The first phase in this approach is pre-processing. This pre-processing step as well as the whole approach are explained in the following subsections.

A. Phrase Extraction

A phrase is defined as "a small group of words standing together as a conceptual unit, typically forming a component of a clause" in Oxford dictionary[1]. As another definition from a Turkish dictionary[2], phrase is defined as "birkaç sözcükten oluşan ifade" (an expression composed of several words). According to Oxford dictionary, phrases can be divided into noun, verb, adjective, adverbial, and prepositional phrases[3]; however, only adjective, noun, and verb phrases are the focus of this work. According to Oxford dictionary, a noun phrase is a word or group of words containing a noun and functioning in a sentence as subject, object, or prepositional object such as "*inanılmaz bir*

[1] http://www.oxforddictionaries.com

[2] https://www.seslisozluk.net/

[3] https://en.oxforddictionaries.com/grammar/phrases

performans" (an unbelievable performance); A verb phrase is a verb with another word or words indicating tense, mood, or person such as "*gözlerimizi boyadılar*" (they deceived us); An adjective phrase is a phrase whose head is an adjective such as ``nasıl böyle saçma" (how silly like this).

At the first phase of the suggested methodology, a phrase list is generated by extracting collocations--trigrams and quadrigrams--using patterns in Table I, from 270,000 sentences in Turkish movie reviews (detailed explanation of the movie dataset is provided in Section IV.A). In this table, numbers inside parentheses are the number of phrases extracted by each pattern; moreover, one sample phrase has been provided for each pattern. The employed patterns (trigrams and quadrigrams) could extract 5213 phrases which are generally meaningful in Turkish; however, bigrams and 5-grams could extract more phrases, which is left as future work. Fig. 2 illustrates the percentage of each part of speech (POS) in extracted patterns. As seen in this figure, adjectives play the most important role and verbs play the least important role among other parts of speech. In this figure, number *P* upon a POS tag bar means that *P*% of phrases includes word(s) with the mentioned POS tag. In order to extract phrasal expressions from text, different methods could be used. One is exploited in this work, which is evaluated in Section IV. Extracting related words together in dependency parse tree is another method which was experimented in [28], but using those kind of phrases--which are usually separated by other words—for sentiment analysis purposes would be very challenging.

TABLE I. Patterns Used for Extracting Phrases from Sentences. Numbers Inside Parentheses Are the Number of Phrases Extracted by each Pattern

Triples	quadruples
adv adj verb (750) çok güzel anlatıyor	adv adj adj noun (436) çok iyi bir şekilde
adv adj noun (972) çok eğlenceli vakit	adv adj noun noun (394) bir güzel oyuncu hikayesi
adj noun verb (676) iyi iş çıkarmış	adj adj noun verb (491) abartılacak bir şey yoktu
adv adv verb (306) çok çok beğendim	adv adv adj verb (525) Neyse çok iyi diyemem
Adv adv adj (310) çok çok sevimli	adv adj noun verb (371) çok büyük saygı duyuyorum

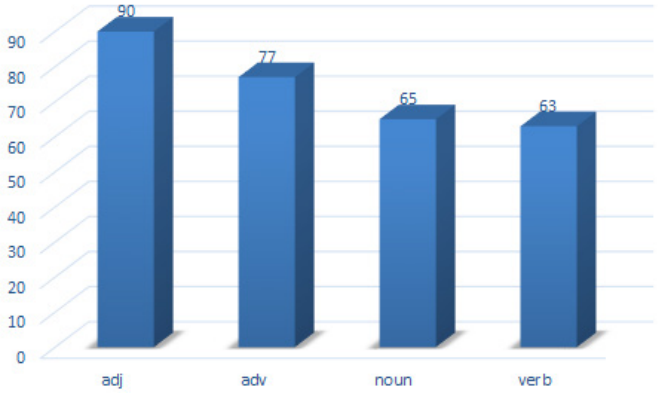


Fig. 2. The contribution of each POS tag in generation of phrases.

At this point a question might raise in mind that how well the suggested patterns can extract the existing phrases in the text. To answer this question, we manually extracted all correctly formed phrases from 100 randomly chosen sentences in Turkish movie reviews and obtained the following results:

Automatically extracted correctly-formed phrases from text: 171

existing bigram and 5-gram correctly formed phrases in text: 93

existing trigram and quadrigrams correctly formed phrases in text: 189

The recall value of the existing correctly formed trigram and quadrigrams is 90% (171 ÷ 189), and the overall recall is 60% (171 ÷ 282); Note that the total number of existing phrases in the above-mentioned senetnces is 282 (189 + 93). The reason for not very high performance is due to ignoring bigrams and 5-grams. Moreover, 10% of trigrams and quadrigrams could not be extracted by the proposed patterns.

There exist some works in the literature which attempt to extract key-phrases [29][30] from text, but those phrases are different from the ones extracted in the current work because any potential phrase (not only key phrases) are extracted in the current work.

In order to extract the above-mentioned phrases, an NLP tool for Turkish named ITU parser [31] is exploited to assign POS tags to the words in a sentence. Other NLP techniques such as lemmatization or normalization were not used.

Note that the collocated expressions are not necessarily compositional. As defined by Manning and Schütze [32], an expression is compositional if its overall meaning can be estimated based on the meaning of its parts. For example, the meaning of non-compositional phrase, "göz boyamak" (to deceive), cannot be estimated according to its words (literally means coloring the eyes); so, knowing that this phrase has negative polarity and catching it in the text helps estimate the polarity of the text including the phrase.

B. Basic Features for Phrase Classification

The list of features for phrase classification is provided in Table II, and explained below.

- N-grams: This method computes the co-occurrence probability of terms (words) with each other in a phrase. The goal is to distinguish correctly formed phrases from incorrectly formed ones. If the co-occurrence probability of included terms in a phrase is high, most probably they constitute a correctly formed phrase. As mentioned in [32], N-gram language model can be computed by probabilities given in Eq. (1).

$$log(P(t_i t_j t_k)) = log(P(t_i)) + log(P(t_j \mid t_i)) + log(P(t_k \mid t_i t_j)) \quad (1)$$

  $P(t_i)$ is the probability of seeing the term t_i in a phrase, $P(t_j \mid t_i)$and $P(t_k \mid t_i t_j)$ are respectively conditional probabilities of seeing t_j and t_k after seeing the given terms t_i and $t_i t_j$ in a phrase, and $P(t_i, t_j, t_k)$ is the probability of having correctly formed phrase with three terms: t_i, t_j, and t_k. For example, in the phrase "daha fazla olmalıydı" (it should be more (better)), extracted by the pattern [Adv Adv Verb], *log(P(daha))*, *log(P(daha|fazla))*, and *log(P(olmalıydı|daha fazla))* are computed. A similar equation could be written for quadruples.

- Hit number in a search engine: In this feature, each phrase is searched in Google search engine to capture its hit number. The higher the number of hits for a phrase, the higher the probability of correct formation.
- Document frequency: This feature counts the number of times each phrase appears among 270,000 Turkish sentences (unlabelled) as Turkish movie reviews.

After training a classifier by using the above mentioned features, all phrases are classified as "correctly formed" and "incorrectly formed". By the help of this classification, incorrectly formed phrases are removed from the list. This classifier has been trained by 1,000 phrases manually labeled as "correctly formed" and "incorrectly formed". The labeling task has been done by two (plus one) native Turkish speakers. The agreement of two labelers is 85.4%, and the third labeler helped

in labeling 14.6% of phrases which were not agreed by two labelers. The input of this classification task is a set of 5213 phrases extracted by the patterns of Table I and the output is a set of 4950 correctly formed phrases. A correctly classified sample is "üstüne yok doğrusu" (Actually there is no higher level upon it) and an incorrectly formed phrase which was misclassified as correctly formed is "bir film günün en ..." (the most ... of a movie day). Note that an incorrectly formed phrase very unlikely appears in a Turkish sentence. Also in some cases, not all words of a phrase are extracted by the proposed patterns; extracting only some (not all) words of a correctly formed phrase makes it incorrectly formed. The classification method used in this work is Logistic classifier which is used for its high generalization accuracy; the classification tool is WEKA, which is a known java-based machine learning tool, and the evaluation method is 5-fold cross-validation. In this evaluation method, the training set is divided into five equal parts, the first four parts (80%) are used as training set and the remaining 20% of data are supposed as test set. This task is repeated for five times for different 80/20 percent of training data.

TABLE II. Features Extracted for Classifying Phrases As Positive, Negative, or Neutral

Phrase Extraction	N-grams Hit number in Google Document frequency
Polarity Classification	Appearing in Pos/Neg sentences Pos/neg word count

C. Features for Phrase Polarity Classification

The classification features for phrase extraction and polarity classification are listed in Table II. First set of features have been used for phrase extraction (explained in Section III.B) and the rest of features have been used for polarity classification of phrases which are explained below.

- Appearing in Positive/Negative sentences: This feature counts the number of times a phrase appears in 2,700 positive and negative sentences--as a subset of movie reviews. The details of this subset are given in Section IV.A.
- Positive/negative word count: This feature captures the number of positive and negative terms appeared in a phrase. Two Turkish polarity lexicons are used for this purpose: Polar Word list and SentiTurkNet. In polar word list, words are already separated as positive and negative; In SentiTurkNet, similar to SentiWordNet, three polarity scores are assigned to each Turkish synset. A Turkish word is assumed as positive (or negative) if the average positivity (or negativity) score of its synsets is greater than their average negativity (or positivity) score. This feature is assumed as a baseline for phrase lexicon generation as it simply counts the number of positive and negative terms in a phrase.

D. Polarity Classification of Phrases

After classifying each phrase as *correctly formed* or *incorrectly formed*, the *correctly formed* phrases are classified as positive, negative, or neutral. For this purpose, two classification tasks (listed below) are carried out by using features listed in Table II. Similar to the first classification task, the classifier, evaluation method, and classification tool are respectively logistic regression, 5-fold cross validation, and WEKA.

- Classifying phrases as subjective and objective (neutral): In this classification, the output list of the phrase extraction phase (correctly formed phrases) is classified as objective and subjective; in other words, objective phrases are removed from the list. The input of this classification is a set of 4950 phrases and the output is a set of 2092 subjective phrases ignoring 2858 objective (neutral) ones. A correctly classified sample is "nasıl böyle saçma" (how silly like this) and an objective phrase which is incorrectly classified as subjective is "tabii romantik komedi" (of course a romantic comedy). The training set for this classification is a set of 800 correctly formed phrases which have been manually labelled as subjective and objective by two (plus one) native speakers, with 88% agreement on the labels of the two labelers, and getting help from the third labeller on 12% of labels which were not agreed at least by two labelers. The labels of remaining 4150 phrases (4950-800) are estimated by the trained classifier.
- Classifying subjective phrases as positive and negative: In this classification task, the output of previous step (subjective phrases) are classified as positive and negative. The input of this classification task is a set of 2092 phrases and the output is a set of 1591 positive and 501 negative phrases. The lower number of negative phrases is due to the lower number of negative reviews and sentences in movie reviews. The training set for this classification is a set of 500 correctly formed phrases which have been manually labelled as positive and negative by three native speakers of Turkish, with 83% agreement among two (plus one) labellers, getting help the third labeller on 17% of data which were not agreed by at least two labellers. The labels of remaining 1692 phrases (2092-500) were estimated by the trained classifier.

A correctly classified positive phrase is "tek işe yarar ..." (the only useful ...); a correctly classified negative phrase is "kesinlikle çok gereksiz bir..." (Absolutely a very unnecessary ...). A positive phrase that has been misclassified as negative is "izlediğim en iyi gerilim" (The best intensity movie that I have ever watched).

Note that instead of two binary classification (objective/subjective and positive/negative), one ternary (positive/negative/neutral) classification task has been also accomplished which is explained in Section IV.B.

IV. Experimental Setup

This section evaluates the proposed methodology by classification accuracy, extrinsic evaluation, and confusion matrix. Note that there is no Turkish polar phrase lexicon, therefore the generated list is new to Turkish. A subset of the generated lexicon is illustrated in Table III, and the complete and cleaned list can be provided for researchers via email. In this table, the column named "composition" shows the polarity of constitutive words in a phrase. A phrase may be composed of positive and objective (PosObj) words, negative and objective (NegObj) words, only objective words (Obj), or positive and negative words (PosNeg). Note that PosNeg phrases exist only in negative set; No positive phrase include both positive and negative words.

TABLE III. A Small Subset of Positive and Negative Phrases

Phrases	composition	tag
etkileyecek bir konu (an impressing subject)	PosObj	P
farklı bir eser (a different work)	Obj	P
olumsuz dersem yalan olur (I cannot say it is negative)	NegObj	P
büyük bir ayıp (a big shame)	NegObj	N
bir anlamı yok (does not have any meaning)	Obj	N
daha iyi olmalı (It should be better)	PosObj	N
iyi bir felaket! (a good disaster!)	PosNeg	N

We also investigated the distribution of positive, negative, and objective words in positive/negative phrases, which is illustrated in Fig. 3 and 4. In Fig. 3, x axis is the number (and percentage) of negative (or positive in Fig. 4) and y axis is the number (and percentage) of positive (or negative in Fig. 4) words in generated positive (or negative in Fig. 4) phrases. For example the number 33% in coordinate [0,0] of Fig. 3 means that 33% of positive phrases has zero positive and zero negative words.

As seen in Fig. 3, majority of positive phrases are composed of objective words, or objective plus positive words but they do not include negative words; however, positive words can be seen in negative phrases. In summary, it is usual to see positive (or negative) words in positive (or negative) phrases but the contribution of positive words in negative phrases (21%) is much more than the contribution of negative words in positive phrases (zero). Numbers upon each circle shows the percentage of phrases included in it.

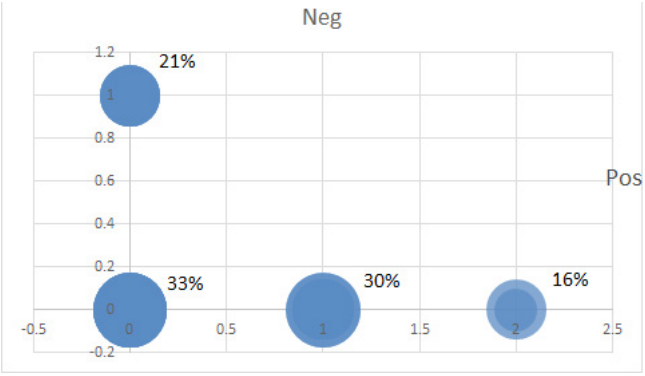


Fig. 3. Distribution of polar words in positive phrases. Numbers upon each circle show the percentage of phrases included in them.

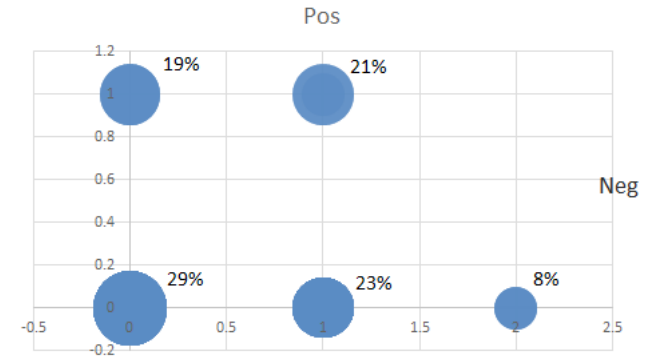


Fig. 4. Distribution of polar words in negative phrases. Numbers upon each circle show the percentage of phrases included in them.

A. Datasets

As mentioned in previous sections, the proposed approach has been applied on documents of movie domain in Turkish, which are more formal than some other data types such as tweets. Applying the suggested methodology on other kinds of textual data such as social media text would be challenging, as those data (e.g., tweets) are usually informal and noisy, including abbreviations and useless text.

Sentiment analysis is a domain-dependent task, therefore a given term may have different polarities in different domains; e.g., the term "big" is positive for *room size* in hotel domain but negative for *battery size* in camera domain. Although extracting polar terms (or phrases) from one domain and applying on another may have some drawbacks, however, in resource-lean languages such as Turkish we have to accept these weaknesses; moreover, some of the extracted phrases from movie domain are domain independent. In this work, two datasets have been used, first one for extracting the phrases and the second one for an extrinsic evaluation.

- Turkish movie reviews[4]. We have manually labelled 1,000 randomly chosen documents from this dataset as positive, negative, or neutral in our previous work [28]. We also labelled 2,700 sentences appearing in these documents as positive, negative, or neutral. Only the labels of sentences are used in this work. The distribution of [neutral, positive, and negative] sentences and documents are [50%, 30%, 20%] and [52%, 29%, 19%] respectively. The average length of each document and each sentence in this domain are respectively 23 and 9 words. The labeling task is accomplished by three (plus one) people and the agreement among at least two labellers is 81% for sentence level analysis. Again, the fourth labeller helped tag those sentences which were not agreed by at least two labellers. The already assigned rating scores to each movie review are not used in this work because we require labels in the sentence level but existing rating scores of movie reviews are available only at document level. We preferred manual labelling which is also more accurate than the rating scores. Each sentence or document is labelled as positive, negative, or neutral, if it conveys a positive, negative, or neutral polarity to the reader.
- Turkish restaurant reviews[5]. This dataset was used as training set in Semeval 2016-task 5 [33], which has been already labelled with three tags: positive, negative, and neutral. The aim of this task is to estimate the polarity label of each aspect appearing in a sentence. This dataset includes 239 documents and 1104 sentences, which has been used in the current work for evaluating the generated lexicons in sentence-level sentiment analysis. The average length of each document and each sentence in this domain are respectively 26 and 8 words.

B. Evaluation of Phrase Polarity Lexicon

In order to separate polar phrases from non-polar ones, one ternary (positive/negative/neutral) and three binary classification tasks (correctly/incorrectly formed, objective/subjective, and positive/negative) have been accomplished. The intuition behind this is that incorrectly formed phrases must be excluded from the extracted list, then the remaining list should be classified as positive, negative, or neutral. The classification accuracies for binary classification of phrases as correctly formed and incorrectly formed are listed in Table IV, and classification accuracies for binary and ternary classification of correctly formed phrases are listed in Table V. Moreover, confusion matrices for both binary (positive/negative) and ternary (positive/negative/neutral) classification of correctly formed phrases are provided in Tables VI to IX.

TABLE IV. Binary Classification of Turkish Phrases as Correctly Formed and Incorrectly Formed by Logistic Classifier Using 5-Fold Cross Validation On Training Data (%)

Feature name	correct/incorrect
N-grams	76.4
Hit number	70.45
Doc. freq.	72.20
All features	79.40

[4] These reviews are collected from www.beyazperde.com which are available in http://sentilab.sabanciuniv.edu/resources/

[5] http://metashare.ilsp.gr:8080/repository/browse/semeval-2016-absa-restaurant-reviews-turkish-train-data-subtask-2/ef952246940f11e5886b842b2b6a04d76a1959c4385a46bda776dd510ac3522e/

TABLE V. The Accuracy Of Binary And Ternary (Positive/Negative/Neutral) Classification of Turkish Phrases by Logistic Classifier Using 5-Fold Cross Validation on Training Data (%)

Feature name	ternary	subj/obj	pos/neg
pos/neg sentences	73.42	70.01	88.04
pos/neg words	71.02	68.22	85.16
Both features	74.43	72.90	91.31

TABLE VI. Confusion Matrix for Binary (Pos/Neg) Classification of Turkish Phrases with All Features

True	Estimated	
	positive	negative
Positive	0.93	0.07
Negative	0.18	0.82

TABLE VII. Confusion Matrix for Binary (Subjective/Objective) Classification of Turkish Phrases with All Features

True	Estimated	
	subjective	objective
Subjective	0.80	0.20
Objective	0.21	0.79

TABLE VIII. Confusion Matrix for Binary (Correctly/Incorrectly Formed) Classification of Turkish Phrases With All Features

True	Estimated	
	corr. formed	incorr. formed
corr. Formed	0.83	0.17
incorr. Formed	0.20	0.80

TABLE IX. Confusion Matrix for Ternary (Positive, Negative, and Neutral) Classification of Turkish Phrases with All Features

Feature name	Positive	Negative	Objective
Positive	0.79	0.05	0.16
negative	0.11	0.68	0.21
Objective	0.17	0.15	0.68

C. Extrinsic Evaluation

In order to evaluate the generated polarity lexicon, an extrinsic evaluation is carried out. The generated lexicon as well as other two Turkish lexicons, polar word set and SentiTurkNet, are used to estimate the polarity of Turkish restaurant reviews. This set includes 1104 Turkish sentences in restaurant domain. This dataset has been used as a benchmark in Semeval competition task 5 -Aspect based sentiment analysis. In this dataset, the goal is to estimate the polarity of aspects appearing in a sentence which have been tagged with three labels: positive, negative, and neutral.

The obtained accuracies with and without using the generated polarity lexicon are given in Table X. In this table, the abbreviations STN, PWS, and PL respectively stand for SentiTurkNet, Polar Word Set, and Phrase Lexicon. This sentiment analysis task simply searches for polar words in a sentence. No NLP technique except tokenization and word cleaning is employed in this system, as the goal is only to measure the usefulness of the generated lexicon. As seen in Table X, adding the phrase polarity lexicon increases the classification accuracy only by two percentage points. The reason (of low increment) can be the low number of idioms and multi-word polar phrases used in the sentences of restaurant domain. Moreover, catching a phrase in a sentence is not always straightforward. The appearance order of constituting words of a phrase in a sentence should be the same as its order in the phrase, so that the sentiment analysis system can find the phrase in the sentence. Note that phrases are extracted from movie domain but applied on restaurant domain. Extracting phrase from one domain does not necessarily makes them domain dependent. For example the phrase "nasıl böyle saçma" (how silly like this), can be used for any domain; however, there exist domain dependent phrases such as "iyi seyirler" (happy watching) which can be used only in movie domain.

TABLE X. The Accuracy of Binary (Positive/Negative) and Ternary (Positive/Negative/Neutral) Classification of Turkish Restaurant Reviews by Logistic Classifier Using 5-Fold Cross Validation on Training Data (%)

Lexicons used	Binary	Ternary
STN +PWS	73.02	67.23
STN+PWS+PL	75.17	69.22

V. Applying the Proposed Method on Other Languages

Since the grammar of natural languages is different from each other, in order to extract phrasal expressions from different languages, different patterns should be exploited. For example, in Turkish, verbs generally appear at the end of sentence, whereas in English, they usually appear in the beginning, after the subject. That is why in the suggested patterns for Turkish (Table I), verb is the last POS tag.

In this section, we examine how well the proposed methodology works on English. Below, necessary updates for applying the suggested methodology on English are explained.

Phrase Extraction: The suggested patterns in Table I should be adapted to English as done in Table XI. Note that the order of POS tags (especially verb) is changed. These patterns are used to extract candidate phrasal expressions from English movie reviews v2.0 [35]. As a result, 2588 raw phrases are extracted from the corpus.

Features and Classification: Classification process is the same as what was accomplished for Turkish. The first input of classification tasks is a set of 2588 raw phrases and the final output is a set of 295 negative and 534 positive phrases. In terms of features, two features, 'Hit number' and 'N-grams' are exactly the same as those used for Turkish, but remaining features use English resources. The feature 'document frequency' searches the generated phrases among randomly chosen 20000 sentences (ublabelled) from English movie reviews[6]. The feature 'appearing in pos/neg sentences' use 4200 sentences extracted from movie reviews, labelled as positive and negative. The feature, 'pos/neg words' benefit from three English polarity lexicons: SenticNet [17], SentiWordNet [18], and Liu's polarity lexicon [20]. In Liu's lexicon, positive words are separated from negative ones.

TABLE XI. Patterns of Table I Adapted to English. Numbers Inside Parentheses Show the Number of Phrases Extracted by Each Pattern

Triples	quadruples
verb adv adj (502) like very much	adv adj adj noun (286) very hard unsolvable problem
adv adj noun (372) very interesting effect	adv adv noun noun (226) very long time friend
verb noun adj (576) love you much	verb noun adv adv (191) put it somewhere else
verb adj noun (310) violate human rights	verb adv adv adj (125) support very very much

In SenticNet, we suppose a word as positive if its overall polarity score is greater than 0, or negative, otherwise; and in SentiWordNet, we suppose a word as positive if the average positivity polarity score

[6] http://www.cs.cornell.edu/people/pabo/movie-review-data/

of all synsets of the word is greater than its average negativity score, or negative, otherwise.

Evaluation: In this phase, we only provide classification accuracy for different classification tasks and omit other results. Table –XII and Table XIII provide these accuracies. As seen in Table XII, similar to Turkish, 'N-grams' is the most effective and 'Hit number' is the least effective feature; however, 'Hit number' feature in English has a little higher accuracy than in Turkish due to tremendous amount of English text in web compared to Turkish text in it. Moreover, overall accuracy in English is slightly greater than in Turkish.

TABLE XII. Accuracy of Binary Classification of English Phrases As Correctly Formed and Incorrectly Formed by Logistic Classifier Using 5-Fold Cross Validation on Training Data (%)

Feature name	correct/incorrect
N-grams	74.94
Hit number	72.05
Doc. freq.	73.32
All features	79.95

TABLE XIII. The Accuracy of Binary and Ternary (Positive/Negative/Neutral) Classification of English Phrases by Logistic Classifier Using 5-Fold Cross Validation On Training Data (%)

Feature name	ternary	subj/obj	pos/neg
pos/neg sentences	72.02	67.21	88.83
pos/neg words	69.22	66.92	85.66
Both features	72.33	70.80	92.01

In Table XIII, only binary classification of English phrases into positive and negative has slightly higher accuracy than the similar classification in Turkish, due to richer polarity lexicons and resources in English; other two classification tasks (subjective/objective and the ternary classification) have a few percentage points lower than the same tasks in Turkish.

VI. Discussion and Comparison

To the best of our knowledge, there is no work in Turkish to generate polar phrases, and other works have been applied on different datasets and languages (e.g., English). Therefore only similar works in English are reported below to provide a relatively fair comparison.

For generating polar phrases, Agrawal et al. [12] could achieve the accuracy of 70% in ternary (positive/negative/objective) classification and 84% in binary (positive/negative) classification of English phrases experimented on MPQA as the dataset. In [34], accuracies in neutral/polar classification range from 65% to 76% and 69% to 83% in polarity classifications for different datasets. In the current work, the best classification accuracies for ternary (positive/negative/objective), polar/objective, and positive/negative classification of Turkish phrases are respectively 74%, 73%, and 91%, whereas the same accuracies for English phrases are respectively 72%, 70%, and 92%. Due to different datasets used in the above-mentioned related work and the current one, the comparison may not be totally fair.

The most similar previous work to this one has been accomplished in [6]; the main difference between these two works is that the suggested approach in this paper for phrase extraction and annotation is semi-automatic but the annotation of phrases in [6] is manual (although phrase extraction is automatic and pattern-based).

According to the results reported in Section IV, the following conclusions can be extracted.

- The proposed approach for phrases, outperforms the baseline approach--counting the number of positive and negative terms in phrase--by 1 to 3 percentage points. This issue emphasizes the effect of non-compositional phrases in sentiment analysis, in which the polarity of the whole phrase cannot be estimated based on the polarity of its parts.
- The best classification accuracy in both Turkish and English phrases has been obtained in binary classification of phrases into positive and negative.
- In correctly/incorrectly formed classification of phrases, the N-gram feature obtained the highest accuracy. This finding approves the assumption that the higher the co-occurrence probability of a word-pair, the higher the probability of correct phrase formation by this pair.
- The highest per-class accuracies (confusion matrix values) belong to the positive class and lowest accuracies belong to the negative class. Generally positive expressions are more clearly expressed by people, compared to the negative expressions.
- Catching phrases and idioms in a sentence is not as easy as catching unigrams and bigrams in it as in some cases, phrases are separated by other words in the sentence.

VII. Conclusion and Future Work

In this work, a semi-automatic methodology is proposed to build phrase polarity lexicons. The proposed methodology consists of several methods such as word co-occurrence probability. Because the polarity of phrases cannot usually be estimated based on the polarity of its parts, covering phrases in sentiment analysis in a very challenging task. The generated lexicon is freely available for research community. Although the paper mostly focused on Turkish, the proposed methodology is language-independent and can be applied on other languages with small changes. The future work consists of adding polar idioms to existing polarity lexicons, and considering language issues such as negation in generated phrases.

References

[1] B. Liu, Sentiment Analysis and Opinion Mining, New York, NY: Cambridge University Press, June 2015.

[2] G. A. Miller, WordNet: a lexical database for English, Communications of the ACM 38 (11), 1995, pp. 39-41.

[3] P. D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, 2002, pp. 417-424.

[4] V. Hatzivassiloglou, K. R. McKeown, Predicting the semantic orientation of adjectives, in: ACL '98 Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, 1997, pp. 174-181.

[5] J. Martineau, T. Finin, Delta tfdf: An improved feature space for sentiment analysis, in: Proceedings of the Third International ICWSM Conference, 2009, pp. 258-261.

[6] S. Kiritchenko, S. M. Mohammad, Happy accident: A sentiment composition lexicon for opposing polarity phrases, in: Proceedings of 10th edition of the Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia, 2016.

[7] R. Dehkharghani, Y. Saygin, B. Yanikoglu, K. Oazer, Sentiturknet: a Turkish polarity lexicon for sentiment analysis, Language Resources and Evaluation. 50, 2016, pp. 1-19.

[8] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in: EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing – Vol. 10, 2002, pp. 79-86.

[9] A. Meena, T. Prabhakar, Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis, ECIR 2007: Advances in Information Retrieval, 2007, pp. 573-580.

[10] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: HLT '05 Proceedings of the conference

on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 347-354.

[11] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis, Computational Linguistics. 35 (3), 2009, pp. 399-433.

[12] A. Agarwal, F. Biadsy, K. R. Mckeown, Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams, in: EACL '09 Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 2009, pp. 24-32.

[13] J. Yi, T. Nasukawa, R. Bunescu, W. Niblack, Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques, in: ICDM: Third IEEE International Conference on Data Mining. ICDM. 2003, pp. 427-434.

[14] T. Nasukawa, J. Yi, Sentiment analysis: Capturing favorability using natural language processing, in: K-CAP '03 Proceedings of the 2nd international conference on Knowledge capture, ACM, 2003, pp. 70-77.

[15] A. C.-R. Tsai, C.-E. Wu, R. T.-H. Tsai, J. Y.-j. Hsu, et al., Building a concept-level sentiment dictionary based on commonsense knowledge, IEEE Intelligent Systems 28 (2), 2013, pp. 22-30.

[16] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, S. Bandyopadhyay, Enhanced senticnet with affective labels for concept-based opinion mining, IEEE Intelligent Systems 28 (2), 2013, pp. 31-38.

[17] E. Cambria, D. Olsher, D. Rajagopal, Senticnet 3: a common and commonsense knowledge base for cognition-driven sentiment analysis, in: Twenty eighth AAAI conference on artificial intelligence, 2014, pp. 1515-1521.

[18] M. Hu, B. Liu, Mining and summarizing customer reviews, in: KDD '04 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2004, pp. 168-177.

[19] S.-M. Kim, E. Hovy, Determining the sentiment of opinions, in: COLING '04 Proceedings of the 20th international conference on Computational Linguistics, 2004, Article no. 1367.

[20] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: LREC, Vol. 10, 2010, pp. 2200-2204.

[21] O. Bilgin, Ö. Çetinoğlu, K. Oazer, Building a wordnet for Turkish, Romanian Journal of Information Science and Technology 7 (1-2), 2004, pp. 163-172.

[22] H. Kanayama, T. Nasukawa, Fully automatic lexicon expansion for domainoriented sentiment analysis, in: EMNLP '06 Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 2006, pp. 355-363.

[23] E. Yıldırım, F. S. Çetin, G. Eryiğit, T. Temel, The impact of NLP on Turkish sentiment analysis, Türkiye Blisim Vakfi Bilgisayar Bilimleri ve Mühendisliği Dergisi 7 (1) (Basılı 8), 2015.

[24] A. G. Vural, B. B. Cambazoğlu, P. Şenkul, Z. Ö. Tokgöz, A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish, E. Gelenbe, R. Lent (Eds.), in: ISCIS, Springer, 2012, pp. 437-445.

[25] M. Thelwall, K. Buckley, G. Paltoglou, Sentiment strength detection for the social web, Journal of the American Society for Information Science and Technology 63 (1), 2012, pp. 163-173.

[26] M. Kaya, G. Fidan, I. H. Toroslu, Sentiment analysis of Turkish political news, in: WI-IAT '12 Proceedings of The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01, 2012, pp. 174-180.

[27] Z. Boynukalın, Emotion analysis of Turkish texts by using machine learning methods, MSc thesis, Middle East Technical University, 2012.

[28] R. Dehkharghani, B. Yanikoglu, Y. Saygin, , K. Oazer, Sentiment analysis in Turkish at different granularity levels, Natural Language Engineering 23 (4), 2017, pp. 535-559.

[29] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, C. G. Nevill-Manning, Kea: Practical automatic keyphrase extraction, in: Design and Usability of Digital Libraries: Case Studies in the Asia Pacific, 2005, pp. 129-152.

[30] Y. Zhang, N. Zincir-Heywood, E. Milios, Narrative text classi_cation for automatic key phrase extraction in web document corpora, in: Proceedings of the 7th annual ACM international workshop on Web information and data management, ACM, 2005, pp. 51-58.

[31] G. Eryiğit, ITU Turkish NLP web service, in: Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Association for Computational Linguistics, Sweden, 2014, pp. 1-4.

[32] C. D. Manning, H. Schütze, Foundations of statistical natural language processing, MIT Press Cambridge, MA, USA, 1999.

[33] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, Orphee declercq, v_eronique hoste, marianna apidianaki, xavier tannier, Natalia loukachevitch, evgeny kotelnikov, nuria bel, salud mar_a jim_enez-zafra, and gülşen eryiğit. semeval-2016 task 5: Aspect based sentiment analysis, in: Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval, Vol. 16, 2016, pp. 19-30.

[34] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis, Computational linguistics 35 (3), 2009, pp. 399-433.

[35] B. Pang and L. Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Proceedings of the 42nd annual meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004, pp. 271.

Rahim Dehkharghani

Dr Rahim Dehkharghani received his PhD in Computer Science and Engineering, Artificial Intelligence branch, from Sabanci Universty, Istanbul, in 2015 and his MSc degree in Computer Engineering, Software branch from Shahid Beheshti University, Tehran, in 2007. His main research area is Natural Language Processing and Data Mining. He specifically works on sentiment analysis of textual data. He is a faculty member of Computer Engineering group at University of Bonab (Bonab, Iran) since 2016.

A New Feature Selection Method based on Intuitionistic Fuzzy Entropy to Categorize Text Documents

M. B. Revanasiddappa, B. S. Harish *

Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka (India)

Received 4 January 2018 | Accepted 23 March 2018 | Published 7 April 2018

Abstract

Selection of highly discriminative feature in text document plays a major challenging role in categorization. Feature selection is an important task that involves dimensionality reduction of feature matrix, which in turn enhances the performance of categorization. This article presents a new feature selection method based on Intuitionistic Fuzzy Entropy (IFE) for Text Categorization. Firstly, Intuitionistic Fuzzy C-Means (IFCM) clustering method is employed to compute the intuitionistic membership values. The computed intuitionistic membership values are used to estimate intuitionistic fuzzy entropy via Match degree. Further, features with lower entropy values are selected to categorize the text documents. To find the efficacy of the proposed method, experiments are conducted on three standard benchmark datasets using three classifiers. F-measure is used to assess the performance of the classifiers. The proposed method shows impressive results as compared to other well known feature selection methods. Moreover, Intuitionistic Fuzzy Set (IFS) property addresses the uncertainty limitations of traditional fuzzy set.

Keywords

High Dimensionality, Feature Selection, Intuitionistic Fuzzy Entropy, Text Categorization.

DOI: 10.9781/ijimai.2018.04.002

I. Introduction

In recent years, rapid development in internet technology generated massive amount of text documents by private and public sectors. To handle such a huge amount of text documents, automatic text categorization has become a popular technology to manipulate and manage [1][2]. Text Categorization (TC) is a process of automatically categorizing unknown text documents into one or more pre-defined classes by their contents. TC has been successfully proposed for many applications viz., retrieving useful information in search engines, spam filtering, document organization, automatic document indexing etc. Due to these applications automatic text categorization is an important research area in text mining and information retrieval [3] [4]. Text categorization process includes feature extraction, pre-processing, feature selection and categorization. In feature extraction, the features are extracted from the text documents [5]. Each term (word) of the text document is considered as a feature and most of the features are unwanted and irrelevant. Further, during pre-processing, tokenization, stop-word elimination and stemming are employed to eliminate irrelevant and unwanted features [6]. The pre-processed text documents are represented in machine understandable form by employing a representation model. Then, feature selection method selects the most informative features from the representation model [7]. Feature selection plays high influence on the performance of classifiers and it is mainly used for dimensionality reduction [8][9]. Finally, selected feature subset is fed into a classifier to categorize the text documents. The text categorization suffers from high dimensionality of feature space. Due to this, the performance of the classifiers degrades and also takes more time for categorization [10][11]. The reduction of high dimensional feature matrix is a significant challenge in text categorization. To tackle this challenge, many researchers have proposed various feature selection methods [8-9][12-15].

Feature selection methods are generally partitioned into 3 groups: filter, wrapper and hybrid [11][15]. Presently there are various feature selection methods reported in the literature viz., Document Frequency [14], Term Frequency-Inverse Document Frequency [16], Term Strength [14], Mutual Information [17], Information Gain [18], Chi-Square [11][13], Ambiguity Measure [19], Term Frequency-Relevance Frequency [20], Symbolic Feature Selection [21], Distinguish Feature Selection [12], Entropy based Feature Selection [22] and many more. Among these feature selection methods, entropy based feature selection method computes the amount of uncertainty and the quality of information content present in the text. The concept of entropy was introduced by Shannon [23] and is called as Shannon entropy (information entropy). It is fundamentally based on information theory, which estimates the entropy value of each feature using probability. A lower value of entropy indicates higher contribution of information in decision making process and larger value of entropy indicate lesser contribution of information. The concept of entropy is described in several manners and applied in different areas [24].

The extended version of Shannon entropy is fuzzy entropy, which is non-probabilistic entropy [25]. It adopts a new term named match-degree to estimate entropy value [26][27]. Match degree satisfies the four properties of de Luca-Termini axioms [25][27]. The probability of the entropy is computed using the number of occurring terms. In

* Corresponding author.
E-mail addresses: revan.cr.is@gmail.com (M. B. Revanasiddappa), bsharish@sjce.ac.in (B. S. Harish)

contrast, the match degree in fuzzy entropy is computed using the membership values of the occurring terms [26]. The fuzzy entropy value does not only consider the number of features, but also considers the actual distribution of features by membership function. Hence, fuzzy entropy reflects more information in the actual distribution of the features than Shannon entropy in the feature space [27].

Fuzzy Entropy is based on the fuzzy set, which measures the fuzziness. The Fuzzy set is used to solve various real world problems, which mainly deals with impreciseness and vagueness [29]. In fuzzy set theory, the membership value of a feature varies between '0' to '1'. Fuzzy set assumes that the non-membership is complementary of membership, but in real world this assumption fails due to hesitation. This hesitation originates, while defining the membership function, due to lack of precise knowledge and it is another type of uncertainty. To address this hesitation, Atanassov [30] developed an Intutionistic Fuzzy Set (IFS), which is an enhanced version of Fuzzy Set. Unlike Fuzzy set which considers only the membership degree, the IFS considers the degree of non-membership and the hesitation along with degree of membership. In IFS, the degree of non-membership is less than or equal to the complement of the degree of membership due to the hesitation degree. The IFS gives mathematical model to deal with the vagueness arising from the inherent uncertainty and imprecision or insufficiency of imperfect information. Considering these advantages, many researchers have recently developed IFS based clustering techniques. Chaira [31] proposed a novel Intuitionistic fuzzy c-means (IFCM) clustering method. IFCM gives more accurate clusters, and shows better performance in the presence of uncertainty and it clusters the data with less iteration than the traditional FCM [32].

Motivated by the significant advantages of IFCM and Fuzzy Entropy, in this article a new feature selection method called Intuitionistic Fuzzy Entropy-Feature Selection (IFE-FS) is proposed for text categorization. This method selects feature subsets based on intuitionistic fuzzy entropy for text document categorization. Basically, it contains two phases: In the first phase, Intuitionistic membership degree is computed with the help of the IFCM clustering method. In second phase, the Intuitionistic fuzzy entropies on the basis of the Intuitionistic membership degree via match degree are estimated. Further, entropy values are arranged in ascending order and then the feature subset selection is pruned by the threshold value. Finally, selected feature subset is considered as input to the classifiers to categorize the text documents. More popular classifiers such as K-Nearest Neighbor (KNN) [33], Support Vector Machine (SVM) [34] and Radial Basis Function-Neural Network (RBF-NN) [35] are used. These classifiers are used due to their widespread use in the area of text mining and give competitive results on standard benchmark datasets [33]. The proposed method is experimented on 3 standard benchmark datasets viz., 20 Newsgroups, Reuters-21578 and TDT2. The performance of IFE-FS method is compared with CHI-Square ($\chi 2$) [11], Mutual Information (MI) [7], Information Gain (IG) [18] and Entropy based Feature Selection (EFS) [22]. All these feature selection methods have their variant characteristics and also are the well-known feature selection methods for text categorization.

The main contribution of this article is as follows:

1. Proposes a new feature selection method (IFE-FS) based on Intuitionistic Fuzzy Entropy, which reduces high dimensionality of feature matrix and enhances the performance of classifiers.
2. Resolves uncertainty (hesitation) limitation of Fuzzy Entropy by making use of IFS.
3. Conducts extensive experiments on three standard benchmark datasets. The experimental results acknowledge that the IFE-FS method shows outstanding performance in terms of F-measure.

The rest of the article is organized as follows: existing and entropy based feature selection methods are reviewed in section II. The preliminary outlines on the Intuitionistic fuzzy set are presented in section III. The proposed method (IFE-FS) along with illustration is described in detail in section IV. The experimental results and discussion are presented in section V. Finally, the conclusion along with future work is presented in section VI.

II. Literature Review

High dimensionality of the feature matrix is a major challenge during text categorization. Feature selection plays an effective role to identify the relevant features to minimize dimensionality. Feature selection method gives reduced text feature collection, reduces storage size, lesser model building and computation time, and better model interpretation [9][15][36]. However, many studies [37][38] [39] indicate that there are no feature selection methods to provide accurate discriminative feature for text categorization. Given the importance of feature selection for text categorization, many methods were proposed and some of them are presented below.

Recent researchers adopted the concept of entropy in feature selection methods to select the discriminate features. The concept of Entropy is based on the information theory proposed by Shannon which is also called as Shannon Entropy [23]. It measures the expected uncertainty of probability distribution with the results predicted by random experiments. Tang et al., [40] proposed a feature selection method based on the information theory, which uses Jeffreys-Multi-Hypothesis (JMH) divergence information measure. This method ranks the original features and maximizes the discriminative capacity for text categorization. Largeron et al., [22] proposed a feature selection method called Entropy based Category Coverage Difference (ECCD), which is based on the entropy. This method computes the occurrence of terms inside various classes with the assistance of entropy. Cai and Song [41] used the maximum entropy modeling with different feature selection methods to categorize text documents and also proposed novel feature selection method named as "count difference". This method considers features of both relevant and irrelevant classes to compute the frequency differences between relative documents. Vaghela et al., [42] presented an entropy based feature selection method, which uses InfoDist and Pearson's Correlation parameters. This method selects the features using InfoDist, which adopts conditional entropy to compute the relevancy of feature and category. Further, it eliminates the irrelevant features using Pearson's correlation. Liu and Song [43] presented the portion set of key words based on proximity degree. Key words are selected based on entropy, semantic field and association degree. Later, fuzzy classification is used to categorize documents.

Many researchers have identified the significance of entropy and have developed fuzzy entropy measures from various perspectives [44]. Fuzzy entropy has been widely used in pattern recognition, image processing and clustering analysis. De and Termini [25] proposed the first non-probabilistic entropy based on the fuzzy set theory. Khushaba et al., [45] described the fuzzy entropy in terms of match degree. The fuzzy entropy uses match degree instead of probability to estimate value of the entropy. Parkash et al., [46] used the principle of maximum weighted fuzzy entropy to develop two new weighted fuzzy entropy measures to remove the redundancy. Luukka [47] proposed a feature selection method based on fuzzy entropy, in which fuzzy entropy is estimated by using the membership degree in fuzzy set. This method minimizes the dimensionality of feature space and also enhances the efficiency of classifiers. Ahmadizar et al., [48] proposed a hybrid feature selection method, which has two stages. In the first stage, dimensionality of the feature matrix is reduced using the fuzzy entropy feature selection method. Fuzzy entropy values are ranked in ascending order and then the feature subset with lowest entropy value is selected. In the second stage, ant colony optimization is used to

select the features from feature subset for text categorization.

All the above mentioned methods fail to handle uncertainty that arises while defining the membership function. To handle this uncertainty, Atanassov [30] developed Intuitionistic Fuzzy Set (IFS) in 1986, which is an advanced version of fuzzy set. IFS considered degree of non-membership and hesitation along with degree of membership. In IFS, non-membership degree is computed by employing intuitionistic fuzzy complement generator [49][50]. Intarapaiboon [51] proposed a framework for text categorization based on similarity measure of intuitionistic fuzzy sets. In this framework, each document is represented in terms of the IFS, where the IFS uses sugeno's integral method to represent the document. Different types of similarity measures using IFS are described in [52][53]. Szmidt and Kacprzyk [54] proposed a feature selection method based on the IFS for text categorization. This method uses the degree of membership, non-membership and hesitation to resolve the imbalance and overlap class problem of categorization.

Chaira [31] proposed a novel Intuitionistic fuzzy c-means (IFCM) clustering method. The IFCM was evaluated on various CT scan brain images and achieved better performance than Type 2 fuzzy and traditional Fuzzy c-means algorithms. In literature, many researchers applied IFCM clustering method to solve image processing problems [55-58]. However, very less amount of work is reported on text categorization [51][54]. In addition to clustering methods, researchers developed entropy methods based on IFS. Szmidt and Kacprzyk [59] depicted entropy as far as non-probabilistic type of entropy. Burillo and Bustince [60] depicted entropy as far as degree of intuitionism of an intuitionistic fuzzy set. Hung and Yang [61] proposed two IFS based entropy measures and provided the axiomatic definition of entropy for IFS. Vlachos and Sergiadis [62] proposed new Intuionistic Fuzzy Entropy and also presented the connection between intuitive and mathematics of entropy for fuzzy set and IFS.

From literature survey, it is observed that a lot of work is reported on entropy and fuzzy entropy based feature selection methods for text categorization. In addition, it is also observed from the literatures that IFS addresses the limitation of fuzzy set by considering the hesitation degree. In this article, we developed a new Feature Selection method based on Intuitionistic Fuzzy Entropy (IFE). To the best of our knowledge this work is first of its kind, where a feature selection method based on IFE to categorize the text documents is proposed. The next section describes the construction of Intuitionistic Fuzzy Set for IFCM.

III. Construction of Intuitionistic Fuzzy Set (IFS)

In this section, the mathematical background on IFS is explained. The Intuitionistic Fuzzy Set, in general sense is portrayed by the membership degree, the non-membership degree and the hesitation degree [30].

Generally the fuzzy set A is defined on text document D_i as

$$A = \{\langle T_l, \mu_A(T_l)\rangle \mid T_l \in D_i\} \tag{1}$$

Where, $\mu_A(T_l) \rightarrow [0,1]$ denotes the membership degree of A. T_l is the l^{th} feature in the i^{th} document D_i. The membership value $\mu_A(T_l)$ defines the degree of belongingness of $T_l \in D_i$ in A. An Intuitionistic fuzzy set $\tilde{A}$ is described on document D_i as:

$$\tilde{A} = \{\langle T_l, \mu_{\tilde{A}}(T_l), \vartheta_{\tilde{A}}(T_l)\rangle \mid T_l \in D_i\} \tag{2}$$

Where, $\mu_{\tilde{A}}(T_l)$ and $\vartheta_{\tilde{A}}(T_l)$ denotes membership and non-membership degree of l^{th} feature T_l respectively. In the fuzzy set, another uncertainty emerges while defining the membership function due to imprecise knowledge. IFS handles this uncertainty by considering the hesitation degree. Unlike in fuzzy set, where the non-membership degree is calculated by taking the complement of membership degree, IFS computes the non-membership degree with the help of intuitionistic fuzzy complement generator. In this work, the Sugeno's intuitionistic fuzzy complement generator is used to calculate the non-membership degree [49]. The Sugeno's intuitionistic fuzzy complement is computed as

$$\vartheta_{\tilde{A}}(T_l) = \frac{1-\mu_{\tilde{A}}(T_l)}{1+\lambda\mu_{\tilde{A}}(T_l)} \tag{3}$$

Where, λ is the constant and $\lambda > 0$. When we set $\lambda = 1$, IFS becomes the traditional fuzzy set. Hesitation degree of a term $T_l \in D_i$ in $\tilde{A}$ is computed as

$$\pi_{\tilde{A}}(T_l) = 1 - \mu_{\tilde{A}}(T_l) - \vartheta_{\tilde{A}}(T_l) \tag{4}$$

Where $\pi_{\tilde{A}}(T_l)$ is the hesitation degree of l^{th} feature T_l. It is evident from the equation (4)

$$0 \le \pi_{\tilde{A}}(T_l) \le 1, \forall T_l \in D_i$$

Non-membership degree is computed from Sugeno's intuitionistic fuzzy complement. Thus, using Sugeno's intuitionistic fuzzy complement, the IFS becomes

$$\tilde{A}_\lambda = \left\{ \left\langle T_l, \mu_{\tilde{A}}(T_l), \frac{1-\mu_{\tilde{A}}(T_l)}{1+\lambda\mu_{\tilde{A}}(T_l)} \right\rangle \mid T_l \in D_i \right\} \tag{5}$$

IV. Proposed Method

A new feature selection method based on Intuitionistic Fuzzy Entropy (IFE) named as Intuitionistic Fuzzy Entropy-Feature Selection (IFE-FS) is proposed in this section. The Intuitionistic Fuzzy Entropy estimates the value of entropy on the basis of intuitionistic membership degree via match degree. The intuitionistic membership degrees are computed from the Intuitionistic Fuzzy C-Means (IFCM) Clustering method. The IFCM is based on the Intutionistic Fuzzy Set (IFS), which considers the degree of membership, non-membership and hesitation. The proposed method consists of two phases: in the first phase, an intutionistic membership value is computed using IFCM. In the second phase, entropy is estimated using the intutionistic membership values and a feature subset is selected on the basis of the entropy value.

A. Intuitionistic Fuzzy C-Means (IFCM) Clustering Method

Intuitionistic Fuzzy C-Means (IFCM) is an advanced version of the traditional Fuzzy C-means (FCM) and it is based on the IFS. Unlike, FCM which clusters input document based on the membership value, IFCM considers the non-membership and hesitation degree along with the membership degree.

Let us assume that there are k number of pre-defined classes $C_a, a = 1,2,3,...,k$. Each class contains n number of documents $D_i, i = 1,2,3,...,n$, and m number of features (terms) $T_l, l = 1,2,3,...,m$. Then the total number of documents is denoted by $z = [k \times n]$. The text document contains sequence of terms (features), where each term is treated as a feature. In text document, most of the features are irrelevant and redundant. These features results in high dimension in feature space and also degrade the performance of classifier. It is necessary to use pre-processing techniques to eliminate irrelevant and redundant features. The most common pre-processing task is

tokenization, elimination of stop-word and stemming. Tokenization is the process of partitioning the text into terms (features), called tokens. In the process of stop-word elimination, the features that do not have important information are eliminated from the feature space. For example, features like: “a”, “is”, “the”, etc., occur very frequently in all text documents and do not convey any meaning for class prediction. The process of stemming is to reduce the inflected terms to their root form. The stemming process helps to group the frequencies of different inflection to single term. For example, the features: “loves”, “loved” and “loving” have the similar meaning as of its root “love”. Further, the preprocessed text documents are represented using the Term Document Matrix (TDM) form. The TDM is considered as the input to the IFCM clustering method. IFCM assigns membership values to a document with respect to each class. But in this work, the aim is to compute the intuitionistic fuzzy entropy of each feature with respect to each class. So in order to compute the intuitionistic fuzzy entropy of each feature, we applied IFCM to each feature rather than the document. The objective function (J) of IFCM is as follows

$$J(U,V)=\sum_{j=1}^{c}\sum_{l=1}^{m}\mu_{jl}^{p}d_{jl}^{2} \tag{6}$$

Where, U is the membership matrix, V indicates the cluster centers matrix, μ_{jl} is the intuitionistic membership degree of l^{th} term in j^{th} cluster, p is the fuzzy coefficient. $d_{jl}=d(T_l,v_j)$ describes the distance measure between cluster center v_j and term T_l, v_j means j^{th} cluster center, T_l means l^{th} term and c is number of cluster. The intuitionistic fuzzy membership (μ_{jl}) is the combination of membership degree (μ_{jl}^{*}) and hesitation degree (π_{jl}).

$$\mu_{jl}=\mu_{jl}^{*}+\pi_{jl} \tag{7}$$

The membership degree μ_{jl}^{*} is computed as

$$\mu_{jl}^{*}=\frac{1}{\sum_{g=1}^{c}\left[\frac{d_{lj}^{2}}{d_{\lg}^{2}}\right]^{1/(p-1)}} \tag{8}$$

In order to compute the hesitation degree, firstly the non-membership degree (ϑ_{jl}) has to be computed by using the intuitionistic fuzzy complement generator. In this article, we used the Sugeno’s intuitionistic fuzzy complement generator [49] to compute the non-membership degree. The non-membership degree (ϑ_{jl}) is computed using the equation

$$\vartheta_{jl}=\frac{1-\mu_{jl}^{*}}{1+\lambda\mu_{jl}^{*}}. \tag{9}$$

Where, λ is the constant and $\lambda>0$. The value of non-membership degree changes by varying the value of λ. The hesitation degree computed using the membership degree (equation (8)) and the non-membership degree (equation (9)) is given by

$$\pi_{jl}=1-\mu_{jl}^{*}-\vartheta_{jl} \tag{10}$$

The Intuitionistic membership degree is computed using the equation (7) and further cluster center is calculated using the following equation

$$v_j=\frac{\sum_{l=1}^{m}\mu_{jl}^{*}T_l}{\sum_{l=1}^{m}\mu_{jl}^{*}} \tag{11}$$

The IFCM optimizes the objective function by continuously updating the membership degree and the cluster center until it meets the convergence criteria value ε, i.e., $\left|J^{(t)}-J^{(t-1)}\right|<\varepsilon$. Here, t is the iteration and ε is the user specified convergence criteria. Further, using this membership value, match degree is computed to estimate the entropy.

B. Intuitionistic Fuzzy Entropy (IFE)

The Intuitionistic Fuzzy Entropy is the non-probabilistic entropy in which we use match degree to estimate the entropy value. The match degree in IFE is computed using the intuitionistic membership value. IFE maximizes the capacity of discriminative features and generates more affluent information. Match degree is described as

$$X_{al}=\frac{\mu_a(T_l)}{\sum_{a=1}^{k}\mu_a(T_l)} \tag{12}$$

Where, $\mu_a(T_l)$ indicates the membership of l^{th} feature T_l in a^{th} class C_a. The match degree X_{al} is the ratio of intuitionistic membership of each feature T_l in class and by the summation of intuitionistic membership of feature T_l in all classes. The match degree X_{al} in intuitionistic fuzzy entropy uses the membership values to calculate the matching degree of feature in class. Later the Intuitionistic fuzzy entropy (IFE) of feature T_l in class C_a is computed using the following equation

$$IFE_{al}=-X_{al}\log X_{al} \tag{13}$$

Further, each features’ IFE value of class is summed to calculate the overall Intuitionistic fuzzy entropy of each feature

$$IFE_l=\sum_{a=1}^{k}IFE_{al} \tag{14}$$

The lower entropy value of a feature indicates major contribution about the classes. Therefore, we select r number of low ranked features. Here, the threshold value (r) indicates the number of features selected and also r varies depending on the datasets. But during experimentation, the major issue is in choosing an accurate threshold value. It will be varied multiple times before selecting the best value. The selected feature subset is then given to the classifiers to categorize the text documents. We used three different classifiers: K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Radial Basis Function Neural Network (RBF-NN). The performance of IFE-FS method is measured based on the F-measure. Algorithm 1 explains the individual steps involved in the proposed method.

In order to provide more meaningful theory of our proposed method, the next section illustrates the proposed method in detail by considering a simple term document matrix.

Algorithm 1: Intuitionistic Fuzzy Entropy Feature Selection (IFE-FS)

Data: k Number of class with n number of documents and m number of terms (features), Fuzzy Coefficient (p), Convergence Criteria (ε), λ is constant and Threshold (r)

Result: Class label

Step 1: Initialize cluster centers v_j randomly

Initialize number of iteration t=0

Repeat

Step 2: Compute membership degree using equation (8)

Step 3: Compute non-membership degree using equation (9)

Step 4: Compute hesitation degree using equation (10)

Step 5: Compute Intuitionistic Fuzzy membership degree using equation (7)

Step 6: Compute Intuitionistic Fuzzy cluster center using equation (11)

Step 7: Compute objective function using equation (6)

Until $\left|J^{(t)} - J^{(t-1)}\right| < \varepsilon$ is satisfied

Step 8: Compute match degree using equation (12)

Step 9: Compute class-wise Intuitionistic Fuzzy entropy for each feature using equation (13)

Step 10: Compute overall Intuitionistic Fuzzy entropy for each feature using equation (14)

Step 11: Select feature subset by using Threshold value (r)

Step 12: Selected feature subset will be the input to classifiers

C. Illustration

This section illustrates the individual steps involved in the proposed feature selection method. For illustration we considered an example of Term Document Matrix (TDM) of size 10 x 8, where 10 documents (D) are distributed among 3 classes (C) with 8 unique features (T). The same is presented in Table I.

The proposed feature selection method is based on the Intuitionistic Fuzzy Entropy, which uses the Intuitionistic Fuzzy C-Means (IFCM) clustering method to compute the intuitionistic membership degree for entropy estimation. So we computed each features' membership degree rather than that of the documents in all classes. The objective function of IFCM is shown in equation (7).

TABLE I. Term Document Matrix (TDM)

Documents	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
D_1	2	2	3	0	0	8	2	1	
D_2	3	1	3	0	0	0	1	6	C_1
D_3	1	1	1	1	1	1	0	1	
D_4	1	1	1	1	1	1	1	0	
D_5	1	1	2	0	0	0	0	1	
D_6	1	1	1	0	3	0	1	0	C_2
D_7	1	1	1	1	5	1	1	2	
D_8	1	1	1	1	1	0	0	1	
D_9	1	1	1	0	3	0	1	0	C_3
D_{10}	1	1	1	1	2	1	1	1	

In step 1, we initialized the cluster centers randomly. Since there are 3 classes and 10 documents, the matrix size will be 3 x 10, which is presented in Table II. Using these cluster centers, the membership degree is calculated.

In step 2, each feature membership degree is computed using equation (9) with respect to the class. Here, we set Fuzzy Coefficient p=2. The dimension of membership matrix is 3×8 and it is shown in Table III.

In step 3, the non-membership degree is computed using the Sugeno's Fuzzy Complement generator, which is mentioned in equation (10). Here, for illustration we set $\lambda = 0.5$. The obtained matrix will be in the form 3×8 as shown in Table IV.

In step 4, the hesitation degree is computed, using the membership and non-membership degree, using equation (11). The obtained resultant matrix is presented in Table V. Further, the intutionistic fuzzy membership degree is computed using equation (8) in step 5 and the membership values are presented in Table VI.

After the 1st iteration, the cluster centers are updated using equation (12) in step 6. Now the updated cluster centers are shown in Table VII.

The objective function value after 1st iteration is J=77.2455, which is computed using equation (7) in step 7. Similarly, same steps are repeated for every iteration. After the 15th iteration, the computed intuitionstic membership degrees and cluster centers are presented in Table VIII and Table IX respectively, and the objective function value is 76.6000. The steps 2 to 7 are repeated until it satisfies the user defined convergence criteria.

TABLE II. Cluster Centers

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	Classes
2.3709	2.3669	0.9335	0.8242	0.8542	0.6182	1.3264	0.8154	0.6182	1.0377	C_1
1.6367	1.0015	0.6376	0.9946	0.3689	0.7297	1.0364	0.6127	0.7297	1.0077	C_2
2.1788	1.5690	0.9856	0.8224	0.4923	1.4360	2.7697	0.8068	1.4360	1.3980	C_3

TABLE III. Membership Degree Matrix

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.7263	0.2395	0.6315	0.2294	0.2302	0.3569	0.1748	0.4503	C_1
0.1508	0.6232	0.1821	0.5726	0.2797	0.3270	0.6864	0.2641	C_2
0.1229	0.1373	0.1864	0.1980	0.4901	0.3161	0.1388	0.2856	C_3

TABLE IV. Non-Membership Degree Matrix

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.1556	0.6189	0.2292	0.6312	0.6302	0.4860	0.7006	0.3910	C_1
0.7328	0.2360	0.6910	0.2787	0.5715	0.5183	0.1856	0.5896	C_2
0.7721	0.7516	0.6854	0.6704	0.3530	0.5302	0.7495	0.5647	C_3

TABLE V. HESITATION DEGREE MATRIX

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	**Classes**
0.1181	0.1416	0.1393	0.1394	0.1396	0.1571	0.1246	0.1587	C_1
0.1164	0.1408	0.1269	0.1487	0.1488	0.1547	0.1280	0.1463	C_2
0.1050	0.1111	0.1282	0.1316	0.1569	0.1536	0.1117	0.1497	C_3

TABLE VI. INTUITIONISTIC MEMBERSHIP DEGREE MATRIX

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	**Classes**
0.8444	0.3811	0.7708	0.3688	0.3698	0.5140	0.2994	0.6090	C_1
0.2672	0.7640	0.3090	0.7213	0.4285	0.4817	0.8144	0.4104	C_2
0.2279	0.2484	0.3146	0.3296	0.6470	0.4698	0.2505	0.4353	C_3

TABLE VII. UPDATED CLUSTER CENTERS

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	**Classes**
2.5155	2.6050	0.9634	0.8486	0.9868	0.7969	1.3746	0.8556	0.7969	1.0558	C_1
1.9649	1.0953	0.7366	0.9331	0.4029	0.7802	1.3586	0.6445	0.7802	1.0729	C_2
2.1477	1.4133	0.9483	0.8438	0.4132	1.2625	2.5368	0.7663	1.2625	1.3451	C_3

TABLE VIII. INTUITIONISTIC MEMBERSHIP DEGREE MATRIX AFTER 15TH ITERATION

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	**Classes**
0.8444	0.3811	0.7708	0.3688	0.3698	0.5140	0.2994	0.6090	C_1
0.2672	0.7640	0.3090	0.7213	0.4285	0.4817	0.8144	0.4104	C_2
0.2279	0.2484	0.3146	0.3296	0.6470	0.4698	0.2505	0.4353	C_3

TABLE IX. CLUSTER CENTERS OF 16TH ITERATION

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	**Classes**
2.5155	2.6050	0.9634	0.8486	0.9868	0.7969	1.3746	0.8556	0.7969	1.0558	C_1
1.9649	1.0953	0.7366	0.9331	0.4029	0.7802	1.3586	0.6445	0.7802	1.0729	C_2
2.1477	1.4133	0.9483	0.8438	0.4132	1.2625	2.5368	0.7663	1.2625	1.3451	C_3

TABLE X. INTUITIONISTIC MEMBERSHIP DEGREE MATRIX AT 32ND ITERATION

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	**Classes**
0.8444	0.3811	0.7708	0.3688	0.3698	0.5140	0.2994	0.6090	C_1
0.2672	0.7640	0.3090	0.7213	0.4285	0.4817	0.8144	0.4104	C_2
0.2279	0.2484	0.3146	0.3296	0.6470	0.4698	0.2505	0.4353	C_3

TABLE XI. CLUSTER CENTERS AT 32ND ITERATION

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	**Classes**
2.3891	2.8796	0.9821	0.8299	1.0720	0.8305	1.4171	0.9003	0.8305	1.0617	C_1
2.1315	1.0463	0.7994	0.9161	0.3644	0.9130	1.7058	0.6594	0.9130	1.1555	C_2
2.1289	1.0456	0.7998	0.9160	0.3634	0.9137	1.7088	0.6598	0.9137	1.1562	C_3

In this illustration, the algorithm converges in the 32nd iteration. The final Intuitionistic membership values and cluster centers are shown in Table X and Table XI respectively and the objective function value is 76.6067.

In step 8, the final Intuitionistic membership values are given to the match degree to compute the actual distribution of features with respect to the class by using equation (13). The obtained match degree of the feature is shown in Table XII.

In step 9, class-wise Intuitionistic Fuzzy Entropy is estimated for each feature with respect to the class by using equation (14) and the estimated entropy value are presented in Table XIII. Later, all rows are summed up to compute the overall Intuitionistic Fuzzy Entropy for each feature by using equation (15), which is shown in Table XIV with a dimension of 1×8.

The obtained entropy value of each feature is sorted in ascending order to select the features with lower entropy values, which is shown in Table XV. The features are ranked $T_1 \succ T_3 \succ T_7 \succ T_2 \succ T_8 \succ T_4 \succ T_5 \succ T_6$ based on their entropy value. The lower entropy

TABLE XII. MATCH DEGREE MATRIX

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	**Classes**
0.7467	0.1616	0.6273	0.2249	0.2733	0.3135	0.1517	0.4597	C_1
0.1266	0.4197	0.1864	0.3874	0.3631	0.3433	0.4244	0.2701	C_2
0.1265	0.4186	0.1861	0.3876	0.3635	0.3431	0.4237	0.2701	C_3

TABLE XIII. INTUITIONISTIC FUZZY ENTROPY

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	**Classes**
0.0946	0.1279	0.1270	0.1457	0.1539	0.1579	0.1242	0.1551	C_1
0.1136	0.1582	0.1359	0.1595	0.1597	0.1594	0.1579	0.1535	C_2
0.1136	0.1583	0.1359	0.1595	0.1597	0.1593	0.1580	0.1535	C_3

TABLE XIV. OVERALL INTUITIONISTIC FUZZY ENTROPY

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
0.3219	0.4444	0.3989	0.4648	0.4734	0.4767	0.4402	0.4622

TABLE XV. FEATURES RANKED IN ASCENDING ORDER

0.3219	0.3989	0.4402	0.4444	0.4622	0.4648	0.4734	0.4767
T_1	T_3	T_7	T_2	T_8	T_4	T_5	T_6

TABLE XVI. SELECTED DISCRIMINATIVE FEATURES

T_1	T_3	T_7
0.3219	0.3989	0.4402

value indicates higher relevant feature with respect to the class and contributes with more information. The higher entropy value indicates less contribution to the class. Thus, we select lower entropy values for categorization.

In next step, we select the number of features based on Threshold value (r). The Threshold value (r) is used to prune the feature subset. The r value is selected after multiple iterations by considering distinct values for r times and we found that r =3 is the most suitable value for the given illustration. Here r value is less than total number of features. The selected discriminative features are presented in Table XVI.

Finally, the selected feature subset is considered as the input to the classifiers to categorize the text documents

V. EXPERIMENTS

This section presents detailed experimentations carried out to demonstrate the efficiency of the proposed method.

A. Datasets Description

In order to assess the effectiveness of the Intuitionstic Fuzzy Entropy-Feature Selection (IFE-FS) method, experiments are conducted on 20-Newsgroups [63], Reuters-21578 [64] and TDT2 [65] standard benchmark datasets. The 20-Newsgroups dataset is a set of 18846 online newsgroup documents, split evenly into 20 different classes with 26214 features. The Reuters-21578 dataset is from newswire, which contains 8293 documents and is non-uniformly divided into 65 categories with 18933 features. The TDT2 (Topic Detection and Tracking) dataset consists of 9394 documents and 36771 features, spread across 30 categories. In each dataset, distribution of features is varied according to their corresponding classes.

B. Experimental Setup

During the experimentation, it is necessary to split the dataset into training and testing set (to validate the proposed method). The large training data results in overfitting of the model. On the other hand, small training data results in underfitting the model. The whole reason for split comes from the fact that, we often have limited and finite data. So we want to make the best use of it and train on as much data as we can. In this paper, to validate the proposed IFE-FS method, we split the dataset into training and testing set in 60:40 ratios respectively. The training set is 60% documents of each class of dataset, used to build our proposed method. On the other hand, 40% testing set is applied on proposed model to assess the performance.

The performance of the proposed method mainly depends on the IFCM parameter values. Authors in [31] investigated the parameter values of IFCM. Based on the backdrop of [31], in this paper we initialized the parameters as follows: fuzzy coefficient p =2, convergence criteria $\varepsilon = 0.0001$ and λ =0.5. The performance of the IFE-FS method is compared against the four widely used feature selection methods viz., Chi-Square, Mutual Information (MI), Information Gain (IG) and Entropy based Feature Selection (EFS). We conducted experiments on three standard benchmark datasets using KNN, SVM, and RBF-NN classifiers. We used the F-Measure metric to assess the performance of the classifiers. The F-Measure is widely used in text categorization, which indicates the overall categorization performance and also combined effectiveness measure determined by precision and recall.

A number of features are selected based on the threshold value (r), where r indicates the number of features. Initially, we conducted experiments by fixing r value as 500 empirically. Further, we varied the value of r from 500 to 7000, with an increment of 500. However, decrease in the value of r below 500 and increase in value of r above 7000, does not yield good results. Hence, we restricted the value of r between 500 to 7000.

C. Experimental Results

Initially, we conducted the experiment without employing feature selection method. The number of features obtained for 20-Newsgroups is 26214, for Reuters-21578 there is 18933 and TDT2 has 36771 features. The F-measure using KNN, SVM and RBF-NN classifiers on three standard benchmark datasets are presented in Table XVII.

Fig. 1 depicts the categorization performance of the proposed method (IFE-FS) using KNN classifier in terms of F- measure on 20-NewsGroups dataset. It can be observed that each feature selection method obtained their best results with variant number of features, in terms of F-measure. The IFE-FS method achieved better result of 0.662 for 2500 features compared to other feature selection methods using KNN classifier. Our proposed feature selection method identifies discriminative features when the value of r is 2500, which leads to achieve the maximum result. Similarly, we conducted the same set of experiments using SVM and RBF-NN classifiers on 20-NewsGroups.

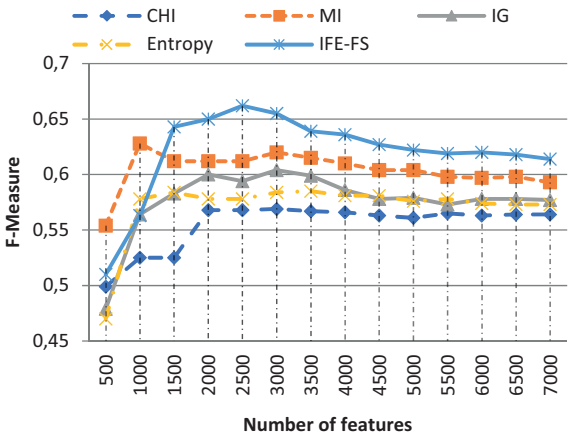


Fig.1. Performance comparisons using KNN classifier (20-NewsGroups dataset).

Fig. 2 and Fig. 3 show the performance comparison of the IFE-FS method with existing feature selection methods on 20-NewsGroups for SVM & RBF-NN classifier. The proposed IFE-FS method performed better compared to existing feature selection methods, when the number of features are ranging from 4500 to 7000. In Fig. 3, the performance of RBF-NN based on MI, IG and Entropy, completely coincides with each other. However, the F-measure curve of RBF-NN based on the proposed IFE-FS method is significantly higher than that of the existing feature selection methods. It is evident from Fig, 1, 2 and 3 that the performance of IFE-FS method in terms of F-measure is superior to that of the existing feature selection methods when value of r is 2500 for KNN, 5000 for SVM and 5500 for RBF-NN classifiers.

Further, the same set of experimentations were carried out on Reuters-21578 and TDT2 dataset. In Reuters-21578, the performance of the proposed IFE-FS method with that of the existing feature selection methods using KNN, SVM and RBF-NN classifiers are shown in Fig. 4, 5 and 6, respectively. It can be observed from Fig. 4-6 that the performance of the proposed IFE-FS method is lower than other feature selection methods, when the number of features ranging from 500 to 2500.

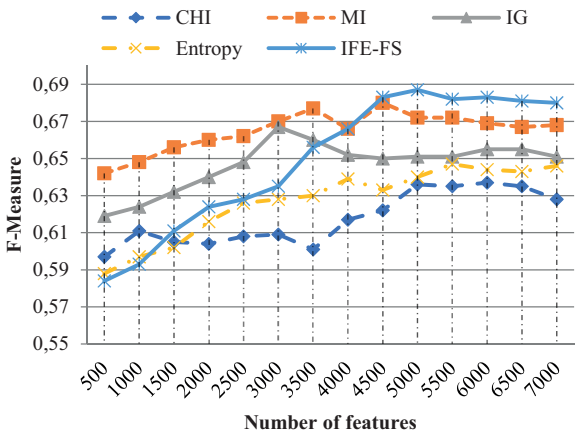


Fig. 2. Performance comparisons using SVM classifier (20-NewsGroups dataset).

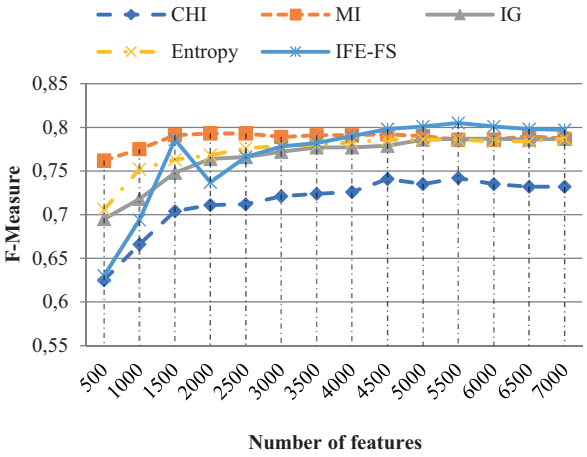


Fig. 3. Performance comparisons using RBF-NN classifier (20-NewsGroups dataset).

From Fig. 4-6, we can note that the performance of IFE-FS method in terms of F-measure is significantly higher than other existing feature selection methods, when the value of r is ranging from 2500 to 7000. From Fig. 4, 5 and 6, the KNN classifier obtains a F-measure of 0.711 when r =2500, the SVM classifier obtains a F-measure of 0.751 when r =2500 and RBF-NN classifier obtains a F-measure of 0.874 when r =6000.

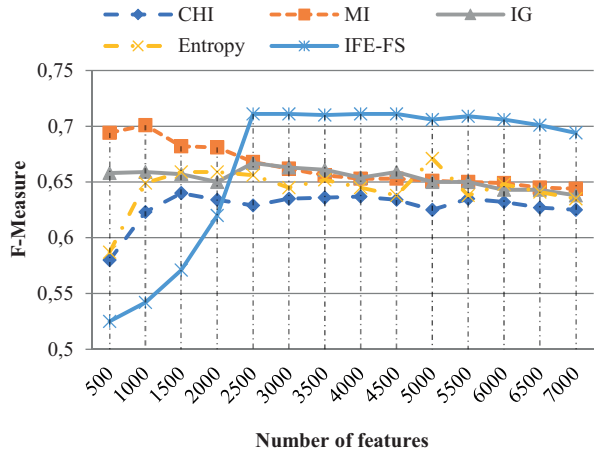


Fig. 4. Performance comparisons using KNN classifier (Reuters-21578 dataset).

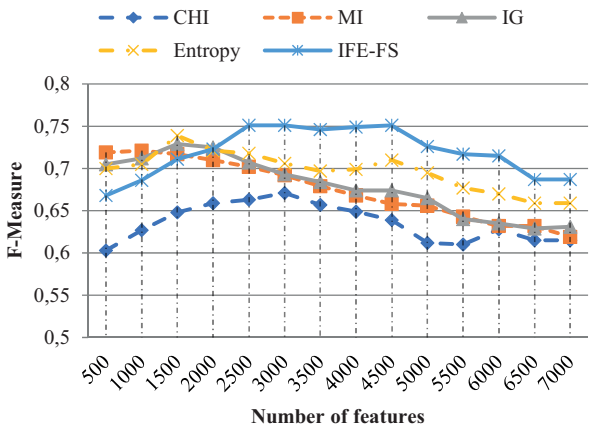


Fig. 5. Performance comparisons using SVM classifier (Reuters-21578 dataset).

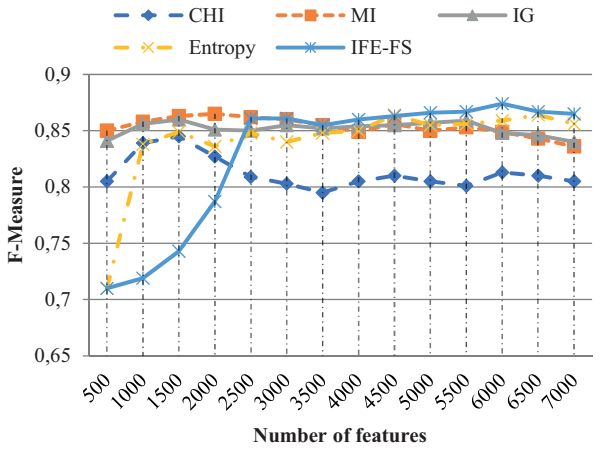


Fig. 6. Performance comparisons using RBF-NN classifier (Reuters-21578 dataset).

Fig. 7, 8 and 9 show the comparison of the proposed Intuitionstic Fuzzy Entropy-Feature Selection (IFE-FS) method with the existing feature selection method using KNN, SVM and RBF-NN classifiers on TDT2 dataset. It can be observed from Fig. 7-9, that the proposed IFE-FS method outperforms other existing feature selection methods. Fig. 7 shows that F-measure curve of KNN classifier based on the proposed IFE-FS method is higher than that of the other existing feature selection methods, when the value of r is ranging from 2500 to 6000. Similarly, the proposed IFE-FS method achieved best result when value of r is ranging from 3000 to 6000 using SVM. From Fig. 9, we can observe that the proposed IFE-FS method achieved best result when the value of r is in the range of 5500 and 6000 using RBF-NN classifier. The proposed IFE-FS method obtained a result of 0.837 using KNN classifier when the value of r is 3500, SVM achieved 0.915 result when the value of r is 3500 and RBF-NN achieved 0.967 result when the value of r is 5500, which are presented in Fig. 7, 8 and 9.

From Fig. 1-9, we can infer that, in terms of F-measure, the performance of IFE-FS method improved when the value of r is ranging from 2500 to 6000. In addition, the proposed IFE-FS method identifies discriminative features when the value of r is ranging from 2500 to 6000, which lead to achieve maximum result for all three classifiers. Table XVII presents the F-measure results using KNN, SVM and RBF-NN classifiers without feature selection method on the three datasets. From Table XVII and Fig. 1-9, we can observe that the result of the proposed IFE-FS method with all three classifiers is superior than the results of the same classifier without feature selection.

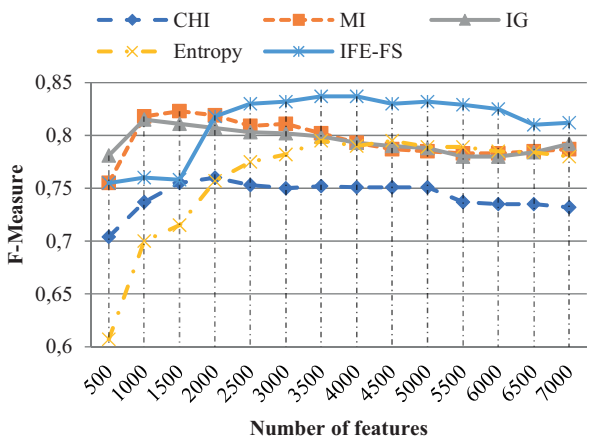


Fig. 7. Performance comparisons using KNN classifier (TDT2 dataset).

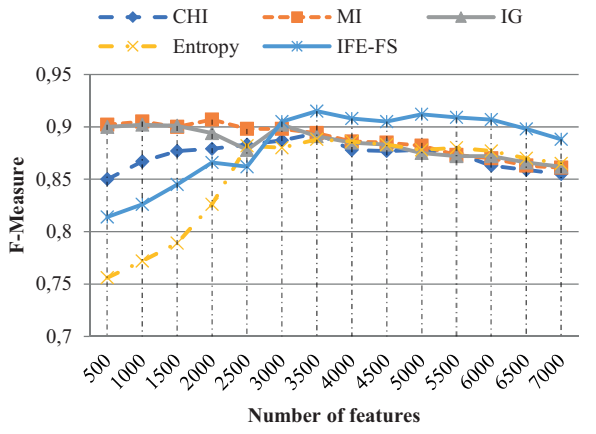


Fig. 8. Performance comparisons using SVM classifier (TDT2 dataset).

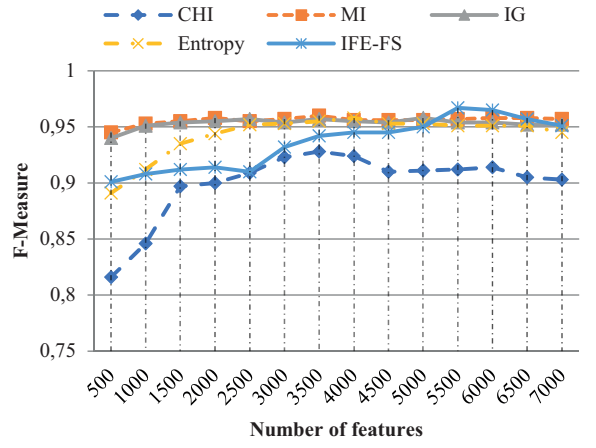


Fig. 9. Performance comparisons using RBF-NN classifier (TDT2 dataset).

Besides, the following observations were made during the experimentation:

- From experimental results, it is noted that the IFE-FS method obtained better results by considering less number of features from the original feature set.
- From Fig. 1-9, the proposed IFE-FS method achieved significantly better results on all standard benchmark datasets using KNN, SVM and RBF-NN classifiers.

TABLE XVII. Performance of Classifiers with and Without Feature Selection Method on Three Datasets

Dataset	Classifiers	Without Feature Selection		With Feature Selection	
		Number of Features Selected	F-Measure	Number of Features Selected	F-Measure
20-Newsgroups	KNN	26214	0.469	2500	0.662
	SVM	26214	0.618	5000	0.687
	RBF-NN	26214	0.716	5500	0.805
Reuters-21578	KNN	18933	0.484	2500	0.711
	SVM	18933	0.570	2500	0.751
	RBF-NN	18933	0.810	6000	0.874
TDT2	KNN	36771	0.745	3500	0.837
	SVM	36771	0.745	3500	0.915
	RBF-NN	36771	0.891	5500	0.967

D. Discussions

The proposed IFE-FS method performance is measured in terms of F-Measure. It is evident from Fig, 1-9 that the proposed IFE-FS method performs superior compared to Chi-Square, MI, IG and Entropy based Feature Selection (EFS) methods using KNN, SVM and RBF-NN classifiers on the three standard datasets. The Chi-Square, estimates the lack of independence between terms and class. The proposed IFE-FS method computes the distribution of terms in each class. Thus, the proposed IFE-FS method obtained higher result than Chi-Square on all the three datasets. The MI estimates the mutual dependency of two terms using joint probability distribution and marginal probability distribution. The proposed IFE-FS method is non-probabilistic, which estimates the intuitionistic fuzzy information by intuitionistic fuzzy set. IG evaluates the quantity of bits of information acquired by knowing the presence or absence of a term in the document for categorization prediction. However, the proposed IFE-FS method does not consider absence or presence of term in a document.

On the other hand, the EFS method considers only randomness uncertainty and estimates the entropy of the terms in a class, based on probability. However, the proposed IFE-FS method considers uncertainty like randomness, ambiguity and vagueness to provide the terms with importance in the class. Moreover, it maximizes the discriminative capacity of the features and produces rich information. The proposed IFE-FS method shows higher performance using RBF-NN classifier compared to other KNN and SVM classifier. It can be concluded from the experiments that the proposed IFE-FS method shows significant improved performance using classifiers by selecting a number of discriminative features based on the distribution of terms in the classes. The proposed IFE-FS method performs better compared to other well known feature selection methods. However, the performance of the proposed method mainly depends on the number of features, initial cluster centers and parameter values. These parameter values differ across the datasets and it makes very difficult task to select the parameter values.

VI. Conclusion

In this article, a new feature selection method called Intuitionistic Fuzzy Entropy-Feature Selection (IFE-FS) is proposed for text categorization. The IFE is based on the intuitionistic fuzzy set. Unlike traditional Fuzzy Entropy, which considers only membership degree, Intuitionistic Fuzzy Entropy considers non-membership and hesitation degree along with membership degree. Thus, it handles uncertainty to better extent. To evaluate the performance of proposed IFE-FS method, extensive experimentations were conducted on three standard benchmark datasets: 20-NewsGroups, Reuters-21578 and TDT2. From the experiments, we can conclude that the IFE-FS method reduces high dimensionality of feature matrix and enhances the performance of classifier. It can also be concluded that IFE-FS selects good subsets of features, which contains most discriminative features. The experimental results reveal that the IFE-FS method outperformed other feature selection methods in terms of F-measure.

In future, we intend to use optimization technique in the proposed IFE-FS method to automatically select the threshold value r. Additionally, we intend to add the kernel method in IFCM, which could improve the categorization performance.

References

[1] J. H. Du, "Automatic text classification algorithm based on Gauss improved convolutional neural network", *Journal of Computational Science*, *21*, pp.195-200, 2017.

[2] M. Kang, J. Ahn and K. Lee, "Opinion mining using ensemble text hidden Markov models for text classification", *Expert Systems with Applications*, *94*, pp.218-227, 2018.

[3] A. M. Jalil, I. Hafidi, L. Alami and E. Khouribga, "Comparative study of clustering algorithms in text mining context", *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(7), pp.42-45, 2016.

[4] A. Moreno and T. Redondo, "Text analytics: The convergence of big data and artificial intelligence", *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(6), pp.57-64, 2016.

[5] A. Onan, S. Korukoğlu and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification", *Expert Systems with Applications*, *57*, pp.232-247, 2016.

[6] A. K. Uysal and S. Gunal, "The impact of preprocessing on text classification", *Information Processing & Management*, *50*(1), pp.104-112, 2014.

[7] J. Huang, G. Li, Q. Huang and X. Wu, "Joint feature selection and classification for multilabel learning", *IEEE Transactions on Cybernetics*. 2017.

[8] D. Agnihotri, K. Verma and P. Tripathi, "Variable Global Feature Selection Scheme for automatic classification of text documents", *Expert Systems with Applications*, *81*, pp.268-281, 2017.

[9] M. Labani, P. Moradi, F. Ahmadizar and M. Jalili, "A novel multivariate filter method for feature selection in text classification problems", *Engineering Applications of Artificial Intelligence*, *70*, pp.25-37, 2018.

[10] Y. Lu and Y. Chen, "A Text Feature Selection Method Based on the Small World Algorithm", *Procedia Computer Science*, *107*, pp.276-284, 2017.

[11] B. Tang, S. Kay and H. He, "Toward optimal feature selection in naive Bayes for text categorization", *IEEE Transactions on Knowledge and Data Engineering*, *28*(9), pp.2508-2521, 2016.

[12] A. K. Uysal, "An improved global feature selection scheme for text classification", Expert Systems with Applications, 43:82–92, 2016.

[13] D. Zhu and K. W. Wong, "An evaluation study on text categorization using automatically generated labeled dataset", *Neurocomputing*, *249*, pp.321-336, 2017.

[14] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization", In *ICML,* Vol. 97, pp. 412-420, 1997.

[15] A. Rehman, K. Javed, and H. A. Babri, "Feature selection based on a normalized difference measure for text classification", *Information Processing & Management*, *53*(2), pp.473-489, 2017.

[16] L. P. Jing, H. K. Huang and H. B. Shi, "Improved feature selection approach TFIDF in text mining", In *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on,* Vol. 2, pp. 944-946, 2002. IEEE

[17] Y. Xu, G. J. Jones, J. Li, B. Wang and C. Sun, "A study on mutual information-based feature selection for text categorization". *Journal of Computational Information Systems*, *3*(3), pp.1007-1012, 2007.

[18] C. Lee and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization", *Information processing & management*, *42*(1), pp.155-165, 2006.

[19] S. S. Mengle and N. Goharian, "Ambiguity measure feature-selection algorithm", *Journal of the American Society for Information Science and*

Technology, *60*(5), pp.1037-1050, 2009.

[20] M. Lan, S. Y. Sung, H. B. Low and C. L. Tan, "A comparative study on term weighting schemes for text categorization", In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on,* Vol. 1, pp. 546-551, 2005. IEEE

[21] M. B. Revanasiddappa, B. S. Harish and S. Manjunath, Document classification using Symbolic classifiers", In *Contemporary Computing and Informatics (IC3I), 2014 International Conference on,* pp. 299-303, November 2014. IEEE

[22] C. Largeron, C. Moulin and M. Géry, "Entropy based feature selection for text categorization", In *Proceedings of the 2011 ACM Symposium on Applied Computing,* pp. 924-928, 2011. ACM

[23] C. E. Shannon, "A mathematical theory of communication", *ACM SIGMOBILE Mobile Computing and Communications Review*, *5*(1), pp.3-55, 2001.

[24] S. V. Vaseghi, "*Advanced digital signal processing and noise reduction*", 2008. John Wiley & Sons

[25] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory", *Information and control*, *20*(4), pp.301-312, 1972.

[26] H. M. Lee, C. M. Chen, J. M. Chen and Y. L. Jou, "An efficient fuzzy classifier with feature selection based on fuzzy entropy", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *31*(3), pp.426-432, 2001.

[27] C. Palanisamy and S. Selvan, "Efficient subspace clustering for higher dimensional data using fuzzy entropy", *Journal of Systems Science and Systems Engineering*, *18*(1), pp.95-110, 2009.

[28] X. Liu and W. Pedrycz, "*Axiomatic fuzzy set theory and its applications",* Vol. 244, 2009. Heidelberg: Springer

[29] L. A. Zadeh, "Probability measures of fuzzy events". *Journal of mathematical analysis and applications*, *23*(2), pp.421-427, 1968.

[30] K. T. Atanassov, "Intuitionistic fuzzy sets", *Fuzzy Sets and Systems*, *20*(1), pp.87-96, 1986.

[31] T. Chaira, "A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images", *Applied Soft Computing*, *11*(2), pp.1711-1717, 2011.

[32] K. Atanassov, "Intuitionistic Fuzzy Sets: Theory and Applications", 1999. Springer, Heidelberg

[33] B. S. Harish, D. S. Guru and S. Manjunath, "Representation and classification of text documents: A brief review", *IJCA, Special Issue on RTIPPR (2)*, pp.110-119, 2010.

[34] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features", Machine Learning: ECML-98. pp.137-142, 1998.

[35] E. P. Jiang, "Semi-supervised text classification using RBF networks", In *International Symposium on Intelligent Data Analysis,* pp. 95-106, August 2009. Springer Berlin

[36] S. Goswami, A. K. Das, A. Chakrabarti and B. Chakraborty, "A feature cluster taxonomy based feature selection technique", *Expert Systems with Applications*, *79*, pp.76-89, 2017.

[37] H. T. Ng, W. B. Goh and K. L. Low, "Feature selection, perceptron learning, and a usability case study for text categorization", In *ACM SIGIR Forum* (Vol. 31, No. SI, pp. 67-73, July 1997. ACM

[38] F. Sebastiani, "Machine learning in automated text categorization", *ACM computing surveys (CSUR)*, *34*(1), pp.1-47, 2002.

[39] S. Vora and H. Yang, "A comprehensive study of eleven feature selection algorithms and their impact on text classification", In *Computing Conference,* pp. 440-449, 2017.

[40] B. Tang, S. Kay and H. He, "Toward optimal feature selection in naive Bayes for text categorization", *IEEE Transactions on Knowledge and Data Engineering*, 28(9), pp.2508-2521, 2016.

[41] J. Cai and F. Song, "Maximum entropy modeling with feature selection for text categorization", In *Asia Information Retrieval Symposium*, pp. 549-554, January 2008. Springer Berlin Heidelberg

[42] V. B. Vaghela, K. H. Vandra and N. K. Modi, "Entropy based feature selection for multi-relational Naive Bayesian Classifier", *Journal of International Technology and Information Management*, *23*(1), p.2, 2014.

[43] W. Liu and N. Song, "A fuzzy approach to classification of text documents", *Journal of Computer Science and Technology*, *18*(5), pp.640-647, 2003.

[44] J. L. Fan and Y. L. Ma, "Some new fuzzy entropy formulas", *Fuzzy sets and Systems*, *128*(2), pp.277-284, 2002.

[45] R. N. Khushaba, A. Al-Jumaily and A. Al-Ani, "Novel feature extraction method based on fuzzy entropy and wavelet packet transform for myoelectric control", In *Communications and Information Technologies, 2007. ISCIT'07. International Symposium on,* pp. 352-357, October 2007. IEEE

[46] O. Parkash, P. K. Sharma and R. Mahajan, "New measures of weighted fuzzy entropy and their applications for the study of maximum weighted fuzzy entropy principle", *Information Sciences*, *178*(11), pp.2389-2395, 2008.

[47] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier", *Expert Systems with Applications*, *38*(4), pp.4600-4607, 2011.

[48] F. Ahmadizar, M. Hemmati and A. Rabanimotlagh, "Two-stage text feature selection method using fuzzy entropy measure and ant colony optimization", In *Electrical Engineering (ICEE), 2012 20th Iranian Conference on,* pp. 695-700, May 2012. IEEE

[49] M. Sugeno and T. Terano, "A model of learning based on fuzzy information", *Kybernetes*, *6*(3), pp.157-166, 1977.

[50] R. R. Yager, "On the measure of fuzziness and negation part I: membership in the unit interval", *Int. J. Gen. Syst.* **5**, pp.221–229, 1979.

[51] P. Intarapaiboon, "An application of intuitionistic fuzzy sets in text classification", In *Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on,* Vol. 1, pp. 604-608, April 2014. IEEE

[52] Y. Li, D. L. Olson and Z. Qin, "Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis", *Pattern Recognition Letters*, *28*(2), pp.278-285, 2007.

[53] G. A. Papakostas, A. G. Hatzimichailidis and V. G. Kaburlasos, "Distance and similarity measures between intuitionistic fuzzy sets: A comparative analysis from a pattern recognition point of view", *Pattern Recognition Letters*, *34*(14), pp.1609-1622, 2013.

[54] E. Szmidt and J. Kacprzyk, "Using intuitionistic fuzzy sets in text categorization", In *International Conference on Artificial Intelligence and Soft Computing,* pp. 351-362, 2008. Springer, Berlin, Heidelberg

[55] H. M. Zhang, Z. S. Xu and Q. Chen, "On clustering approach to intuitionistic fuzzy sets", *Control and Decision*, *22*(8), p.882, 2007.

[56] D. Iakovidis, N. Pelekis, E. Kotsifakos and I. Kopanakis, "Intuitionistic fuzzy clustering with applications in computer vision", In *Advanced Concepts for Intelligent Vision Systems* (pp. 764-774), 2008. Springer Berlin/Heidelberg

[57] Z. Xu, J. Chen and J. Wu, "Clustering algorithm for intuitionistic fuzzy sets", *Information Sciences*, *178*(19), pp.3775-3790, 2008.

[58] Z. Xu and J. Wu, "Intuitionistic fuzzy C-means clustering algorithms", *Journal of Systems Engineering and Electronics*, *21*(4), pp.580-590, 2010.

[59] E. Szmidt and J. Kacprzyk, "Entropy for intuitionistic fuzzy sets", *Fuzzy Sets and Systems*, *118*(3), pp.467-477, 2001.

[60] P. Burillo and H. Bustince, "Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets". *Fuzzy Sets and Systems*, *78*(3), pp.305-316, 1996.

[61] W. L. Hung and M. S. Yang, "Fuzzy entropy on intuitionistic fuzzy sets", *International Journal of Intelligent Systems*, *21*(4), pp.443-451, 2006.

[62] I. K. Vlachos and G. D. Sergiadis, "Intuitionistic fuzzy information–applications to pattern recognition", *Pattern Recognition Letters*, *28*(2), pp.197-206, 2007.

[63] 20newsgroups: http://people.csail.mit.edu/jrennie/20Newsgroups/

[64] Reuters-21578: http://www.daviddlewis.com/resources/testcollections/reuters21578/

[65] X. Hu, X. Zhang, C. Lu, E. K. Park and X. Zhou, "Exploiting Wikipedia as external knowledge for document clustering", In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 389-396, 2009. ACM

M. B. Revanasiddappa

He received his B.E degree in Information Science and Engineering and M.Tech degree in Software Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India. He is currently pursuing Ph.D degree in Computer Science from Visvesvaraya Technological University, Belagavi, Karnataka, India. His area of research includes Machine Learning, Text Mining and Soft Computing.

B. S. Harish

He obtained his B.E in Electronics and Communication (2002), M.Tech in Networking and Internet Engineering (2004) from Visvesvaraya Technological University, Belagavi, Karnataka, India. He completed his Ph.D. in Computer Science (2011); thesis entitled "Classification of Large Text Data" from University of Mysore. He is presently working as an Associate Professor in the Department of Information Science & Engineering, JSS Science & Technology University, Mysuru. He was invited as a Visiting Researcher to DIBRIS - Department of Informatics, Bio Engineering, Robotics and System Engineering, University of Genova, Italy from May-July 2016. He delivered various technical talks in National and International Conferences. He has invited as a resource person to deliver various technical talks on Data Mining, Image Processing, Pattern Recognition, Soft Computing. He is also serving and served as a reviewer for National, International Conferences and Journals. He has published more than 50 International reputed peer reviewed journals and conferences proceedings. He successfully executed AICTE-RPS project which was sanctioned by AICTE, Government of India. He also served as a secretary, CSI Mysore chapter. He is a Member of IEEE (93068688), Life Member of CSI (09872), Life Member of Institute of Engineers and Life Member of ISTE. His area of interest includes Machine Learning, Text Mining and Computational Intelligence.

An Adapted Approach for User Profiling in a Recommendation System: Application to Industrial Diagnosis

Fatima Zohra Benkaddour [1]*, Noria Taghezout [2], Fatima Zahra Kaddour-Ahmed [3], Ilyes-Ahmed Hammadi [3]

[1] Laboratoire d'Informatique Oran (LIO), École Normale Supérieure d'Oran (ENSO), Département des Sciences Exactes, filière Informatique, Oran (Algeria)
[2] Laboratoire d'Informatique Oran (LIO), Département d'Informatique, Université Oran1, Oran (Algeria)
[3] Département d'Informatique, Université Oran1, Oran (Algeria)

Received 15 January 2018 | Accepted 31 May 2018 | Published 12 June 2018

Abstract

In this paper, we propose a global architecture of a recommender tool, which represents a part of an existing collaborative platform. This tool provides diagnostic documents for industrial operators. The recommendation process considered here is composed of three steps: Collecting and filtering information; Prediction or recommendation step; evaluating and improvement. In this work, we focus on collecting and filtering step. We mainly use information result from collaborative sessions and documents describing solutions that are attributed to the complex diagnostic problems. The developed tool is based on collaborative filtering that operates on users' preferences and similar responses.

Keywords

Recommender Systems, Decision Support System, Industrial Diagnosis, Collaborative Filtering, Twitter.

DOI: 10.9781/ijimai.2018.06.003

I. Introduction

With the increasing volume of data that exist in information systems, the user can quickly be submerged by an informative mass. Moreover, these data are highly heterogeneous. As a result, targeting relevant information to the user becomes a major concern of a large number of research projects. Therefore, personalization is a suitable solution to this problem. Among the personalization tools, we find the recommendation systems.

Recommendations systems are defined by [1] as systems that can provide customized recommendations to guide the user to interesting and useful resources within an important data space. They play a major role in the information filtering systems dealing with how best it can recommend items or information which is relevant to the user. Recommendation systems can be applied to a variety of applications such as E-Commerce site.

The world is developing daily especially in the industrial diagnostics field. It becomes necessary to support the diverse diagnoses of the industry in order to achieve better performance and a sustainable and profitable industrial efficiency. In any industrial process, it is essential to establish the relative performance with the improvement and development of new technologies.

Generally, the industrial diagnosis helps to promote and develop expertise for a better result and an achievement of the objectives set by the company. Furthermore, industrial documents help to promote and develop diversity and knowledge.

The industrial domain is a field with different orientations and different sectors. It is preferable to move towards a recommendation system that gives to the user the way to acquire any documentation relating to his domain in a reduced time and without difficulties.

A. Problem Statement

The search for solutions to the breakdowns in the industrial environment is a difficult task that requires considerable research time. Finding these solutions in a short time will improve the company productivity. The work presented here is a part of collaborative decision support system [2]. This system is the first system that provides answers to non-woven operators for their diagnosis problems by using a domain ontology, which represents the knowledge source and case-based reasoning.

Due to the availability of industrial documents that describe most solutions to the complex problems of industrial operators arising from collaborative work; a real need for an information filtering tool was felt. The information gathered from collaborative sessions and Web 2.0 tools facilitate the development project of a recommendation system. In this context, the main objective of this work is to improve the search for diagnostic documents for an industrial operator by taking his preferences into account in order to provide better quality in the recommendations.

B. Contribution

The recommendation system in the industrial field has become the main focus for the development and efficiency of the company.

* Corresponding author.
E-mail address: benkaddourfatima@gmail.com

Our study consists of incorporating a collaborative support system with a documents recommender tool. The main goal of this tool is to provide relevant documents to industrial operators in a fast and simple way. For that, we consider a recommendation process of three steps which are: collecting and filtering information, prediction, and evaluating and improvement. In this work, we believe that identifying preferences of users allows more precision to the recommender tool. In this study, we give more importance to the first step.

The recommendation system that will plugin will have access to all the database including users, documents, and users' history data. The recommendation system will take this data as input and predict for example a new time-line for the logged in user.

We summarize our main contribution in the following:

- Implementing an extended architecture of collaborative system detailing the recommended component;
- Recognizing and collecting users 'preferences in two ways: explicitly and implicitly;
- Applying KNN (K Nearest Neighbors) method based on Cosine similarity to recommend documents after preparing and filtering users and documents data.

This paper is organized as follows: In section two, we give some works relative to recommendation systems. Besides, we also present our analysis. In section three, we detail our proposed approach by giving the global architecture of our recommendation tool and some details regarding the process of collecting and filtering information. In section four, we explain the experimental protocol. In section five, we discuss some obtained results. In section six, we present two scenarios of execution to describe in detailed the proposed approach. Finally, in section seven, we conclude with some future works.

II. Related Work

Today, as the Web is rapidly growing at a faster rate, finding relevant information has become extremely difficult. Information or content can be in any form such as music, video, images or text, which are the interest to the users. Therefore Recommendation systems come into picture. Recommendation systems are a sub-category of information filtering systems that help people to find products, correct information, and even other people as well.

There are a lot of works about filtering information and recommendation systems. We present some of them below.

In [1], authors proposed a social user profiling method to recommend online sites. These recommendations relied on some similar interest between users and their followers. The suggested approach was based on an extended matrix factorization model by incorporating both individual and shared users' interests, and multifaceted unsupervised similarities. Some experiences were conducted to show performance of this approach.

Authors presented in [3] a survey of collaborative filtering techniques. They sorted collaborative filtering into three categories: memory based, model-based, and hybrid Collaborative Filtering (CF) algorithms. These later were advanced with their advantages, limits, and solutions. Different metrics of evaluating recommender systems were given in this paper.

In [4], authors resolved the social recommendation (SR) problem by utilizing microblogging data via multi-view user preference learning. User preferences are presented as rating information, social relations, items information, and tagging information, which are the common representation of multi-view information. The items are recommended based on the learnt user preference.

Authors developed an alternating direction method of multipliers (ADMM) scheme to solve the proposed model. They evaluated their approach, by using two real world datasets.

In addition to the works cited above, authors proposed in [5] two recommendation models to solve the complete cold start (CCS) and incomplete cold start (ICS) problems for new items. These models are based on a framework of tightly coupled CF approach and deep learning neural network. They used a deep neural network SADE to extract the content features of the items, also a solution for cold start items (CSI) is provided. The problem of CSI exiting in CF model is solved by taking account of the content features and ratings into prediction.

The two proposed recommendation models are also evaluated and compared with ICS items, and a flexible scheme of model retraining and switching is proposed to deal with the transition of items from cold start to non-cold start status. Some experiments were conducted to support the proposed approach.

In [6], authors developed a multi-criteria approach for a recommender system. This later is developed in order to support decision makers in their activities by managing users 'profiles. An automated technique is used to ensure the evolution of the recommender system.

The paper cited in [7] presents a comparative study of different recommendation techniques. A Content-Based Recommendation is highlighted. This kind of recommendation makes easier to find relevant information to the user based on previous ratings and predictions. Authors gave a comparative study of different techniques; they concluded that hybrid approach will give better results.

Authors presented their work [8] as a comparison between explicit ratings methods. These methods represent users 'preferences. The most adequate method is used to rate web content, and will be utilized by any web recommendation systems.

As shown in [9], authors explored the effect of combining the implicit relationships of the items and user-item matrix on the accuracy of recommendations. They introduced Item Asymmetric Correlation (IAC), as a new method that generates the implicit item relationship based on the user-item matrix. In their work, they used relations as an additional dataset for the Matrix Factorization (MF) technique. This research considered the implicit relationship between items, the correlated items are extracted, and the new dataset is used in MF model as a regularization term.

After our analysis, we can say that a recommender system will be accurate if it takes in consideration users 'preferences. However, users 'preferences represent a significant amount of information. Using relevant information would considerably reduce research time, which would allow a good evolution of the recommender system.

We give a comparison between recommendation approaches in Table I.

Among the existing recommendation approaches, the hybrid approach remains the best.

A. Problems of Recommendation Systems

Recommendations Systems (RSs) have emerged with the evolution of information available on the Web. RSs come to solve the problem of information overload as well as its research. However, these systems encounter performance problems. We classify these problems into three categories: in the first category, we find the lack of information on users / items which is called the cold start problem (found in collaborative filtering). This problem is defined by [33] as the inability of the system to deal with new users or new items due to the lack of prior knowledge. The second category recurs the lack of information on an item, therefore, this item will never be a subject for commendation (sparsity problem). In other terms, inactive users, who have only expressed few ratings or interacted with few items, cause data sparsity, which makes it

TABLE I. Comparison Between Recommendation Approaches

Techniques	Advantages	Limits	References
Collaborative filtering	Does not require any knowledge about the content of the item or its semantics. The quality of the recommendation can be assessed. The higher number of users' accurate the recommendation.	Cold start problem. New items are recommended only if they are already rated by users. Problem of confidentiality. Complexity: in systems with a large number of items and users, the calculation grows linearly. The number of users is relative to the quality of the recommendation provided.	[3]
Content-based recommendation	Does not need for a large community of users to make recommendations. A list of recommendations can be generated even if there is only one user. Quality grows over time. Does not need for information about other users. Considers the unique tastes of users.	User Profile Requirement. Problem of recommendation of images and videos in the absence of Metadata. Content analysis is required to make a recommendation.	[10]
Hybrid approach	It always provides predictions to content of recommendation. It improves the user preferences for suggesting items to users.	Have increased complexity and expense for implementation. Need external information is not usually available.	[3], [7], [10]

hard to make an accurate recommendation, if the system relies only on users' ratings or their interaction records [32]. The third category presents the problem of information overload. This latter can be solved by collecting relevant users 'preferences. In the literature, there are two types of information acquisition explicit and implicit feedback. The explicit way is considered as the most representative indicator of the user's interest in an item [11]. The indirect way allows to observe the users 'behavior towards a given item. As a result, users are not required to rate the items directly. This act will not have any impact on the quality of the provided recommendation. The authors [11] find necessary to capture as much information as possible without the direct intervention of users, in order to [12] better determine their interests and needs. Some research has been conducted in this area, the authors in [13] have found a way to transform users 'behavior in the recommendation platform into explicit information through the "User Interactions Converter Algorithm (UICA)". This research helped to determine users' interest by analyzing and converting their behavior.

B. Real Recommender Systems

Currently, there are wide ranges of recommendation systems that are used in different areas. Table II includes the most popular recommendation systems.

TABLE II. The most Popular Recommendation Systems

Systems	item	References
Netflix	Films	[14]
YouTube	Videos	[15]
Facebook	Persons	[16]
TripAdvisor	Hotel, restaurant	[17]
Book crossing	Books	[18]
Google scholar	Scientific articles	[19]
Amazon	Objects	[21]

III. Proposed Approach

In this section, we describe in details our approach. Fig. 1 represents the general process of our approach. In the following, we give details of each step according to the chronological order of their appearance (shown in Fig. 2).

Fig. 1. General Process of recommendation.

1. Collecting users preferences: This step allows the user to enter his / her personal data by filling out forms that include questions such as: name, first name… and professional information such as years of experience, center of interest, etc. Some information is both static and dynamic since they vary over time according to the age or experience of the user.
2. Collecting preferences via Twitter: Through this step, the user can give his permission to retrieve his social information. Generally, users are more active on their social networks allowing us to recover their preferences indirectly (implicitly). In this study, we only operate on Twitter social network. Choosing Twitter is justified by its provision of public API developers, easy to use. We used Twitter API for collecting users preferences [34]. Furthermore, Twitter has several APIs to query its database, but also to build other services. These APIs are particularly rich by returning almost a hundred variables per query; the data concerning the tweets (date of publication, the text of the message, etc.), the author (creation date of the account, pseudo ...), the

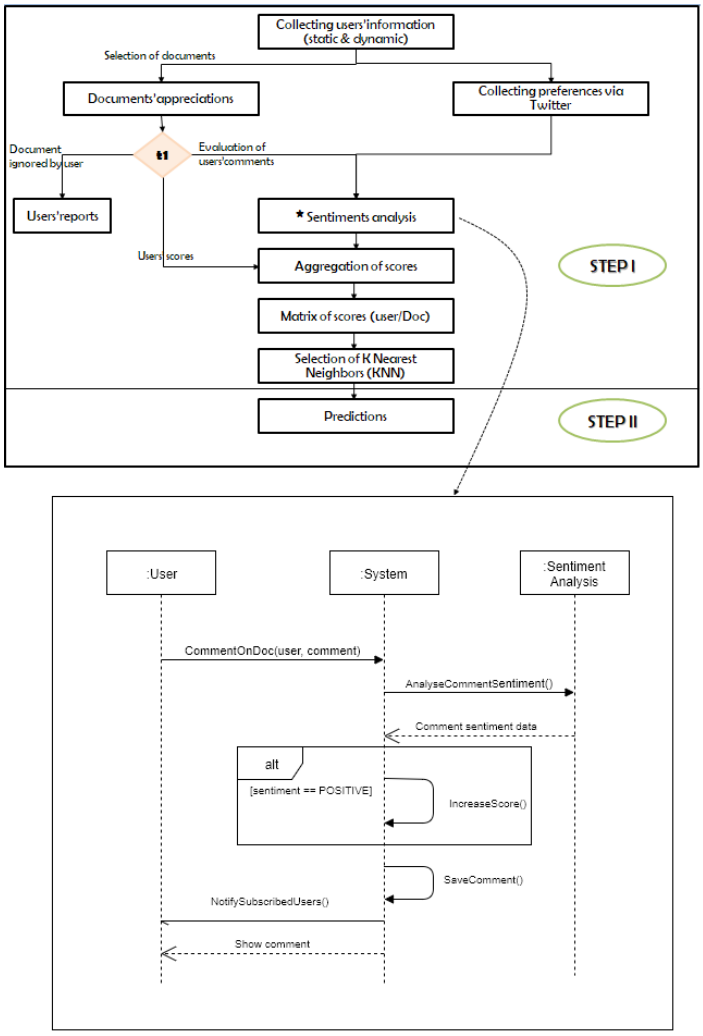


Fig. 2. Detailed process of collecting and filtering information step.

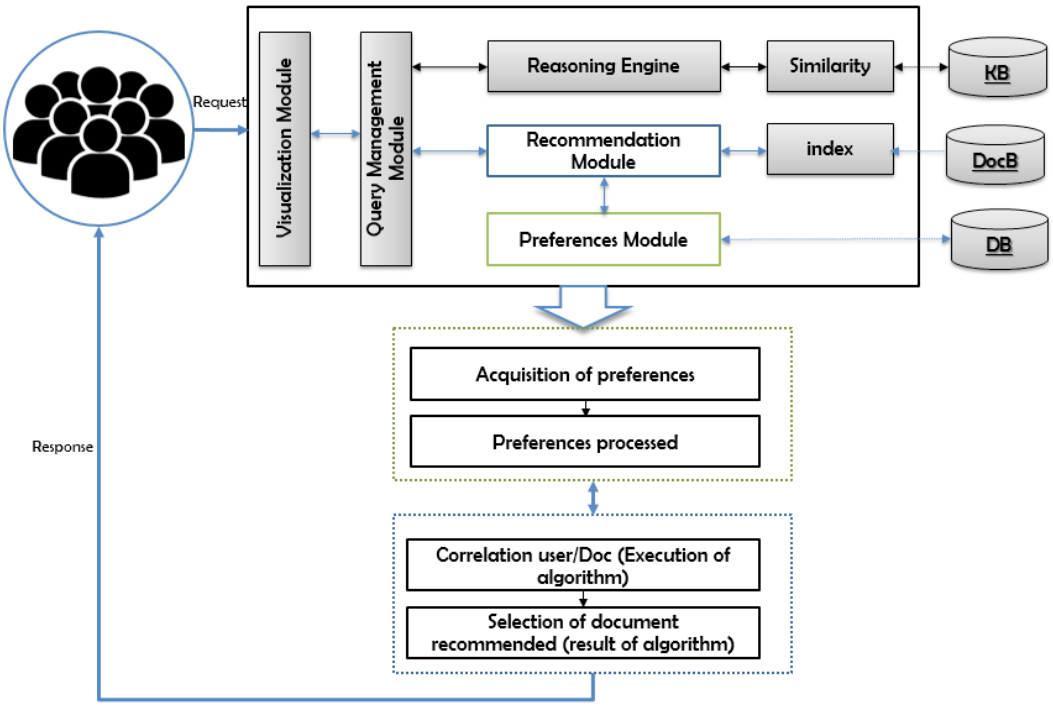


Fig. 3. A resumed overview of the collaborative system architecture.

entities contained in the messages (hashtags, mentions, urls ...) and information location (country, time zone, longitude / latitude).

Moreover, it is open source and allows analyzing users' behaviors through their activities and their posts [20]. This allows us to enrich the user profile with his behavior (We can be inspired by [20]). The user behavior will be considered as dynamic data (example: hashtag can be used as a center of interest) and consequently lead to an evolutionary recommender system over time. Behavioral study will also allow to build communities of users (cited in perspective) according to their behavior on social networks particularly on Twitter.

3. Document appreciations (document notation): the user is invited to evaluate some documents in order to better understand his preferences. The evaluation is done in two possible ways: by assigning a direct score varying from 1 to 10 and / or commenting on the proposed document.
4. Users' reports: If the document is not noted, the user is invited to answer some questions to better understand his expectations.
5. Sentiments analysis: user comments that appear on a given document are analyzed with API to detect feelings based on text processing [35] and categorized as Good where the obtained note vary between 8 to 10, Medium where it is between 5 to 7, and Null where it is between 1 to 4. These notes are utilized (average) with direct evaluation of a document, and constitute the final score of this document (as detailed in the scenario 2).The comment will be stored with the user profile (his history). The system will notify all the subscribed users (users that have also commented on this document), in when the document has a new comment. Posted comments for a given document provide feedback on the document itself. Users who give their opinions are thus bound in one way or another to the user who wants to have opinions on the document in question. Consequently, this link constitutes trusted network. In the case of industrial diagnosis, the trades people are considered to be persons of high confidence that is to say of first degree.
6. Aggregation of scores: here, the average of the notes of the documents is calculated. The matrix of scores (user / Doc) is established.
7. Selection of K Nearest Neighbors (KNN): This step allows to select the K nearest neighbors of the Active User in order to recommend relevant documents firstly. This selection is based on cosine similarity, which is a measure of similarity between users [22]. We opted for the KNN algorithm because we believe that is effective when the number of users is increasing in the platform. Our tool will only calculate the predictions between the k nearest neighbors instead of calculating them among all users. This allows a quick result. Moreover, this algorithm is easy to set up. Remember that this recommendation tool complements an existing CDSS system. The need for documents' recommendation is real.

$$\cos(u1,u2) = \frac{\sum_{i=1}^{d}(note(u1,i) * note(u2,i))}{\sqrt{\sum_{i=1}^{d}(note(u1,i))^2} * \sqrt{\sum_{i=1}^{d}(note(u2,i))^2}} \quad (1)$$

8. Calculating predictions: It represents the second phase of the recommendation process that is given in Fig. 1.

A. Global Architecture

We give below our tool architecture which is a part of a collaborative platform. Fig. 2 shows the detailed recommendation process shown in Fig. 1 (specifically Steps 1 and 2).

The first step in this process is to gather the information in mass of users in static type: name, first name ... and dynamic type such as the age, the experiment and the centers of interests.

Documents are presented for users' evaluation. This evaluation includes the score given to a document and / or the comment posted. The user can, if he wants, give access to his Twitter account.

If the document proposed to the user is not noted a questionnaire will be proposed to better understand the user's expectations (needs).

If the user posts a comment, this latter will be analyzed to explain his opinion and translated into a score through the feelings analysis step. This analysis is purely syntactic. It recognizes the terms of expressions such as good, bad, etc. We use an API for that [35].

The aggregation step groups the direct and indirect notes and normalizes them on a scale from 1 to 10. The User / Doc matrix is drawn at the end of this step.

To recommend a new document, we use the KNN to identify users with similar preferences to the active user. This method is based on Cosine measure. This latter represents the second step of Fig. 1.

Fig. 3 represents the summary architecture of the collaborative decision support system [2] for which we develop the recommendation module.

This figure zooms 2 essential modules the recommendation module and the preferences module.

Some information about the users (also preferences) of the recommendation tool developed in this paper come from this module shown in green. The blue part shows how the two modules influence each other.

All the documents of diagnostic that are stored in the documents base (DocB) have been evaluated by domain experts.

Proposed documents during a collaborative session are evaluated and corrected by the domain experts. These documents represent in the majority of times, detailed technical solutions to the problem faced by the industrial operator. We believe that the recommended documents in addition to the users' preferences influence the quality and the accuracy of the recommendation tool.

Table III summarizes some questions that allow experts to evaluate a diagnostic document.

TABLE III. Questions of Experts for Evaluating Diagnosis Documents

Id	Question
1	What is the machine part that needs to be changed?
2	What is the model of the new part?
3	What is the reference of the new part?
4	What is the machine to repair (reference)?
5	Please introduce the new solution.
6	What is the type of this problem? (Major or repetitive)
7	How is the proposed solution? (Simple to apply, complicated)
8	Do you find the proposed solution applicable?
9	Is the solution solving the problem in final way?

The diagram in Fig. 4 describes the main actions that a user can do on the platform. This latter can search by categories, tags, username or Doc name without being logged in. He can edit his profile to add more information, for example to add his Twitter social data to his profile to get more accurate recommendation based on his latest tweets. He can also view his notifications, and follow other users. The user can also access a book and like, share, Bookmark, comment, and read the document. All of these interactions are saved by the system for future uses by the recommendation system.

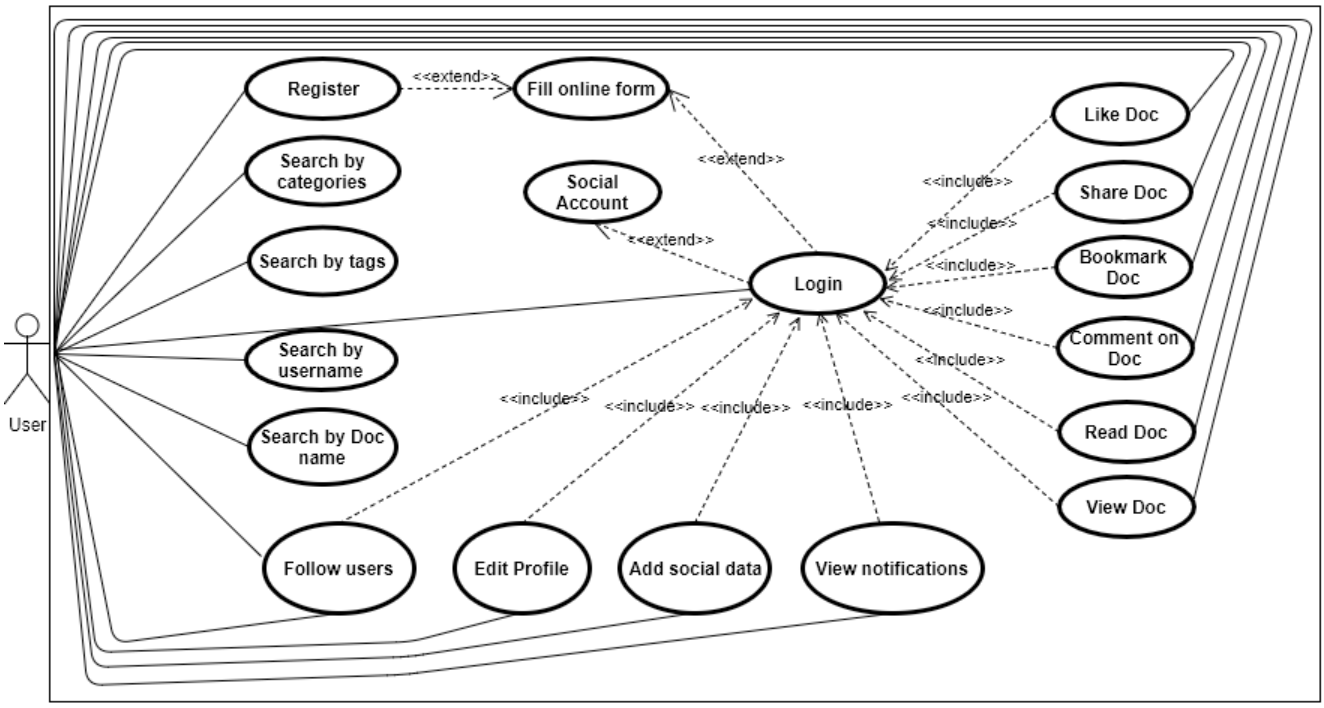


Fig. 4. User use-case diagram.

IV. Experimental Protocol

We conduct experiments "offline" (simulations) to evaluate our recommendation tool thus illustrating its efficiency, its performances and its scalability. In this section, we present the corpus of documents that will be used in the experiments of this work, as well as the evaluation measures.

A. Corpus

The experimental studies are carried out on a set of 40 industrial diagnostic documents and on the Book Crossing corpus. Book Crossing choice is justified by our motivation to analyse the behaviour of the tool developed facing a large number of documents. Book Crossing was collected in 2004 by Cao-Nicolas Ziegler from the Book Crossing community [18]. It is made up of 278,858 anonymous users who have provided 1,149,780 ratings on 271,379 books rated on a scale of 0 to 10. We have modified this corpus in such a way that it can be exploited by our tool. In book crossing the zero (0) is considered as null note. On the other hand, in our work 0 represents an unrated document. The grades assigned to the documents vary between [1, 10]. For that, we made a small modification to the notes in the book crossing corpus so that it can be manipulated by our tool.

B. Sample Used

To perform these experiments, we used a sample of 100 and 1600 votes for 40 diagnosis documents, we also used 10000 documents from Book Crossing to study the behavior of our approach. We make vary the k which is the number of the nearest neighbors according to the cosine similarity.

C. Experiences

In these experiments, we follow the example of calculation presented below. First, we calculate cosine similarity between a group of users. Second, we select K nearest neighbors and calculate predictions.

1) Example of Calculation

Table IV represents the ratings assigned by 6 users to 7 documents. Ux is a new user to whom we want to calculate the interest rate that D7 will bring.

TABLE IV. Matrix of Scores (User/Doc)

	D1	D2	D3	D4	D5	D6	D7
U1	1	2	7	1	8	10	3
U2	10	9	9	9	3	1	8
U3	7	7	8	9	10	1	1
U4	3	3	10	1	2	3	8
U5	8	9	1	8	9	1	1
U6	2	2	10	10	10	9	10
Ux	8	5	5	10	7	10	?

The principle is that users who shared the same preferences in the past are likely to share their preferences in the future.

Knowing that Ux noted D1, D2, D3, D4, D5 and D6 with respective notes 8, 5, 5, 10, 7 and 10. We use cosine measure to determine the similarity between these 7 users.

According to cosine (U, U) similarity:

Cosine (U1, Ux) = 0.777, Cosine (U2, Ux) = 0.774, Cosine (U3, Ux) =0.852, Cosine (U4, Ux) = 0.536, Cosine (U5, Ux) = 0.824, Cosine (U6, Ux) = 0.939.

The similarity threshold is set to 0.5, the similarity varies between [0, 1]. We keep the notes of the 3 nearest neighbors (in descending order) and calculate their average. The notes used in this example are the note of U6, U3 and U5 (10, 1 and 1). The average of these ratings is 3.3. As conclusion, document D7 will not be recommended to Ux.

Table V gives detailed notes according to the three nearest neighbors.

TABLE V.
Prediction Calculations According to the 3 Nearest Neighbors

	D1	D2	D3	D4	D5	D6	D7
U3	7	7	8	9	10	1	1
U5	8	9	1	8	9	1	1
U6	2	2	10	10	10	9	10
Ux	8	5	5	10	7	10	3.3

2) Evaluation with a Dataset

Below, we study the behavior of our approach with different mount of data knowing that an operator cannot evaluate the same documents many times. We use three samples: 1600 votes with 40 documents and 40 operators, 100 votes with 40 documents and 40 operators and 10000 votes of Book Crossing.

- *Sample of 1600 votes for diagnostic documents*

  Fig. 5 represents the Active user / Users similarity which varies between [0, 1]. The threshold of the similarity is set to 0.5. As shown in the figure, we notice that users with the respective id 4457, 5667, 832, 998, 593, and 8845 have similarity above the threshold, thus these users are similar to the active user. This result means that these users have shared the same preferences in the past. Hence, they are likely to share their preferences in the future.

  After the study of cosine similarity, we use KNN [22], [23] which determines the k users to calculate the prediction of the documents 'notes to be recommended to a given user.

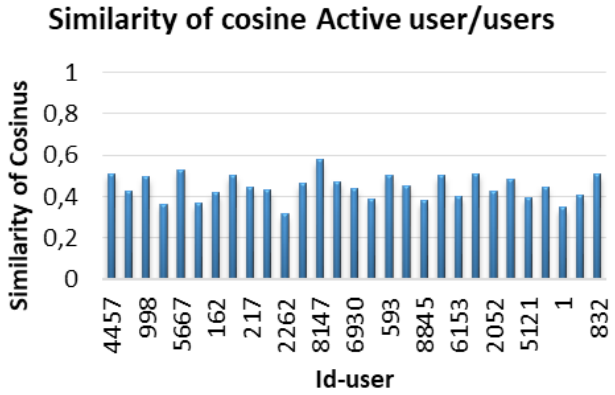


Fig. 5. Similarity of cosine calculated between the Active User and users of the sample.

We use the same set of users and documents. We give respectively the variable k the values 3, 20 and 30. The results obtained are presented in Fig. 6, 7 and 8.

K=3

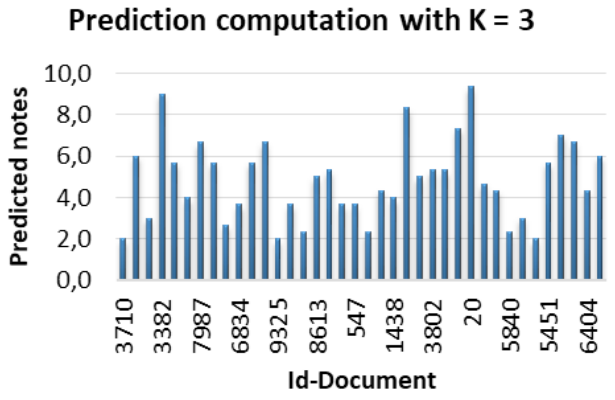


Fig. 6. Predicted notes with k=3.

K=20

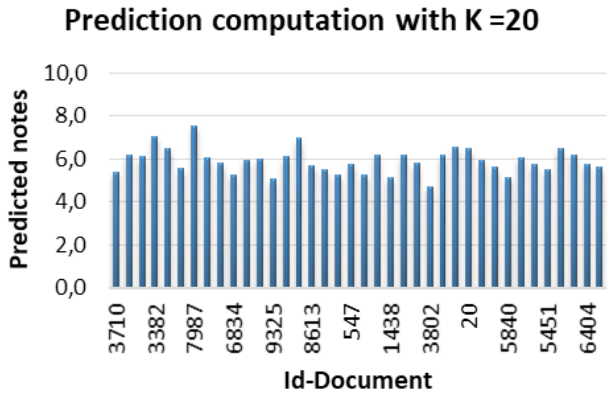


Fig. 7. Predicted notes with k=20.

K=30

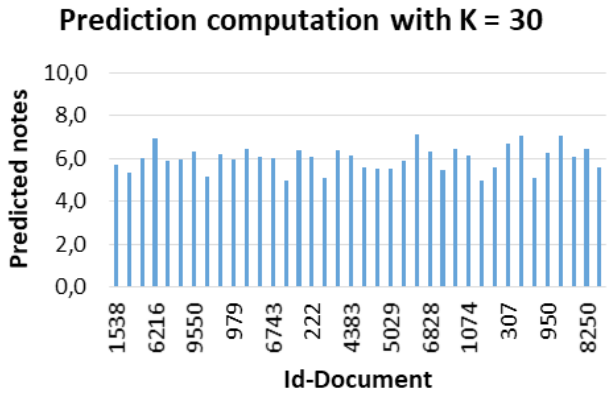


Fig. 8. Predicted notes with k=30.

Fig. 9 shows similarity between users and Active User. We see that the user with ID = 2 is more similar than others to the Active User. The sample presented here contains three (3) users and one hundred (100) votes.

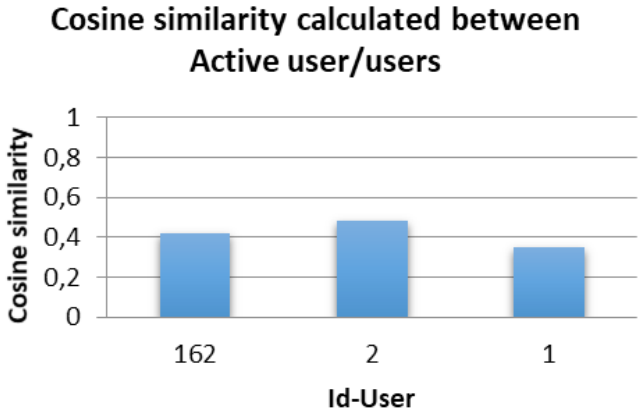


Fig. 9. Cosine similarity is calculated between the Active user and users of the sample.

Fig. 10 shows the calculated prediction for the diagnostic documents.

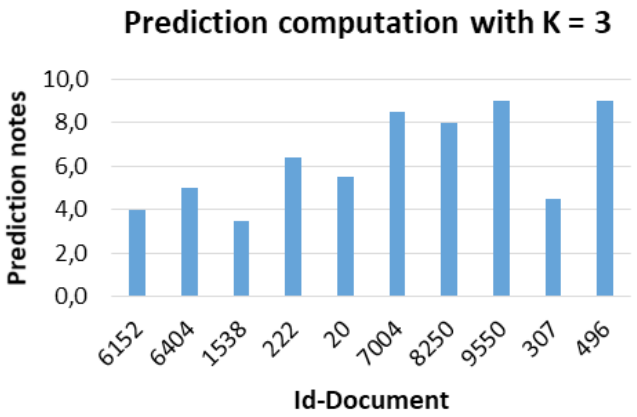


Fig. 10. Predicted notes with k = 3.

- *Sample of 10000 votes from Book Crossing*

  We conducted an experiment with 10000 votes taken from Book Crossing. The results obtained are shown in Fig. 11. We note that some values are null for some documents because Book Crossing interprets the zero as a non-given note.

  We notice that when the K increases, some documents appear as the best predicted documents. We also notice that the best documents predicted with k = 3 always remain the best predicted with k = 30 and k = 20 in the case of 1600 diagnostic documents.

D. Evaluation Measures

In this experimental protocol, we use 2 measurements of predictive evaluations MAE "Mean Absolute Error" and RSME "Root Mean Squared Error", which calculates the accuracy of the predictions against the actual assessment performed by the operator.

Let n be a set of test items, p (u; i) a note prediction of the user u for the item i and n (u; i) the actual score assigned by the user u for the item i [24].

The most commonly used measure is MAE, which is regularly used to evaluate the accuracy of a prediction. It corresponds to the mean absolute error between the actual evaluation and the prediction. The measure is calculated by the following formula:

$$MAE = \frac{1}{n}\sum_{k=0}^{n}\left|P(u,i) - n\,(u,i)\right| \tag{2}$$

The second RMSE measure raises the squared error before summing, which is useful when we want to give more criticality to the important errors [25]. The measure is calculated by the following formula:

$$RMSE = \frac{1}{n}\sum_{k=0}^{n}\sqrt{\left(P(u,i) - n\,(u,i)\right)^2} \tag{3}$$

Table VI gathers the notes assigned to the documents by a set of users as well as the predicted notes by the cosine measure while averaging the notes of the k nearest neighbors.

TABLE VI. PREDICTED AND ASSIGNED NOTES BY EACH USER

Id-user	ID-Document	Predicted Note	Attributed Note
6092	7335	7.1	7
	3382	7.5	7
	6216	7.1	7
	491	6.9	6
	3550	6.9	7
7887	7987	7.6	8
	3382	7	7
	7338	6.5	6
	8250	6.3	6
	6216	6.2	6
11111	7987	6.9	7
	3382	6.8	7
	7338	5.95	6
	8250	5.9	6
	6216	5.9	6
2222	7987	6.9	7
	3382	6.9	7
	7338	5.95	7
	8250	5.9	6
	6216	5.9	6
685	7987	6.4	7
	3382	6.35	6
	7338	6.3	6
	8250	6.2	6
	6216	6.2	6

After the calculation of MAE and RMSE presented above, we obtain the Table VII.

TABLE VII. RECOMMENDATION TOOL EVALUATION

Id-user	Real evaluation	MAE	RMSE
6092	6.8	1.1	2.13
7887	6.6	0.26	0.27
11111	6.4	0.11	0.21
2222	6.4	0.31	0.55
685	6.2	0.73	0.78

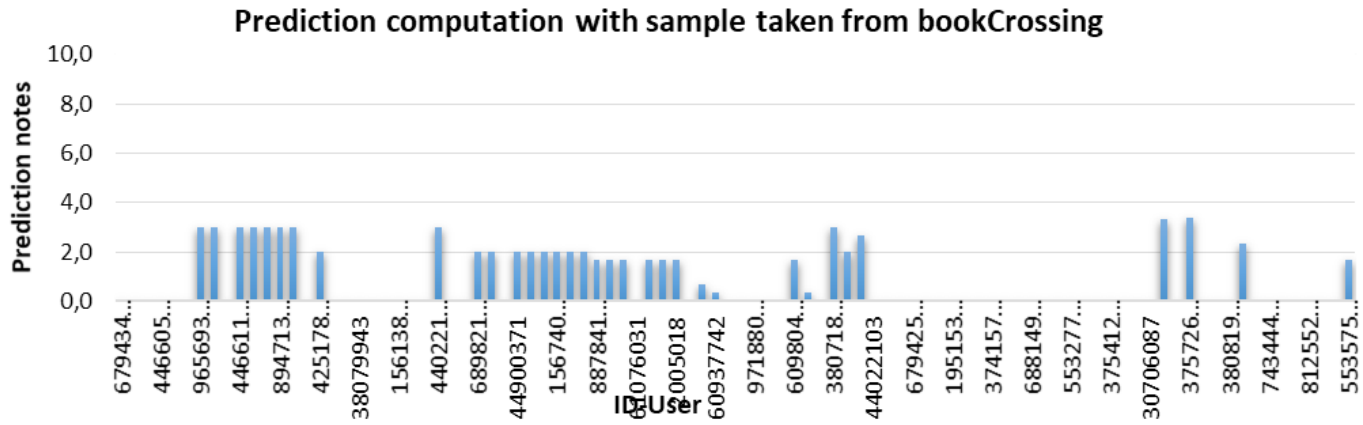


Fig. 11. Predicted notes using Book Crossing.

Fig. 12 shows the results shown in Table VII. We note that the error rate of our recommendation tool is much lower than the evaluation of the users.

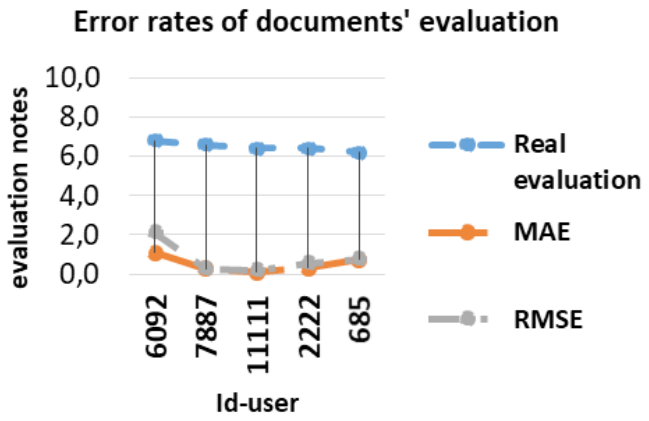


Fig. 12. Error rates of documents evaluation.

V. Results and Discussion

The data pre-processing phase that is presented in this article plays a very important role in the recommendation of the diagnostic documents proposed by the collaborative platform. Extracting relevant elements of recommendation is permitted by the pre-processing phase. This phase allows the acceptability of the provided recommendations to the users. In this phase, users preferences are discerned.

Our approach has some advantages which are summarized in the quality and speed of the recommendations that are provided to the users of the platform, and the use of users preferences especially industrial operators to avoid possible rejections. These recommendations orient to the solutions applied by users in the case of industrial diagnostics; a good recommendation in this case is translated into an effective resolution of the problem in order to allow a greater amount of production and a gain of money.

In order to demonstrate the effectiveness of the new tool and its influence in the collaborative platform, precisely in reasoning engine as shown in Fig. 3 (section 3) we use the same cases presented in [26].

We note that the reasoning engine performance is not affected only by the similarity measure but by the quality of the documents provided also. The quality for us is expressed in terms of relevance. This tool helps to increase the documents relevance by evaluating them through the domain experts as explained in section 3.We can justify this by increasing the number of relevant documents saved after integrating the recommendation tool. Besides, it will help to gain the confidence of operators and thus encourage them to reuse the tool to improve their profiles.

The objective of this evaluation is to study the impact of each prediction method on the performance of the recommendation tool. The methods studied are: FCS (Standard Collaborative Filtering) [24], D-BNCF-KNN Densified Behavioural Network based. Collaborative Filtering where only the direct neighbors [27] are involved in the calculation of the predictions, as well as our approach [28].

Table VIII shows MAE results of the compared methods. We can see that MAE is deteriorating in the case of D-BNCF-KNN. The application of the FCS brings an improvement compared to D-BNCF-KNN, we also notice that our approach is slightly than FCS.

TABLE VIII. MAE Results

Recommendation model	MAE
FCS	0.763
D-BNCF-KNN	1.074
Our work	**0.502**

VI. Scenario/Overview

In this section, we give an overview of the main Web features, mobile application for both the user and admin, and two scenarios of execution. The platform focuses on both Admin and User as the main users of the system. We use different documents which are not necessarily documents of industrial diagnosis. These documents are provided from different sources as Book Crossing.

- **Admin**: The admin logs in the system if he already has an account. He has the privilege to create, update and delete documents. He can also create, update, disable or delete other users from the platform. The admin has also the ability to see more features in the platform like the sentiment detection on users' comments. The admin has also all the features that a simple user has.
- **User**: The user logs in the system if he already has an account otherwise he creates a new one with his email that he needs to verify in order to interact with the system. After that, the user is logged in, he can navigate in the platform by reading, liking, bookmarking, sharing, and commenting on documents. He can follow other users and be followed by other users also.

A. Web Application

Fig. 13 shows an overview of sign in user. He can learn about all the platform features and choose to join or not. The user can login if he has already an Account. Otherwise, the user must provide a unique user-name that does not exists in the database with a valid email and password. After that, the user is registered. He will be redirect to a page that informs him about a verification link that was sent to his email inbox. Then, he will be redirected to a page to confirm his email.

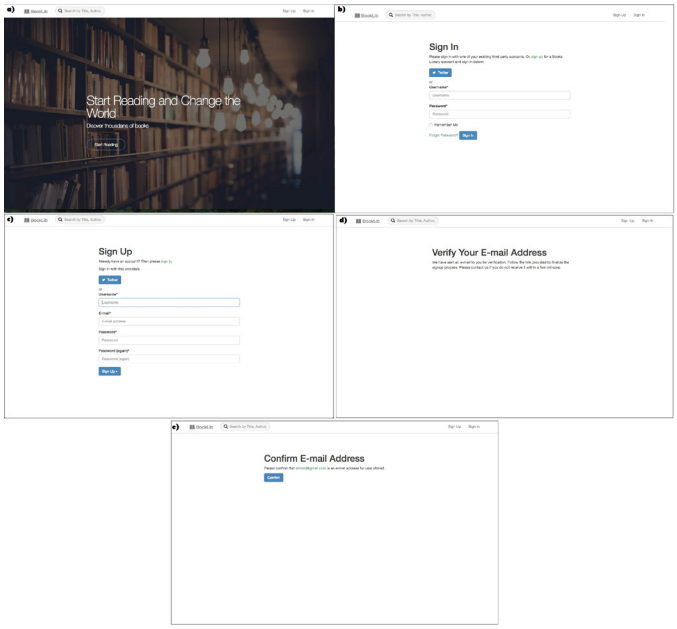


Fig. 13. An overview of sign in user process.

After the login the user is directly redirected to his time-line that contains all the recommended documents (books) according to his

profile as shown in Fig. 14. The recommendation shown here does not represent the final result, we recall that the aim of this work is to conceive and prepare the ground for the documents recommendation. The user can get more information about a document, and in this case the document is a book, by going to his details. The user can view his profile or the profile of another user. The user can view the documents that he started to read, his bookmarked documents, the users that he follows, and the users that are following him. The logged in admin can view the comments and sentiment as shown in Fig. 14 part "d".

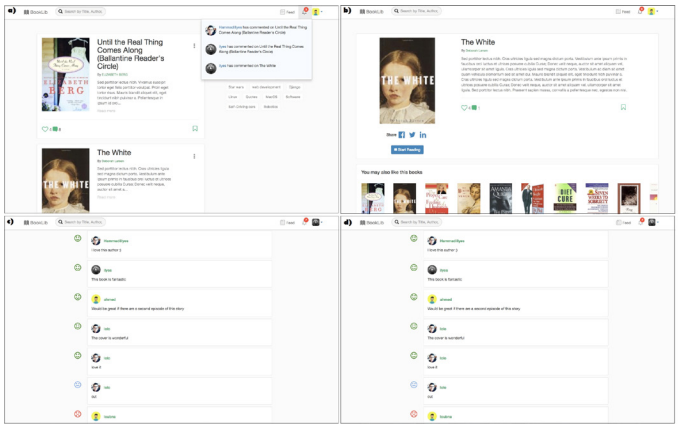


Fig. 14. User profile and documents details.

Moreover, it can access the admin panel by navigating to its url. The logged in admin can manage users and documents in the admin panel in the users and documents sections as shown in Fig. 15.

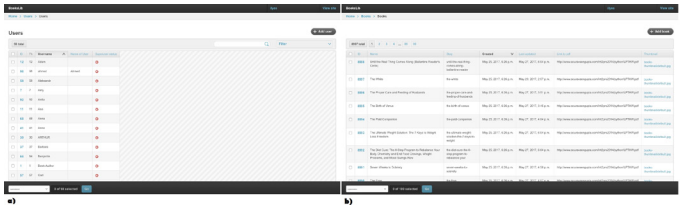


Fig. 15. Admin features.

B. Mobile Application

With the mobile app, the user can perform the same actions like in the Web application as shown in Fig. 16.

C. Scenarios

Here, we give two (02) scenarios of execution.

1) Scenario 01: Case of New User

The new user really represents a new operator who has just been recruited to the non-woven company "INOTIS", thing that remains unlikely. We have previously stated that this recommendation tool operates with a collaborative decision support system launched over the last 3 years on the INOTIS pilot company. We specify that operator data comes mainly from this system. The collection of user data lasted about one year. We mentioned as a future work the launch of this tool in the Web which opens the door to other operators from the world wide to join it. Each new operator goes through the registration step on the platform where he is invited to fill the requested information.

The user is not obliged to complete his profile; the registration form is designed in two categories "personal information of the user" which is required as: surname, first name, email address and password. The second category is "personal preferences" which is optional such as: favorite authors and centers of interest.

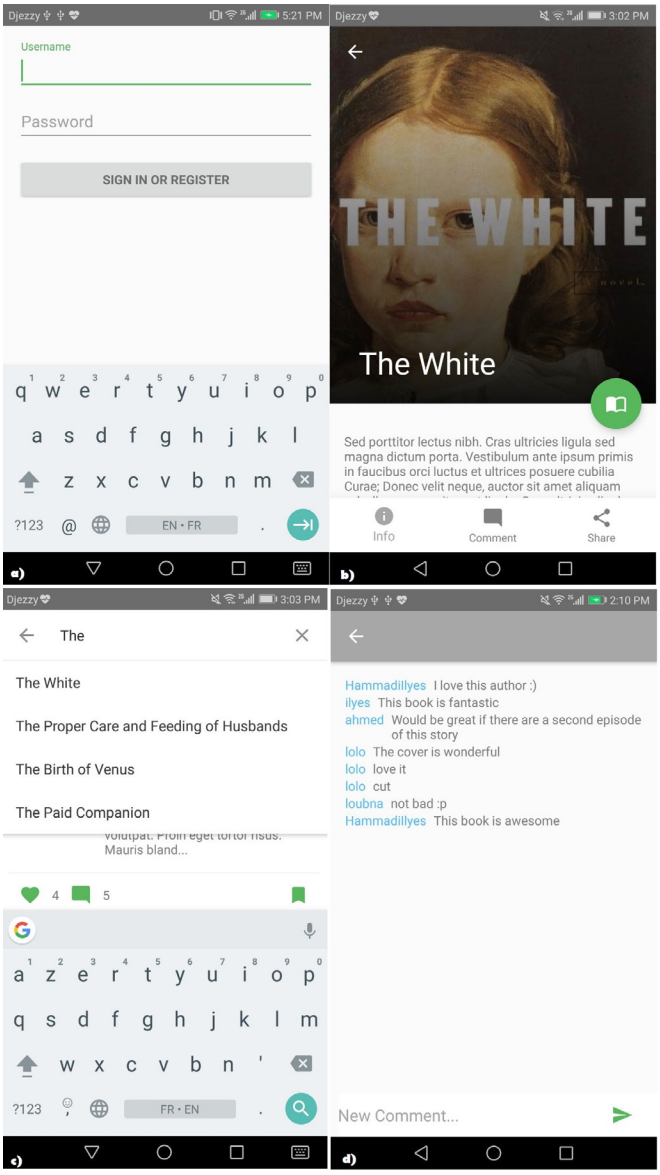


Fig. 16. An overview of the Web application.

Upon registration, he is redirected to his private area where he finds documents to note.

- If the new user doesn't complete his data profile (but he introduces his center of interest). The system uses this latter for example: the operator introduces HP machine, water pump. The system recovers the documents which contain these keywords in their description (knowing that a document has Keywords as a descriptor) in this example document whose id=38 is recovered, and some new documents.
- If this user doesn't insert his center of interest, the system proposes popular documents (i.e. popular documents are the top rated documents by other users of the platform) and new documents recently added to the documents base (the number of documents which are proposed to the evaluation is 10).

At any time, this operator can complete his profile by his personal preferences. As soon as it is done, he receives a recommendation according to the preferences expressed (as detailed in scenario 2).

2) Scenario 2: Case of an Old User (in 8 Steps)

In the case of an old user (registered user), the preferences are already

listed in the database. He can complete or update them by adding other areas of interest. The user has agreed to access his Twitter account.

The personal data of the user can be completed from his Twitter account such as: age ... for confidentiality, we cannot give a real example. It is enough to call the user "operator x".

The confidentiality of the data is ensured in the following 3 cases:

- In the case of an old operator who has already participated in the collaborative sessions by giving expertise (diagnosis). Data will not be used outside the recommendation platform.
- The notes attributed to the documents are visible only internally, that is to say only in the recommendation platform.
- When an operator comments on a document, the comment is only readable by the operators of the platform.

However, the operator is free to post a tip or an opinion on his social network for example: when he participates at a conference, he posts a tweet with hashtags. These will be exploited to enrich his list of centers of interest as explained below:

"#adhesive innovation enabling next generation hygiene products", adhesive is added as new center of interest for the operator who posts this tweet.

A set of 4 documents recently added to the database are proposed to the evaluation. The ratings assigned to D1, D2, D3 and D4 are 2, 4, 0, and 9 as a direct notation. The user has posted a comment to D4. The comment is evaluated as positive with the rating of 8/10, and then will be added to the user's profile. Users commenting on D4 will be notified.

The D3 has not been evaluated, a questionnaire is therefore proposed to the user. The answers to the questions are presented in Table IX.

TABLE IX. Users' Reports Questionnaire

Question	Answer
A part or the entire document is damaged?	No
The content of the document does not fall within your area of expertise?	No
Is the document too long?	Yes

After the assessment of the documents the score is calculated as shown Table X.

TABLE X. Documents Scores

Document	Score
D1	2
D2	4
D3	0
D4	(9+8)/2 = 8,5

In order to recommend a new document to this user, it is necessary to know its similar users having shared the same preferences in the past (document notation). Our tool is based on Cosine's measurement. We take an example of 5 documents noted by 4 users where the D5 is a new document not noted by operator x. We calculate the prediction for the operator x.

The similarity of Cosine is calculated in the following way, we detail the calculations for the user 1 and operator X.

Cos (U1, operator x) = 0.45 (according to the formula 4).

In the same way the other calculations are made. We get Table XI.

TABLE XI. Users / Documents Evaluation

	U1	U2	U3	U4
Operator x	0,45	**0,65**	**0,61**	**0,73**

Where k = 3 and the users closest to our user are U4, U2, and U3. The prediction is calculated according to the average of the notes of the similar neighbors: Operator x prediction (6 + 9 + 5)/3 = 6.66. The D5 will be appreciated by the operator X so it will be recommended (score is greater or equal to the threshold (5/10)).

As new documents coming from collaboration sessions of CDSS are offered to the different users (operators) of the platform according to their areas of interest this proposition is based on the descriptors "keywords" of the documents. The documents will be proposed according to the number of interests (keywords) that contain.

VII. Conclusion and Future Work

In this paper, we have presented a Web 2.0 platform that includes a document recommendation tool for industrial diagnosis. This tool is integrated in the collaborative decision support system in order to provide relevant solutions to industrial operators. The proposed approach is based on three essential steps for the recommendation namely: (i) data collection and filtering, (ii) prediction and (iii) evaluation. The input data for the implemented recommendation tool is varied like (i) considered data from collaborative system experiences such as operator preferences and relevant documents to recommend; (ii) information, preferences retrieved in a direct way through the forms, (iii) indirectly via the social network Twitter, or sentiments detection of comments posted. The prediction calculation is performed between the k-users using Cosine measure. It allowed us to consider similar users to the Active User and therefore obtain considerable reduction of the user / document matrix.

We discussed the results which are obtained from the recommendation tool and collaborative support system (without recommendation tool). This discussion presents the quality of documents that are recommended to industrial operators is not based only on the similarity measure of case-based reasoning [26] but on the number of relevant documents that are considered as inputs also. This comparative study allowed us to measure the number of relevant documents in the document base of collaborative support system.

We presented a comparison, which is conducted between our study and others from literature (FCS: Standard Collaborative Filtering, D-BNCF-KNN: Densified Behavioural Network based Collaborative Filtering where only the direct neighbors). The MAE results obtained from this comparison show that our approach has a low error rate compared to the other approaches.

The experiments are based on diagnostic and Book Crossing documents.

We can improve our work by:

- Using collaboration platform for identifying potential users networks,
- Evaluating the proposed tool with other data sets,

$$\cos(u1,u2) = \frac{\sum_{i=1}^{d}(note(u1,i) * note(u2,i))}{\sqrt{\sum_{i=1}^{d}(note(u1,i))^2} * \sqrt{\sum_{i=1}^{d}(note(u2,i))^2}} = \frac{((3*2)+(7*4)+(8*0)+(4*8*5)}{\left(\sqrt{(3^2+7^2+8^2+4^2+10^2)} * \left(\sqrt{(2^2+4^2+0^2+8*5^2)}\right)\right.} = 0{,}45 \quad (4)$$

- Testing other prediction measures to increase relevance,
- Widening of the recommendation tool to a variety of resource diagnostics (machines) in the field of non-woven, textile and other.
- Testing the satisfaction of the industrial operators;
- Testing the usability of the Web application according to the works presented in [29], [30] as a quantitative means for measuring the user's experience, and by using heuristics as a qualitative means such mentioned in [31].

References

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] F.Z. Benkaddour, N. Taghezout, B. Ascar, "Towards a Novel Approach for Enterprise Knowledge Capitalization Utilizing an Ontology and Collaborative Decision-Making: Application to Inotis Enterprise". International Journal of Decision Support System Technology (IJDSST), 2016, vol. 8, no 1, p. 1-24.

[3] X. Su, and M. Khoshgoftaar. "A survey of collaborative filtering techniques". Advances in Artificial Intelligence, 2009, vol. 2009, p. 4.

[4] H. Lu, C. Chen, M. Kong, H. Zhang, and Z. Zhao, "Social recommendation via multi-view user preference learning". Neurocomputing, 2016, vol. 216, p. 61-71.

[5] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items". Expert Systems with Applications, 2017, vol. 69, p. 29-39.

[6] A. Martin, P. Zarate, and G. Camillieri, "A Multi-Criteria Recommender System Based on Users' Profile Management". In: Multiple Criteria Decision Making. Springer International Publishing, 2017. p. 83-98.

[7] O., Sanjuan Martínez, C., Pelayo G-Bustelo, R., González Crespo, and E., Torres Franco, "Using recommendation system for e-learning environments at degree level". International Journal of Artificial Intelligence and Interactive Multimedia, 2009, vol. 1, no 2. p. 67–70.

[8] V. Satish Kumar, and P. Nymphia, "Survey on content based recommendation system". Int. J. Comput. Sci. Inf. Technol, 2016, vol. 7, no 1, p. 281-284.

[9] E. R, Nuñez-Valdez, J.M.C., Lovelle, O., Sanjuan-Martinez, C.E., Montenegro-Marin, and G. Infante-Hernandez. "Social voting techniques: a comparison of the methods used for explicit feedback in recommendation systems". International Journal of Interactive Multimedia and Artificial Intelligence, 2011, vol. 1, no 4.

[10] A.M., Dakhel, H.T., Malazi, and M., Mahdavi, "A social recommender system using item asymmetric correlation". Applied Intelligence, 2017, p. 1-14.

[11] J. Mathieu, Dyson, Mark, Callaway, and Duncan, "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations". In: ACEEE Summer Study on Energy Efficiency in Buildings. 2012. p. 189-203.

[12] E. R, Núñez-Valdéz, J.M.C., Lovelle, O.S., Martínez, V., García-Díaz, P.O., De Pablos, and C.E.M., Marín, "Implicit feedback techniques on recommender systems applied to electronic books". Computers in Human Behavior, 2012, vol. 28, no 4. p. 1186-1193.

[13] E. R, Nuñez-Valdez, J.M.C., Lovelle, J.M., Cueva, G.I., Hernández, A.J., A.J., Fuente, and J.E., Labra-Gayo, "Creating recommendations on electronic books: A collaborative learning implicit approach". Computers in Human Behavior, 2015, vol. 51, p. 1320-1330.

[14] R., Hastings, and M., Randolph, from https://www.netflix.com/, created in 1997.

[15] S., Chen, C., Hurley, and J. Karim, from https://www.youtube.com/, created in 2005.

[16] M., Zuckerberg, and S. Sandberg, from https://www.facebook.com/, created in 2004.

[17] Mugly, and C. Petersen, from http://www.tripadvisor.fr, created in 2000.

[18] Book Crossing, from: http://www.BookCrossing.com, visited in 01/02/2017 (2001).

[19] Google, "Google scholar", from: https://scholar.google.com, created in 2004.

[20] F., Abel, Q., Gao, G.J., Houben, and K., Tao. Analyzing user modeling on twitter for personalized news recommendations. In International Conference on User Modeling, Adaptation, and Personalization (pp. 1-12). 2011, July. Springer, Berlin, Heidelberg.

[21] Bezos, J., "Amazon.com", from: https://www.amazon.com, created in 1994. Visited in 10/01/2017.

[22] Z. Wen, Recommendation Systems based on Collaborative filtering. CS229 Lecture Notes, 2008.

[23] W. Liang, G. Lu, X. Ji. J. Li, and D. Yuan, "Difference Factor' KNN Collaborative Filtering Recommendation Algorithm". In: International Conference on Advanced Data Mining and Applications. Springer, Cham, 2014. p. 175-184.

[24] U. Sharda and P. Maes, "Social information filtering: algorithms for automating "word of mouth"". In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., 1995. p. 210-217.

[25] S. Holmes, "RMS Error", from http://statWeb.stanford.edu/~susan/courses/s60/split/node60.html. Visited in 02/04/2017 (2000).

[26] F.Z. Benkaddour, N. Taghezout, B. Ascar, "Novel Agent Based-approach for Industrial Diagnosis: A Combined Use between Case-based Reasoning and Similarity Measures". International Journal of Interactive Multimedia & Artificial Intelligence, 2016, vol. 4, no 2.

[27] I. Esslimani, A. Brun, and A. Boyer, from social networks to behavioral networks in recommender systems. In: Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in. IEEE, 2009. p. 143-148.

[28] I. Esslimani, Vers une approche comportementale de recommandation: apport de l'analyse des usages dans un processus de personnalisation. 2010. Thèse de doctorat. Université Nancy II.

[29] Schrepp, M., A. Hinderks, and J. Thomaschewski. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S), International Journal of Interactive Multimedia and Artificial Intelligence, 2017, vol. 4 no. 6, pp. 103-108.

[30] Schrepp, M., A. Hinderks, and J. Thomaschewski. Construction of a Benchmark for the User Experience Questionnaire (UEQ), International Journal of Interactive Multimedia and Artificial Intelligence, 2017, vol. 4 no. 4, pp. 40-44.

[31] Bader, F., E. M. Schön, and J. Thomaschewski. Heuristics Considering UX and Quality Criteria for Heuristics, International Journal of Interactive Multimedia and Artificial Intelligence, 2017, vol. 4 no. 6, pp. 48-53.

[32] Q., Yuan, C., Li, S., Zhao, "Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation". In: Proceedings of the 5th ACM conference on recommender systems. ACM, 2011, pp 245–252.

[33] J. Bobadilla, F., Ortega, A., Hernando, A., Gutierrez, "Recommender systems survey." Knowl-Based Syst vol. 46, 2013, pp.109–132.

[34] Nagard, E., L., "Use the Twitter API to collect tweets with Talend" from: http://www.erwanlenagard.com/general/tutoriel-utiliser-lapi-twitter-pour-collecter-des-tweets-sans-coder-avec-talend-1029. Accessed: 2017-05-20

[35] Sentiment analysis. http://text-processing.com/docs/sentiment.html. Accessed: 2017-05-29

F. Z. Benkaddour

F. Z. Benkaddour is an assistant professor at École Normale Supérieure d'Oran (ENSO). She holds her doctorate thesis in IRAD at university of ORAN1 Ahmed Ben Bella, Algeria in 2017. She received her Master degree on ID-IHM from the same university, in 2013. She is a member of the EWG-DSS (Euro Working Group on Decision Support Systems) since May 2015. She is also a member of the LIO (Laboratoire d'Informatique d'Oran) laboratory. Her research topics include Collaborative Decision Support System, Knowledge Management, Case Based Reasoning Systems, Multi Agents Systems, WEB technologies, and Recommender Systems.

N. Taghezout

N. Taghezout is an assistant professor at ORAN1 University. She holds her doctorate thesis in MITT at PAUL SABATIER UNIVERSITY in France in 2011. She also received another doctorate thesis in Distributed Artificial Intelligence from ORAN University in 2008. She holds a Master degree in Simulation and Computer aided-design. Dr Noria Taghezout conducts her research at the LIO laboratory as a chief of the research group in Modeling of enterprise process by using agents and WEB technologies. Since she studied in UPS Toulouse, she became a member of the EWG-DSS (Euro Working Group on Decision Support Systems). She is currently lecturing Collaborative decision making, Enterprise management and Interface human machine design. Her seminars, publications and regular involvement in Conferences, journals and industry projects highlight her main research interests in Artificial Intelligence.

F. Z. Kaddour-Ahmed

F. Z. Kaddour-Ahmed holds her Master degree on KEWT (Knowledge Engineering and Web Technology) at university of ORAN1 Ahmed Ben Bella, Algeria in 2017. She obtained her Bachelor's degree on CND (Computer Networks and Database) at university of Aïn Témouchent, Algeria in 2015. Her research topics include information systems, WEB technologies, and Recommender Systems.

I. A. Hammadit

I. A. Hammadi is a Full Stack Web Developer at Joboti since September 2017. He holds his Bachelor's degree in Computer Science in 2017 at university of ORAN1 Ahmed Ben Bella. His domains interest are Software Engineering and Development, Cloud Native Technologies, Machine Learning and AI.

An Evolutionary Approach for Learning Opponent's Deadline and Reserve Points in Multi-Issue Negotiation

R. Ayachi*, H. Bouhani, N. Ben Amor

Larodec laboratory, Tunis (Tunisia)

Received 15 February 2018 | Accepted 23 July 2018 | Published 6 August 2018

Abstract

The efficiency of automated multi-issue negotiation depends on the available information about the opponent. In a competitive negotiation environment, agents do not reveal their parameters to their opponents in order to avoid exploitation. Several researchers have argued that an agent's optimal strategy can be determined using the opponent's deadline and reserve points. In this paper, we propose a new learning agent, so-called Evolutionary Learning Agent (ELA), able to estimate its opponent's deadline and reserve points in bilateral multi-issue negotiation based on opponent's counter-offers (without any additional extra information). ELA reduces the learning problem to a system of non-linear equations and uses an evolutionary algorithm based on the elitism aspect to solve it. Experimental study shows that our learning agent outperforms others agents by improving its outcome in term of average and joint utility.

Keywords

Automated Negotiation, Deadline Learning, Differential Evolution, Invasive Weed Optimization, Agent-based Systems.

DOI: 10.9781/ijimai.2018.08.001

I. Introduction

AUTOMATED negotiation aims to imitate the humans' negotiation process using intelligent agents. It is based on three components [1]: (1) the negotiation protocol defining the rules governing the interaction between the negotiating agents such as the number of agents, their actions, etc. When we deal with two agents, we talk about bilateral negotiation. If the negotiation concerns more than two partners, negotiation is then multilateral. (2) The negotiation object corresponding to the set of issues under negotiation. Issues are the characteristics of the negotiation item that are taken into account during the evaluation [7]. A negotiation can be either a single-issue or a multi-issue. (3) The negotiation strategy determining the agents plan for reaching a satisfactory agreement. It includes tactics and decision functions adopted by the negotiating agent.

In this work, we focus on bilateral multi-issue negotiation under a time constraint. In this context, several challenges can be derived out from the fact that negotiators do not reveal their private information (e.g., preference, deadline and reserve point) to their opponents for fear of being exploited.

Several researchers paid attention to the endowment of learning techniques into negotiating agents [1-3]. Most of the proposed learning methods require prior knowledge about the opponent. The challenge is to propose a learning method that only uses available information during the negotiation.

In this paper, we propose a learning agent, so-called Evolutionary Learning Agent (ELA), employing the evolutionary learning approach Differential Evolution Invasive Weed Optimization (DEIWO) [4] to learn its opponent's deadline and reserve points from only his counter-offers in a bilateral multi-issue negotiation. The use of DEIWO allows ELA enhancing its performance even with an important number of issues.

The remainder of this paper is as follows: Section 2 presents basic concepts of bilateral multi-issue negotiation and related work. The new learning approach is detailed in Section 3. The empirical evaluation and the analysis are presented in Section 4.

* Corresponding author.
E-mail address: raouia.ayachi@gmail.com

II. Bilateral Multi-issue Negotiation

As stated above, we are interested in a bilateral multi-issue negotiation framework as it is widely used in the agent's field [5]. In such environment, agents negotiate by exchanging offers and counteroffers, until they either reach a consensus that satisfies each party's private preferences, or withdraw from negotiation without any agreement. The communication between agents is made using the simple alternative offer protocol [6]. Formally, let I be a pair of negotiating agents (generally $I = \{B,S\}$ where B corresponds to a Buyer and S to a Seller). According to the negotiation process, the two negotiating agents act in conflictive roles. An agent receiving an offer at time t needs to decide whether (1) to accept or (2) to reject it and propose a counteroffer at time $t + 1$.

Counter-offers can be made until one agent's deadline (denoted τ^i for agent i) is reached as shown in Fig. 1. Since the negotiating agents have time constraint, they use time-dependent tactics (when preparing their offers) which model the fact that they concede faster as the deadline approaches [6].

Let $J = \{1,..,n\}$ be the set of issues under negotiation. For each issue $j \in J$, any agent $i \in I$ has a lower and upper reserve point, denoted by IP_j^i and RP_j^i, respectively and corresponding to the minimal and maximal utility the agent i is willing to accept during the negotiation.

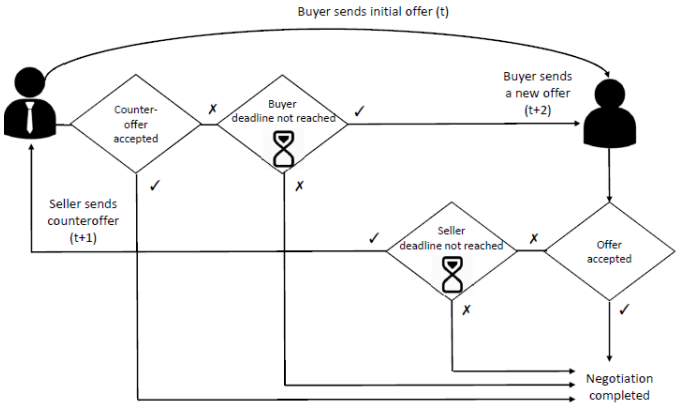


Fig. 1. Negotiation process.

A. Basics of Bilateral Multi-issue Negotiation

An agent *i* starts the negotiation (at $t = 0$) with its initial price (IP_j^i) and at $t = \tau^i$, concedes to its reserve point (RP_j^i). Formally, in each round *t*, agent *i* assigns to each issue $j \in J$ a value $x_j^i[t] \in [IP_j^i, RP_j^i]$ expressed by:

$$x_j^i[t] = IP_j^i + \left(RP_j^i - IP_j^i\right) * \left(\frac{t}{\tau^i}\right)^{\alpha_j^i} \quad (1)$$

where α_j^i is *i*'s concession rate that quantifies the amount an agent concedes towards its opponent during the negotiation [7]. For simplicity reasons, we will use the notation α_j to denote the concession rate for either agent i or its opponent i'. Fig. 2 depicts the behavior of α_j on agent's offers curve. Clearly, the agent's tactics are classified into three classes [8], depending on the value of α_j , namely:

- *Boulware* ($\alpha_j < 1$)*:* where an agent maintains the offered value until the time is almost exhausted.
- *Conceder* ($\alpha_j > 1$)*:* where an agent concedes to its reserve point very quickly.
- *Linear* ($\alpha_j = 1$)*:* where an agent makes a constant rate of concession.

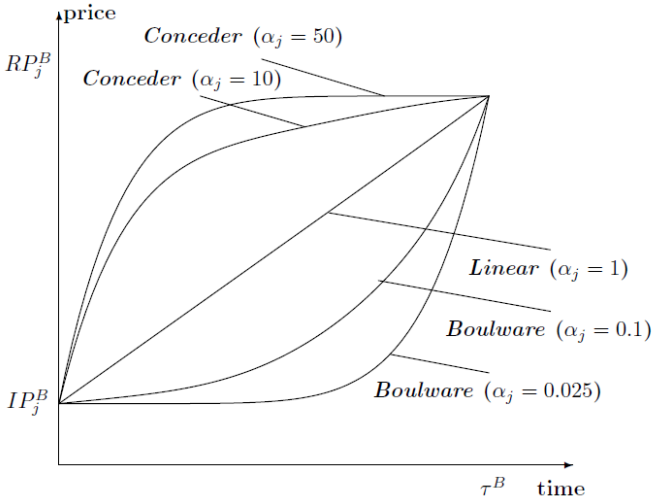


Fig. 2. The concession rate's impact on agent's offering curve.

The opponent's concession rate $\alpha_j^{i\prime}$ can be computed using $IP_j^{i\prime}$ and two successive offers as follows:

$$\alpha_j^{i\prime} = log_{\frac{t}{t-1}}\left(\frac{x_j^{i}[t] - IP_j^{i}}{x_j^{i}[t-1] - IP_j^{i}}\right) \text{ where } t \geq 3 \quad (2)$$

A non-learning agent keeps the value of α_j unchanged until the negotiation ends while a learning one adjusts α_j in order to maximize its utility. Each agent employs an evaluation function $V_j^i: \left[IP_j^i, RP_j^i\right] \rightarrow [0,1]$ which assigns a normalized valuation to each possible value x_j. Formally:

$$V_j^i(x_j) = \frac{RP_j^i - x_j}{RP_j^i - IP_j^i} \quad (3)$$

Thus, the utility function modelling agent i's preferences is a linearly additive function defined by:

$$U^i(x) = \sum_{j=1}^{n} w_j^i \; V_j^i(x_j) \quad (4)$$

where w_j^i is the *j*'s weight of the issue given by agent *i*. The utility function $U^i(x)$ is set to 1 at the beginning of the negotiation and decreases as the deadline approaches.

Since the negotiation is a sequence of alternative offers finishing with an *accept* or a *withdraw*. The response to an offer $x^{i'}[t]$ at time round *t*, denoted by Response (t, $x^{i'}$), is :

$$\text{Response}(t, x^{i\prime}) = \begin{cases} withdraw \; if \;\; t > \tau^i \\ accept \; if \;\; U^i(x^{i\prime}[t]) \geq U^i(x^i[t]) \\ x^i[t] \qquad otherwise \end{cases} \quad (5)$$

B. Related Work on Learning Opponent's Deadline and Reserve Point

In recent years, endowing agents with machine learning techniques has attracted automated negotiation community. Several researchers paid their attention to learn opponent's deadline and reserve point [1, 2, 9–12]. In what follows, we briefly review works related to the learning of opponent's deadline and reserve point. In fact, the learning problem has been deeply studied for bilateral single-issue negotiation. The first investigation was established by Hou [9] to learn the opponent's deadline and reserve point by employing non-linear regression. No mechanism has been used for adapting the concession strategy and this represents the major weakness of the method. Sim et al. [10] proposed a Bayesian Learning (BL) approach called BLGAN that only learns the opponent's reserve point and then employs a genetic algorithm to generate counter-offers. Sim et al. [11] proposed an improved version of BLGAN in which they incorporated a deadline learning method. Compared to BLGAN, Gwak et al. [2] exploited a new conditional probability to update the belief of the opponent's reserve point. In this framework, the concession rate adjustment mechanism is not efficient when opponent's deadline is greater than the learning agent's deadline. Yu et al. [12] proposed a combination of BL and regression analysis in order to estimate the opponent's deadline and reserve point. They defined a set of hypotheses about the values of the opponent's deadline and reserve point and then used the non-linear correlation coefficient to update agent's beliefs.

The research mentioned above only deals with bilateral single-issue negotiation. In contrast, only two works handle multi-issue negotiation. In fact, Zeng el al. [13] used a BL method called *Bazaar* which is a sequential decision making approach modelling beliefs of the opponent's reserve point. Their learning method does not include any mechanism to learn the opponent's deadline, in addition, it requires extra information about the opponent which is not available in most cases. Zhang et al. [14] improved the learning approach proposed in [12] to deal with multi-issue negotiation. To this end, they use strong assumptions about the opponent's preferences since they only deal with conflicting and equally weighted issues. By conflict issues they mean that increasing the value of an issue will help agents to raise their utilities but to decrease the opponents' utilities.

From another side, Coehoorn et al. [15] have used the Kernel Density Estimation (KDE), a non-parametric method for learning opponent's preferences where the initial distribution is based solely on the training data. However, this method was applied in a context of making negotiation trade-offs in bilateral encounters in which an agent concedes on one issue and demands more on another, which falls outside the scope of this paper.

From this review on bilateral single and multi-issue negotiation, it is clear that existing works strongly depend on a set of hypothesis. These assumptions must be made about the underlying distribution function. In what follows, we propose a novel approach to learn opponent's deadline and reserve point for multi-issue negotiation able to learn simultaneously multiple reserve points for multiple issues without any extra information.

III. A Novel Approach for Learning Opponent's Deadline and Reserve Points in Multi-Issue Negotiation

The basic idea of the new negotiation approach is to model the learning process as an optimization problem. To this end, the learning problem should be expressed in mathematical terms, i.e., reduced to a system of non-linear equations, in order to take advantage of recent results on optimization. In fact, the Differential Evolution Invasive Weed Optimization (DEIWO) [4] algorithm will be used to solve the system. Afterwards, our learning agent so called *Evolutionary Learning Agent* (ELA) should be able to adjust its concession strategy according to the estimated value of opponent's deadline and reserve points.

A. Transformation of the Learning Problem

As mentioned in Section 2, deadline and reserve point are inter-dependent terms, i.e., if the deadline is learned, the reserve point can be easily estimated. This is explained by the fact that agents concede to their reserve points when the deadline is reached. Based on (1), we can easily express the relationship between deadline and reserve point as expressed by (6) and (7).

$$RP_j^{i'} = \frac{x_j^{i'}[t] - IP_j^{i'}}{\left(\frac{t}{\tau^{i'}}\right)^{\alpha_j^{i'}}} + IP_j^{i'} \tag{6}$$

$$\tau^{i'} = t.\left(\frac{RP_j^{i'} - IP_j^{i'}}{x_j^{i'}[t] - IP_j^{i'}}\right)^{\frac{1}{\alpha_j^{i'}}} \tag{7}$$

The proposed approach aims to learn the opponent's deadline and reserve points by reducing the learning problem to a system of non-linear equations problem. To this end, we will use (7) in order to build our system. Moreover, the learning agent ELA needs to find the parameters that best fit its opponent's historical offers. Formally, for each issue under negotiation, the following error function should be minimized:

$$f_j\left(\widehat{RP}_j^{i'}[t], \hat{\tau}^{i'}[t]\right) = \sum_{k=1}^{t}\left(t/\left(\frac{x_j^{i'}[k] - IP_j^{i'}}{\widehat{RP}_j^{i'}[t] - IP_j^{i'}}\right)^{\frac{1}{\alpha_j^{i'}}} - \hat{\tau}^{i'}\right)^2 \tag{8}$$

where $\hat{\tau}^{i'}$ and $\widehat{RP}_j^{i'}[t]$ are the learned deadline and j's reserve point at time round t, respectively.

To take advantage of the issues' multitude and (8), we reduce the learning problem to a system of nonlinear equations as follows:

$$\begin{cases} f_1\left(\widehat{RP}_1^{i'}[t], \hat{\tau}^{i'}[t]\right) = 0 \\ \dots \\ f_J\left(\widehat{RP}_J^{i'}[t], \hat{\tau}^{i'}[t]\right) = 0 \end{cases} \tag{9}$$

The deadline's search area is restricted as it remains the same for all cases. While when we deal with multiple reserve points (i.e., one for each issue) we have a good approximation of the opponent's deadline (i.e., close to the exact value) but inaccurate estimations of the opponent's reserve points. To make the results rigorous, the opponent's reserve points should be recomputed using the relation in (6).

Solving a system of non-linear equations corresponds to a multi-objective optimization problem in which all the functions should be minimized. Formally,

$$min \sum_{j=1}^{J} f_j\left(\widehat{RP}_j^{i'}[t], \hat{\tau}^{i'}[t]\right) \tag{10}$$

B. DEIWO Optimization Algorithm

There exists an extensive literature centered on solving optimization problems [16, 17]. Within the panoply of existing methods, we focus on a recent evolutionary algorithm *Differential Evolution Invasive Weed Optimization* (DEIWO) [4] that proved its efficiency for solving systems of non-linear equations. DEIWO has the abilities to overcome local optimal solutions and obtain global optimal solutions and is based on two popular global optimization algorithms, namely: the Invasive Weed Optimization (IWO) and the Differential Evolution (DE).

The first part of DEIWO employs the IWO algorithm which is a numerical optimization method inspired from colonizing weeds [18]. In IWO, weeds refers to feasible solutions of the given problem. They are spread over the search area and are allowed to produce seeds (i.e., new solutions) depending on their fitness. After some iterations, the number of population reaches its maximum, consequently, a mechanism for eliminating plants with poor fitness activates. Compared to the Genetic Algorithm (GA) [19], IWO employs a different way to disperse new individuals. In fact, the generated seeds are randomly dispersed over the search space by normally distributed random numbers with mean equal to zero; but varying variance [18]. The second part of DEIWO exploits DE which provides means for accelerating the optimization [20]. It is based on three operators: *mutation, crossover* and *selection*. These operators keep population diversity and avoid premature convergence.

To improve the quality of the DEIWO output at time round t, we incorporate its output at time round $t-1$. Thus, if the previous output stills optimal, it is not necessary to readjust the opponent's learned parameters. Formally, the improved DEIWO procedure is outlined in Algorithm 1.

Algorithm 1 the Differential Evolution Invasive Weed Optimization Algorithm

1: **input**: Solution at t-1 + the opponent's historical offers + DEIWO parameters;
2: **output**: Best solution at t;
3: Set the generation counter g=0;
4: Initialize the population size;
5: Initialize the population P (g); /*Initialization step*/
6: **while** g < maximum iteration **do** /*IWO phase*/
7: Compute the fitness for each weed in P (g);
8: Produce new seeds based on fitness; /*Reproduction step*/
9: Add new seeds to the population P (g); /*Spatial dispersal step*/
10: **if** population size > maximum population **then**
11: Eliminate weeds with poor fitness; /*Competitive exclusion step*/
12: **end if** /*DE phase*/
13: Perform mutation on P (g);
14: Perform crossover on P (g);
15: Perform selection on P (g);
16: Increment g by 1;
17: **end while**
18: **return** best solution;

In our problem, each weed represents a solution for the system in Eq. (9) and is expressed as follows:

$$w_n = \left(\hat{\tau}^{i'}, \widehat{RP}_1^{i'}, \widehat{RP}_2^{i'}, \ldots., \widehat{RP}_J^{i'} \right) \quad (11)$$

where $\hat{\tau}^{i'}, \widehat{RP}_1^{i'}, \widehat{RP}_2^{i'}, \ldots., \widehat{RP}_J^{i'}$ are feasible solutions for (9). Basically, weeds with the lowest fitness are closest to the optimal solution. To evaluate the quality of a solution, we consider the following fitness function:

$$f(w_n) = \sum_{j=1}^{J} f_j \left(\widehat{RP}_j^{i'}[t], \hat{\tau}^{i'}[t] \right) \quad (12)$$

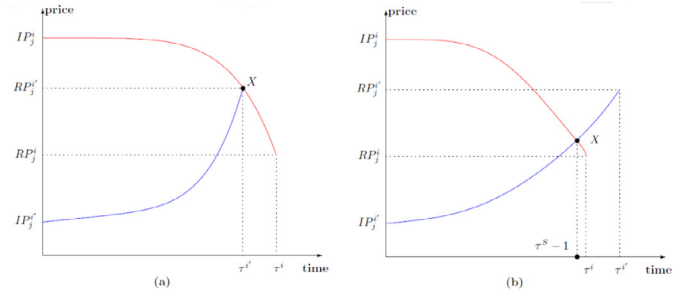


Fig. 3. Two cases of concession strategy: (a) $\hat{\tau}^{i'} < \tau^{i}$ and (b) $\hat{\tau}^{i'} > \tau^{i}$.

C. Concession Rate Adjustment

Learning the opponent's deadline and reserve points is necessary for finding ELA's optimal strategy. In each round, the learning agent uses the predicted values of deadline and reserve points in order to adjust its bidding strategy. Henceforth, it could improve its outcome and avoid disagreements at the end of the negotiation.

Before laying out with the concession rate, let us adapt the offering function. ELA generates counter offers using the following equation proposed in [11]:

$$x_j^i[t] = x_j^i[t-1] + \left(RP_j^i - x_j^i[t-1]\right) * \left(\frac{1}{\tau^i - t + 1}\right)^{\alpha_j} \quad (13)$$

Compared to (1), ELA treats its previous offer $x_j^i[t-1]$ as its new initial point at time round *t*.

The opponent's concession rate $\alpha_j^{i'}$ can be computed using $IP_j^{i'}$ and two successive offers as follows:

$$\alpha_j^{i'} = log_{\frac{t}{t-1}}\left(\frac{x_j^{i'}[t] - IP_j^{i'}}{x_j^{i'}[t-1] - IP_j^{i'}}\right), \text{ where } t \geq 3 \quad (14)$$

A non-learning agent keeps the value of α_j unchanged until the negotiation ends while a learning one adjusts α_j in order to maximize its utility.

Finding the optimal strategy for the learning agent can be analyzed by considering two cases:

- **Case 1** ($\hat{\tau}^{i'} < \tau^{i}$): In this case, the opponent's deadline is smaller than agent *i*'s deadline. Fig.3 (a) depicts the behavior of the learning agent in this case. When the two curves intersect, the negotiating agents reach an agreement. That is why, ELA needs to negotiate with its opponent's as long as possible in order to catch the opponent's best offers. Therefore, the optimal strategy for agent *i* is to make its offers curve cross the opponent's offer curve at t=$\hat{\tau}^{i'}$ (point X). Hence, the optimal strategy can be computed as follows [11]:

$$\alpha_j^i[t] = \left| log_{\frac{\hat{\tau}^{i'} - t + 1}{\tau^i - t + 1}} \left(\max\left\{0, \left| \frac{x_j^{i'}[t-1] - \widehat{RP}_j^{i'}}{x_j^{i'}[t-1] - RP_j^i} \right| \right\} \right) \right| \quad (15)$$

where $\hat{\tau}^{i'}$ and $\widehat{RP}_j^{i'}$ are, respectively, the learned opponent's deadline and issue *j*'s reserve point.

- **Case 2** ($\hat{\tau}^{i'} > \tau^{i}$): In this case, the opponent's deadline is greater than agent i's deadline. Fig. 3(b) depicts the behavior of the learning agent in this case.

Proposition 1 *let X be the optimal point that maximizes the learning agent's utility. The optimal strategy for agent i in this case is to make its offers curve cross the opponent's offer curve at* $t = \tau^i - 1$ *since at* $t = \tau_{i'}$ *agent i concedes to its reserve point which is the worst case. The proposed adjusting formula is as follows:*

$$\alpha_j^i[t] = log_{\frac{\tau^i - t}{\tau^i - t + 1}} \left(\frac{IP_j^{i'} - x_j^i[t-1]}{RP_j^i - x_j^i[t-1]} + \frac{RP_j^{i'} - IP_j^{i'}}{RP_j^i - x_j^i[t-1]} \left(\frac{\tau^i - 1}{\tau^{i'}}\right)^{\alpha_j^{i'}} \right) \quad (16)$$

Proof For each issue under negotiation, crossing the opponent's offers curve at $t = \tau^i - 1$ is expressed as follows:

$$x_i^{\tau^i - 1} = x_{i'}^{\tau^i - 1}$$

$$\overset{so}{\Rightarrow} IP_j^i + (RP_j^i - IP_j^i)\left(\frac{\tau^i - 1}{\tau^i}\right)^{\alpha_j^i} = IP_j^{i'} + (RP_j^{i'} - IP_j^{i'})\left(\frac{\tau^i - 1}{\tau^{i'}}\right)^{\alpha_j^{i'}}$$

$$\overset{so}{\Rightarrow} \left(\frac{\tau^i - 1}{\tau^i}\right)^{\alpha_j^i} = \frac{IP_j^{i'} - IP_j^i}{RP_j^i - IP_j^i} + \frac{RP_j^{i'} - IP_j^{i'}}{RP_j^i - IP_j^i}\left(\frac{\tau^i - 1}{\tau^{i'}}\right)^{\alpha_j^{i'}}$$

$$\overset{so}{\Rightarrow} \alpha_j^i = log_{\frac{\tau^i - 1}{\tau^i}} \left(\frac{IP_j^{i'} - IP_j^i}{RP_j^i - IP_j^i} + \frac{RP_j^{i'} - IP_j^{i'}}{RP_j^i - IP_j^i}\left(\frac{\tau^i - 1}{\tau^{i'}}\right)^{\alpha_j^{i'}} \right) \blacksquare$$

Note that when the opponent's deadline is greater than agent i's deadline (case 2), there is one special case that have to be considered.

If the estimated $\widehat{RP}_j^{i'}$ for agent *i* (*B* or *S*) at time round $\tau_i - 1$ is lower (respectively greater) than RP_j^i, the learning agent does not adjust its concession strategy because it is still possible that the learned reserve point may change in later time rounds.

By considering (15) and (16) as the offering tactic, the adjusting formula becomes as follows:

$$\alpha_j^i[t] = \begin{cases} \left| log_{\frac{\hat{\tau}^{i'} - t + 1}{\tau^i - t + 1}} \left(\max\left\{0, \left| \frac{x_j^{i'}[t-1] - \widehat{RP}_j^{i'}}{x_j^{i'}[t-1] - RP_j^i} \right| \right\} \right) \right| & \hat{\tau}^{i'} < \tau^i \\ log_{\frac{\tau^i - t}{\tau^i - t + 1}} \left(\frac{IP_j^{i'} - x_j^i[t-1]}{RP_j^i - x_j^i[t-1]} + \frac{RP_j^{i'} - IP_j^{i'}}{RP_j^i - x_j^i[t-1]} \left(\frac{\tau^i - 1}{\hat{\tau}^{i'}}\right)^{\alpha_j^{i'}} \right) & \hat{\tau}^{i'} > \tau^i \end{cases} \quad (17)$$

Equality: when the deadline of the opponent is equal to the agent deadline we will use case 2 because if the agent reaches the deadline he will give his reservation value so we need to get a deal before deadline is met at round (deadline - 1).

D. Illustrative Example

In this section, we give an illustrative example of the proposed negotiation model. Let us consider two agents, one buyer (B) and one seller (ELA), negotiating over a service provided by ELA. The issues under negotiation are the service's price and duration. It is obvious that B (respectively, ELA) wants to reduce its costs as much as possible. Hence, B wants to quickly execute the task with the lower price. On the other hand, ELA does not want to waste its computational resources but at the same time wants to maximize the selling price. Table I summarizes agents' parameters for the two issues.

TABLE I. AGENT'S PARAMETERS FOR THE ILLUSTRATIVE EXAMPLE

	Price (DT)			Duration (sec)			Deadline
	IP	RP	α	IP	RP	α	τ
Seller (ELA)	100	55	1	30	10	1	5
Buyer	40	70	1	5	20	1	7

Each agent starts the negotiation with its initial price. Using Eq. (13), ELA's offer at $t = 1$ is computed as follows: $x^{ELA}_{price}[1] = 100 + (55 - 100)\left(\frac{1}{5}\right)^1 = 91$. With the same equation, ELA computes its offer for the second issue. After two rounds, ELA will be able to compute B's concession rate. The sequence of counter offers for both agents is shown in Table II.

TABLE II. AGENT'S OFFERING FOR THE ILLUSTRATIVE EXAMPLE: ELEMENTS IN THE OFFER VECTOR REFER TO THE PRICE AND DURATION, RESPECTIVELY

Round	0	1	2
ELA	[100, 30]	[91, 26]	[82, 26.22]
Buyer	[40, 5]	[44.29, 7.14]	[48.57, 9.28]

Using (14), ELA computes B's concession rate as follows: $\alpha^B_{price} = log_{\frac{2}{1}}\left(\frac{48.57-40}{44.29-40}\right) = 1$.

After that, ELA employs the proposed DEIWO based Learning method to predict B's deadline and reserve points. In this illustrative example, we do not aim to explain how the weeds are generated. However, we clarify how the best solution is selected from a population of weeds.

Let us consider a population containing three weeds, w_1, w_2 and w_3 shown in Table III. To determine w_1's fitness, we need first to compute $f_{price}(w_1)$ and $f_{duration}(w_1)$ using Eq. (8) and then perform their sum.

TABLE III. A POPULATION CONTAINING THREE WEEDS

	$\hat{\tau}^B$	$\widehat{RP}^B_{price}$	$\widehat{RP}^B_{duration}$	Fitness
w_1	6	66	17	0.1587
w_2	7	70	20	0.0002
w_3	8	74	22	0.0161

For example,

$$f_{price}(w_1) = f_{price}(66.6) = \left[1 / \left(\left(\frac{44.29-40}{66-40}\right)^{\frac{1}{1}}\right) - 6\right]^2$$

$$+ \left[2 / \left(\left(\frac{48.57-40}{66-40}\right)^{\frac{1}{1}}\right) - 6\right]^2 = 0.0046$$

and $f_{duration}(w_1) = 0.1547$. Now, (12) is used to compute w_1's fitness as follows: $f(w_1) = 0.0046 + 0.1541 = 0.1587$. The same calculations are made for the other weeds.

The weed $w_2 = (7, 70, 20)$ is selected as the best solution because it has the lowest fitness of the population. Therefore, B's deadline (equal to 7 from w_2) is greater than ELA's deadline (equal to 5). Hence, using (17), ELA adjusts its concession rate for the issue price as follows:

$$\alpha^{ELA}_{price}[3] = log_{\frac{5-3}{5-3+1}}\left(\frac{40-82}{55-82} + \frac{70-40}{55-82}\left(\frac{5-1}{7}\right)^1\right) = 0.204$$

With the same manipulation, S adjusts its concession strategy for the second issue.

In the next rounds, ELA repeats the same procedure. Since ELA's deadline is less than B's deadline, the agreement much be reached at $t = (\tau^{ELA} - 1)$. In a such situation, ELA's offer at $t = (\tau^{ELA} - 1)$ is $X^{ELA}[4] = [57.143, 13.57]$ and B's offer is $X^B[4] = [57.143, 13.57]$. Thus, an agreement is reached with $U(X^{ELA}) = 0.1131$ instead of $U(X^{ELA}) = 0$.

IV. EXPERIMENTAL STUDY

To evaluate a heuristic-based negotiation model, simulations need to be performed. In this section, we evaluate our new learning agent ELA through multiple simulations and scenarios.

Experiments were implemented in Java 7 language, compiled using the Eclipse Java Mars environment and ran on windows 10-64 bits equipped with an Intel Core i7-4750QM (3 GHz) and 16 GB of RAM. We start with describing the experimental protocol then we compare agents.

A. Experimental Protocol

We will evaluate our negotiation model by comparing ELA to the following agents:

1. An agent with complete information which adapts its concession strategy based on available information.
2. A no learning agent that does not learn its opponent's parameters and its concession strategy remains fixed during the negotiation.
3. The Bayesian Learning Agent (BLA) based agent [14] that learns its opponent's reserve utility and deadline in order to adjust its concession strategy.

We propose to study four scenarios (Incomplete, Complete, ELA, BLA) as detailed in Table IV. For each of scenario, 1000 random runs were carried out to show the generality and the robustness of our negotiation model. In each run, *S* and *B* were programmed using the same parameters. Table V defines the negotiation settings. Since BLA and ELA employ different learning techniques, they have additional parameters shown in Table VI.

TABLE IV. FOUR NEGOTIATION SCENARIOS

Scenario	Seller	Buyer
Incomplete	Incomplete information	Incomplete information
Complete	Complete information	Incomplete information
ELA	S learns the opponent's deadline and RPs using ELA	Incomplete information
BLA	S learns the opponent's deadline and RU using BLA	Incomplete Information

TABLE V. THE NEGOTIATION PARAMETERS

	Issues' parameters
min_j	1
max_j	100
Number of issues	4
Preferences	0.25
	Agents' parameters
Parameter	Buyer / Seller
IP_j	$[RP^B_j, maxj]/ [minj, RP^S_j]$
RP_j	$[minj + 5, minj + (maxj - minj)/2]/ [RP^S_j + 10, maxj - 5]$
α_j	[0.1,5]
τ	[10,100]

TABLE VI. BLA & ELA PARAMETERS

BLA parameters										
Parameter	V_{max}		V_{min}		Number of cells					
Value	1		0.2		100					
ELA parameters										
Parameter	P_{init}	P_{max}	σ_{init}	σ_{final}	s_{min}	s_{max}	n	F	CR	MaxIt
Value	20	30	5	1	5	15	3	1	0.5	30

In order to evaluate our negotiation model, we will use four common performance measures [1]:

- **The Average Utility (AU)** considered as the most popular performance measure to gain the utility of outcomes.

$$AU = \frac{1}{N_{success}} \sum_{i=1}^{n} U(X) \quad (18)$$

where $N_{success}$ is the number of deals.

- **The Success Rate (SR)** represents the ability of the negotiation model to reach an agreement.

$$SR = \frac{N_{success}}{N} \quad (19)$$

where N is the number of negotiation.

- **The Average Negotiation Speed (ANS)** measuring the average duration of negotiation. A small ANS reduces costs but also affects the quality of the outcome of an agent.

$$ANS = \frac{1}{N_{success}} \sum_{i=1}^{N_{success}} t^i \quad (20)$$

where t^i is the time round of the agreement if an agreement is reached.

- **The Joint Utility (JU)** measuring the fairness of the outcome. In a competitive environment, JU is better when it is minimal.

$$JU = \sqrt{US * UB} \quad (21)$$

The testbed consists of two negotiation agents with conflicting interests (i.e., one buyer and one seller). Each agent is an instance from a super class Agent. Besides their private information (RP and deadline), the agents incubate three main procedures:

- A procedure for recording the opponent's historical offers.
- A procedure for learning the opponent's parameters.
- A procedure for generating offers based on the learning results.

Fig. 4 depicts a simple negotiation simulation using the Eclipse environment. It illustrates a negotiation between one buyer and one seller under the incomplete information scenario and a deadline range of [9-10]. The negotiation ends by B accepting S's offer. The iteration number corresponds to the number of runs.

B. Results and Analysis

The goal of the proposed negotiation model is to achieve results close to the best scenario which is the complete information scenario. Empirical results were recorded from *S*'s perspective and are shown in Fig. 5 and Fig. 6. Stacked lines are used to represent results, the x-axis indicates deadline ranges and y-axis corresponds to performance measures.

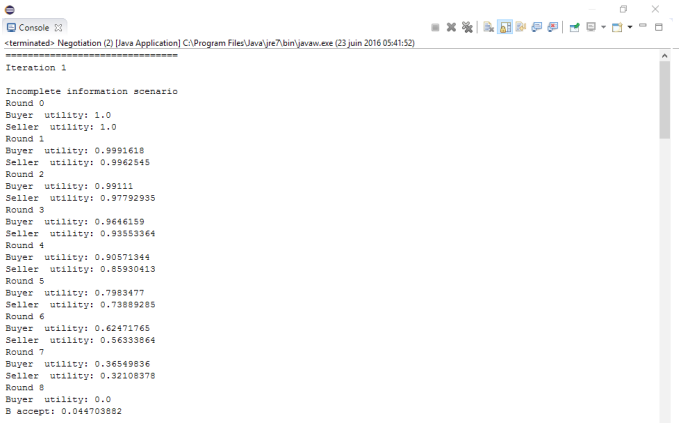


Fig. 4. Eclipse Toolbox.

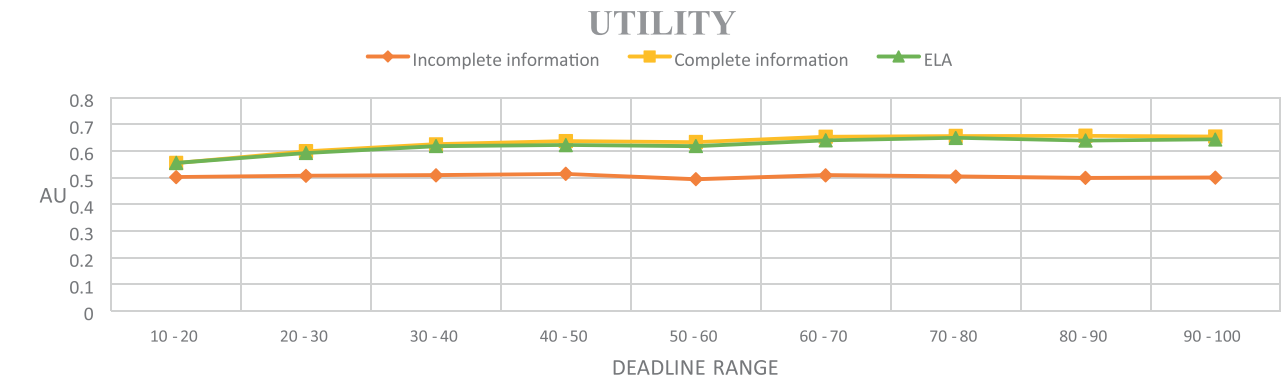
Fig. 5 depicts results of the comparison between ELA and the complete and incomplete scenarios. Obviously, ELA achieves very close results to the optimal scenario. For example, we pinpoint that ELA's average utility is equal to 0.618 for the deadline range [30 − 40], which is very close to the best scenario (equal to 0.625). ELA also achieves little faster ANS than the complete information agent (Fig. 5(c)). This is due to the approximation value of opponent's deadline and reserve points. Compared to the incomplete scenario, ELA always achieves much better AU and JU (Fig. 5(a) and Fig. 5(d)). We can also observe that our learning agent always reaches an agreement (SR = 1) (Fig. 5(b)). In contrast, in the incomplete scenario, *S* does not always reach agreements, especially for short deadlines (i.e., 10 to 40 rounds).

Fig. 6 presents different results relative to the comparison between ELA and BLA agents. Clearly, ELA outperforms BLA. In fact, ELA achieves much better AU and JU (Fig. 6(a) and Fig. 6(c)) since BLA achieves faster ANS (Fig. 6(b)) affecting the quality of its outcome. For example, ELA attains 58.938 of ANS (deadline range [60 − 70]), which is higher than the one of BLA (equal to 49.788). In term of SR, the two agents always reach agreements.

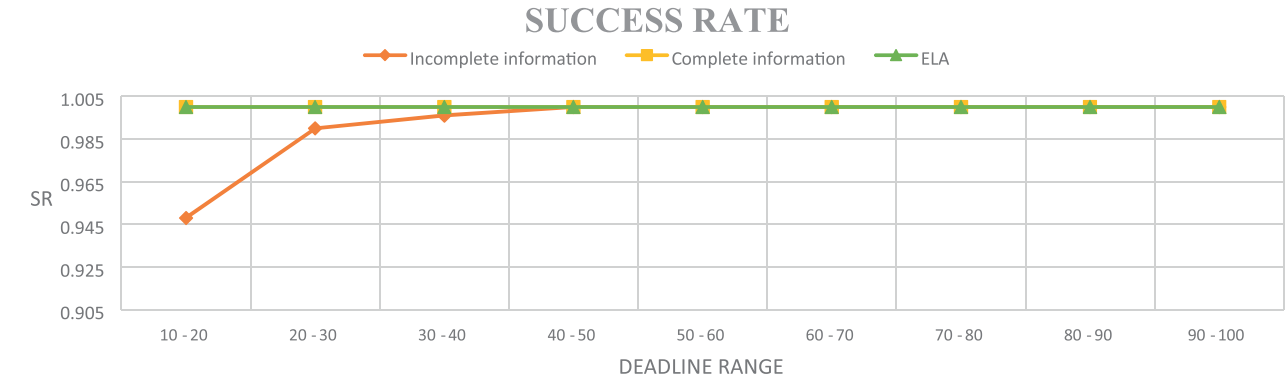
To test the cut-off points of ELA, we proposed to increase the number of issues from 4 to 5, then boosted it to 10 and 20. The results of this experiment are shown in Fig. 7 and Fig. 8.

Fig. 7 represents the success rate of ELA, BLA, complete and incomplete scenarios. It is clear that ELA's success rate follows the same behavior as the number of issues increases. Indeed, ELA achieved 0.984 successful negotiations for 5 issues, then it reached the value of 0.995, which is very close to the complete scenario, when we upgraded the number of issues to 20. While BLA failed from 10 issues since the computational complexity raises by boosting the number of prediction cells used by the Bayesian learning.

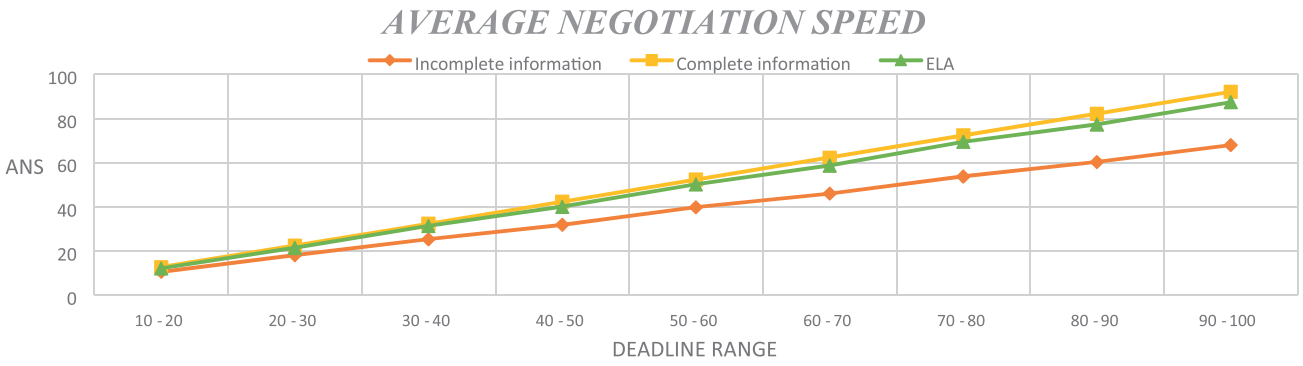
The average utility of the four negotiation strategies is depicted by Fig.8. Clearly, ELA provides better average utility than BLA since the first one adjusts the concession strategy of each issue based on the learned reserve point by DEIWO. Unlike to BLA that uses a single concession rate to adjust the concession strategy for all issues based on the learned reservation utility.



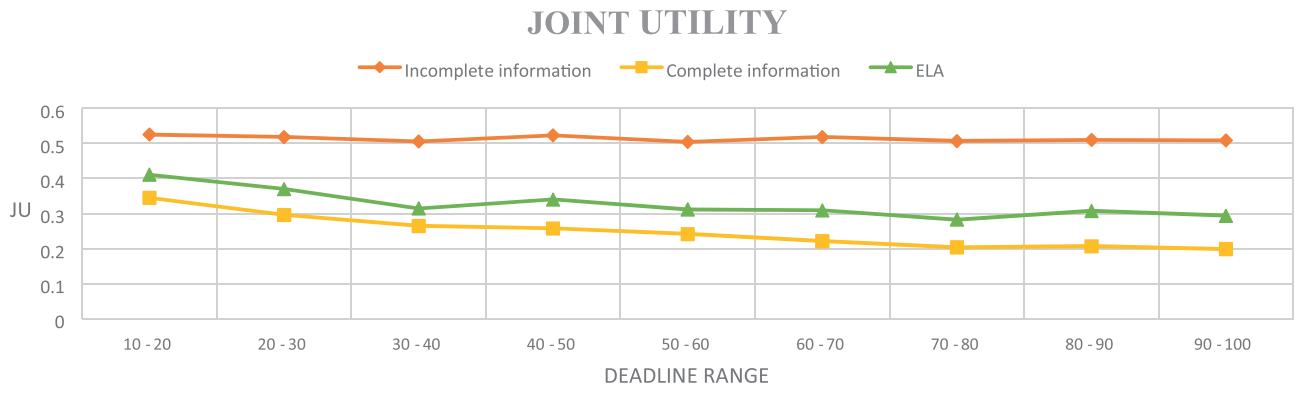
(a)



(b)

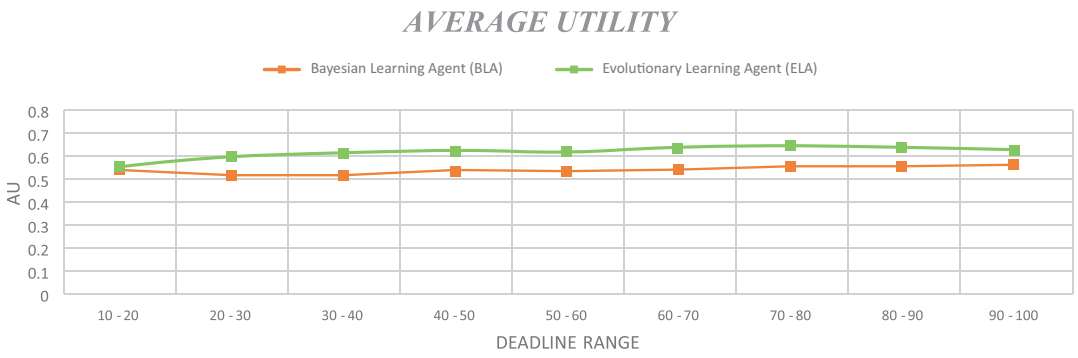


(c)

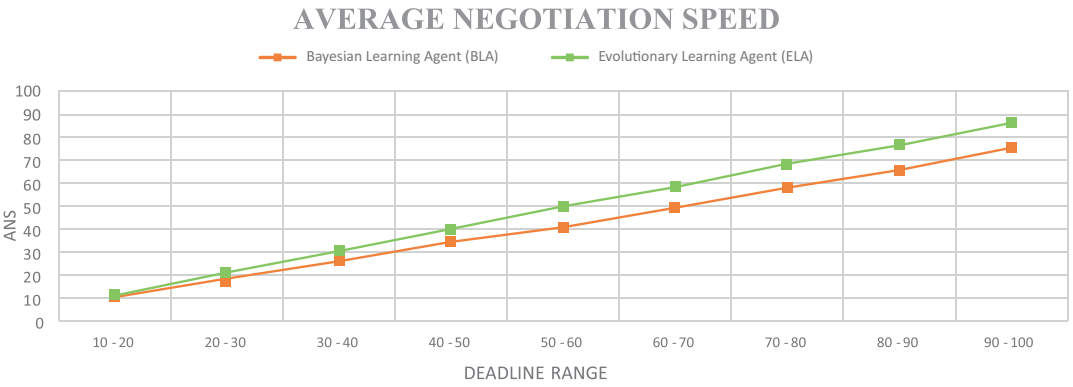


(d)

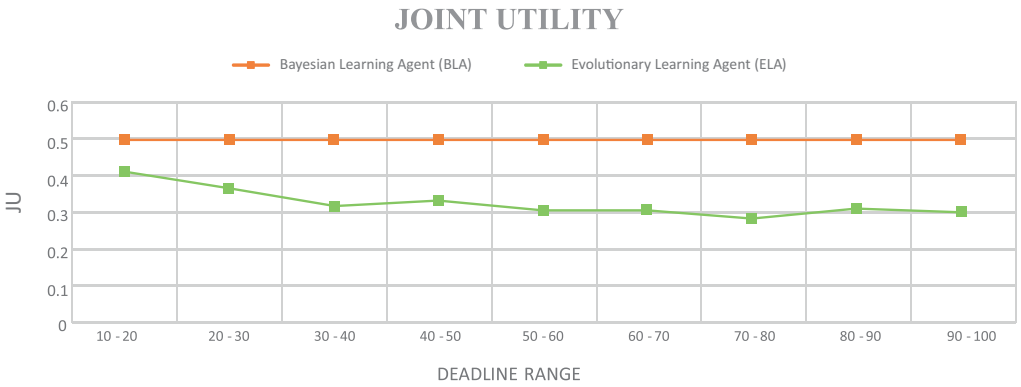
Fig. 5. ELA vs complete and incomplete information agents.



(a)



(b)



(c)

Fig. 6. ELA vs BLA.

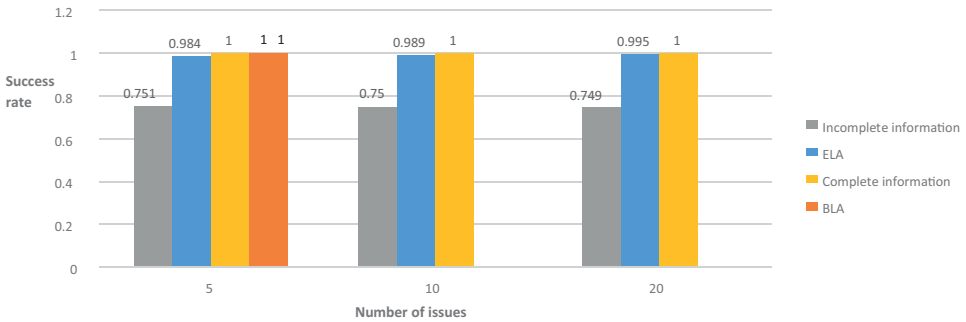


Fig. 7. Success rate for 5, 10 and 20 issues.

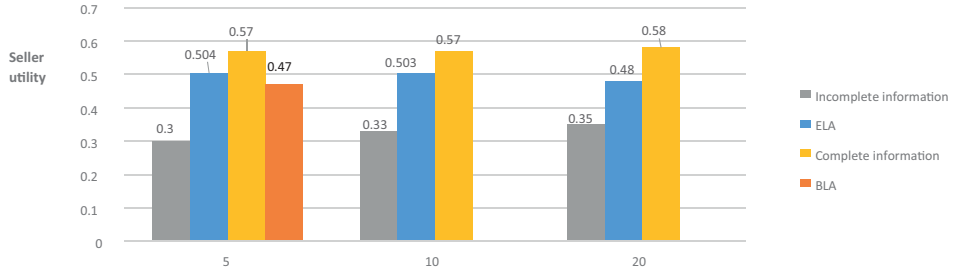


Fig. 8. Seller average utility for 5, 10 and 20 issues.

We have also compared the execution time of BLA and ELA as shown in Table VII. In fact, the execution time of BLA is equal to 3000 seconds for the case of 7 issues, while with ELA, the execution time is only equal to 3 seconds. This confirms that DEIWO used by ELA explores the outcome space faster than the combination of regression analysis and Bayesian learning used by BLA.

Table VII. Execution Time of ELA and BLA

Number of issues	BLA	ELA
1	0.093 ms	0.141 ms
2	0.150 ms	0.125 ms
3	0.39 ms	0.903 ms
4	0.851 ms	1 s : 32 ms
5	0.469 ms	0.185 ms
6	145 s	2442 ms
7	3000 s	3 s

V. Conclusion and Future Work

In this paper, we introduced the ELA agent which learns its opponent's deadline and reserve points in a bilateral multi-issue negotiation. ELA employs an evolutionary optimization algorithm in order to learn its opponent's parameters. A new concession strategy adjustment is performed to improve an agent's outcome. Empirical results showed that ELA gets very close outcomes to the best scenario. Also, we test the limit of our model in order to find out how much it can handle in term of number of issues. Our future work consists in studying in depth multi-issue negotiation and proposing a new negotiation model based on multi-criteria methods such as ANP [21] for inter-dependent issues and extend our negotiation model to multilateral negotiation. Another perspective is to extend our approach to the case of concurrent one side multilateral negotiation in which an agent may engage simultaneously many agents in multiple bilateral negotiation.

References

[1] T. Baarslag, M. J. C. Hendrikx, K. V. Hindriks, and C. M. Jonker, "Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques," Autonomous Agents and Multi-Agent Systems, vol. 30, no. 5, pp. 849–898, 2016.

[2] J. Gwak and K. M. Sim, "Bayesian learning based negotiation agents for supporting negotiation with incomplete information," in Proceedings of the international multiconference of engineers and computer scientists, vol. 1, 2011, pp. 163–168.

[3] C. C. Lee and C. Ou-Yang, "A neural networks approach for forecasting the suppliers bid prices in supplier selection negotiation process," Expert Systems with Applications, vol. 36, no. 2, pp. 2961–2970, 2009.

[4] Y. Zhou, Q. Luo, and H. Chen, "A novel differential evolution invasive weed optimization algorithm for solving nonlinear equations systems," Journal of Applied Mathematics, vol. 2013, no. 3, pp. 1-18, 2013.

[5] C. Siqi and W. Gerhard, "An efficient and adaptive approach to negotiation in complex environments," in ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012), p. 228-233, 2012.

[6] A. Rubinstein, "A bargaining model with incomplete information about time preferences," Econometrica: Journal of the Econometric Society, vol.53, no, 5, pp. 1151–1172, 1985.

[7] T. Baarslag, K. Hindriks, and C. Jonker, "Towards a quantitative concession-based classification method of negotiation strategies," in: Kinny D., Hsu J.Y., Governatori G., Ghose A.K. (eds) Agents in Principle, Agents in Practice PRIMA 2011. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, vol. 7047, pp. 143–158, 2011.

[8] P. Faratin, C. Sierra, and N. R. Jennings, "Negotiation decision functions for autonomous agents," Robotics and Autonomous Systems, vol. 24, no. 3, pp. 159–182, 1998.

[9] C. Hou, "Predicting agents tactics in automated negotiation," in Intelligent Agent Technology, 2004.(IAT 2004). Proceedings. IEEE/WIC/ACM International Conference on. IEEE, 2004, pp. 127–133.

[10] K. M. Sim, Y. Guo, and B. Shi, "Adaptive bargaining agents that negotiate optimally and rapidly," in Evolutionary Computation, 2007. CEC 2007. IEEE Congress on. IEEE, 2007, pp. 1007–1014.

[11] B. Sim, Kwang Mong andShi and Y. Guo, "Blgan: Bayesian learning and genetic algorithm for supporting negotiation with incomplete information," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 39, no. 1, pp. 198–211, 2009.

[12] C. Yu, F. Ren, and M. Zhang, "An adaptive bilateral negotiation model based on bayesian learning." In: Ito T., Zhang M., Robu V., Matsuo T. (eds) Complex Automated Negotiations: Theories, Models, and Software Competitions. Studies in Computational Intelligence, Springer, Berlin, Heidelberg, vol.435, pp.75-93, 2013.

[13] D. Zeng and K. Sycara, "Bayesian learning in negotiation," International Journal of Human-Computer Studies, vol. 48, no. 1, pp. 125–141, 1998.

[14] J. Zhang, F. Ren, and M. Zhang, "An innovative approach for predicting both negotiation deadline and utility in multi-issue negotiation," in PRICAI 2014: Trends in Artificial Intelligence. Springer, 2014, pp. 1076–1088.

[15] R. M. Coehoorn and N. R. Jennings, "Learning on opponent's preferences to make effective multi-issue negotiation trade-offs," in Proceedings of the 6th international conference on Electronic commerce. ACM, 2004, pp. 59–68.

[16] R. Marler and J. Arora, "Survey of multi-objective optimization methods for engineering," Structural and Multidisciplinary Optimization, vol. 26, pp. 369–395, 2004.

[17] G. Venter, Review of Optimization Techniques. John Wiley and Sons, Ltd, 2010.

[18] A. R. Mehrabian and C. Lucas, "A novel numerical optimization algorithm inspiredfrom weed colonization," Ecological informatics, vol. 1, no. 4, pp. 355–366, 2006.

[19] J. H. Holland, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. U Michigan Press, 1975.

[20] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," Journal of global optimization, vol. 11, no. 4, pp. 341–359, 1997.

[21] T. L. Saaty and L. G. Vargas, "Decision making with the analytic network process", International Series in Operations Research & Management Science, 95, Springer US, 2006.

R. Ayachi

R. Ayachi is an assistant professor at Ecole Supérieure des Sciences Economiques et Commerciales de Tunis (ESSEC Tunis), Tunisia and member of LARODEC laboratory (ISG Tunis). Her research interests include decision and reasoning issues.

H. Bouhani

H. Bouhani is a master student at Institut Supérieur de Gestion (ISG Tunis), Tunisia. His research interests are automated negotiation and Multi-agent systems.

N. Ben Amor

N. Ben Amor is a professor at Institut Supérieur de Gestion (ISG Tunis), Tunisia. Her research interests include reasoning and decision making under uncertainty, graphical models, and machine learning.

EVEN-VE: Eyes Visibility Based Egocentric Navigation for Virtual Environments

Muhammad Raees*, Sehat Ullah

Department of Computer Science and IT, University of Malakand (Pakistan)

Received 25 May 2018 | Accepted 1 August 2018 | Published 10 August 2018

Abstract

Navigation is one of the 3D interactions often needed to interact with a synthetic world. The latest advancements in image processing have made possible gesture based interaction with a virtual world. However, the speed with which a 3D virtual world responds to a user's gesture is far greater than posing of the gesture itself. To incorporate faster and natural postures in the realm of Virtual Environment (VE), this paper presents a novel eyes-based interaction technique for navigation and panning. Dynamic wavering and positioning of eyes are deemed as interaction instructions by the system. The opening of eyes preceded by closing for a distinct time-threshold, activates forward or backward navigation. Supporting 2-Degree of Freedom head's gestures (*Rolling* and *Pitching*) panning is performed over the xy-plane. The proposed technique was implemented in a case-study project; EWI (*Eyes Wavering based Interaction*). With EWI, real time detection and tracking of eyes are performed by the libraries of OpenCV at the backend. To interactively follow trajectory of both the eyes, dynamic mapping is performed in OpenGL. The technique was evaluated in two separate sessions by a total of 28 users to assess accuracy, speed and suitability of the system in Virtual Reality (VR). Using an ordinary camera, an average accuracy of 91% was achieved. However, assessment made by using a high quality camera testified that accuracy of the system could be raised to a higher level besides increase in navigation speed. Results of the unbiased statistical evaluations suggest/demonstrate applicability of the system in the emerging domains of virtual and augmented realities.

Keywords

Gesture Recognition, Interactive Systems, Computer Interfaces, Virtual Reality.

DOI: 10.9781/ijimai.2018.08.002

I. Introduction

Navigation is the user's locomotion inside a virtual environment to reach to and inspect an object of interest from close proximity. The two broad metaphors of navigation are Egocentric and Exocentric. Where the former is analogous to moving the world in reference to user's input, the latter is moving the user's viewpoint inside the virtual world. Navigation often becomes a preliminary step to perform other interactions like selection, scaling and rotation. Furthermore, a wide spread 3D scene cannot be viewed from a single static look of virtual camera. Navigation is, therefore, important to explore different parts and portion of a synthetic world. Though, a number of navigation techniques have been proposed based on mouse/keyboard, their traditional menu-commands based interface fails to bring intuition and naturalism in VE. Low-cost vision based tracking on the other hand provides an applicable platform to devise a flexible interface for Human Computer Interactions (HCI). Enormous Gestures based techniques in the literature of 3D interactions provides estimable degree of realism. The latency and fatigue involved in posing gestures, however, are the challenges yet to be covered. Particularly, the time lapse in posing of gesture, adds in imbalance bandwidth between human-computer dialogues which in turns badly affects immersion.

As eyes-based interaction is faster than speech and gestures based interactions [1] therefore, interaction via eyes postures can reduce the speed mismatch. This paper is an attempt to present a convenient and cost-effective eyes' wavering based interaction. The technique; Eyes Visibility based Egocentric Navigation in Virtual Environment (*EVEN-VE*), follows gestures of eyes which are speedier than any other gesture made by any other part of the body. Though worthy research has been carried out on iris recognition but iris varies from person to person, therefore it cannot be used for interaction [2]. The proposed system merely considers wavering of eyes hence processing time is not wasted in tracing user's specific features. Only flickering of eyes and tilting of head are traced by the system to interact with the designed 3D virtual world. *Navigation* is activated by a slight flickering-gesture of both the eyes while *Panning* is performed by tilting of head along the look-vector. To make user aware about the beginning of any action, all interactions are initiated when both the eyes are opened. At the closing state of eyes, threshold times for forward and/or backward navigation are checked. Keeping both eyes closed for extended time greater than that of normal flickering, activates forward navigation. Upholding the closing state of eyes for another equal interval leads to activation of reverse navigation. Panning is performed based on eyes position in dynamic scanned frames. As eyes' positions change with tilting of head, the 2DoF head's movement; *Rolling* and *Pitching*, are traced to pan along z-axis or x-axis respectively. The case-study project EWI was designed using the open libraries to implement and evaluate the system. Based on algorithm of the technique, EWI supports multiple speed sectors for navigation. The system was twice evaluated in two separate sessions. The first session was to evaluate accuracy and learning effect of the system. In second evaluation session, the system performance was assessed in different navigation sectors using two

* Corresponding author.
E-mail address: visitrais@yahoo.com

different cameras. Results of the second evaluation session proved that with a high quality camera both speed and accuracy of the system can be improved. The proposed technique is applicable for all the basic navigation tasks; exploration, searching and inspection [3]. Moreover, due to the least muscular strive and strain involved in wavering of eyes and tilting of head, less fatigue is guaranteed.

The paper is organized into 7 sections. Related work for Navigational interaction is discussed in Section II. Section III elaborates the algorithm of the technique in detail, section IV is about implementation and evaluation details. In Section V, the system is compared with state-of-the-art navigation techniques. Section VI discusses applicability of the technique in VR. Conclusion and future work is discussed in the last section.

II. Literature Review

No doubt, the ceaseless progress in processing and storage of computer systems has made possible the emergence of virtual reality applications in every field. To make possible convenient dialogue with such applications, realistic way of interactions is required. Being the most commonly used interaction task, navigation is focused by most of today's research works to make exploration feasible and flexible. Most of the navigation techniques in the literature of virtual reality are either based on mouse/keyboard or need Head-Mounted Display (HMD) [4]. Where the former techniques lack immersion, the techniques based on HMD always remain a second choice because of the high cost. The bimanual Multi-Finger gestural navigation of [5] works on a constrained tabletop surface and is not suited for hand-held devices. Another hand gesture based navigation; NuNav3D [6] preliminarily needs whole body pose estimation. This makes the system not only difficult to use but also 79% slower than Joypad based navigation [6]. The technique of lee et al. [7] consumes a large amount of processing for finger action recognition as the system ought to pass from three hefty stages; skin-color detection, k-cosine based angle detection and contour's analysis each for finger's state, position and direction. Furthermore, due to variable finger's thickness, accuracy of the system varies from user to user.

With the Drag'n Go [8] the position of screen's cursor casts ray to the target for navigation. The technique is applicable only for navigation in large empty space while askew navigation is not supported. In another related approach [9] dragging mouse in a specific direction shifts the look-at vector of virtual camera. Thus for a lengthy travel the practice is needed to be repeated again and again. Moreover, in the way if the ray collides with an object, navigation is halted and the collided object is mistakenly inspected.

In the head-directed navigation [10], speed and direction are calculated from the pose and direction of the user's head. Though intuition of the system is commendable but unintentional and casual head movements are fallaciously treated as navigation commands. The FmF (Follow my Finger) [11] presents MC (Management Cabin) model for navigation. The system projects 3D view of a virtual world on a 2D table-top device but suffers from disorientation problem.

Trackers based gestures recognition [12] and [13] are the noteworthy navigation techniques which work on natural gestures of hands. Both of the systems treat virtual scene as big object that can be grabbed and moved with hands. The techniques are interesting but remain a rare option due to the hand-worn overload.

The walking in place of [14] correlates pace of the natural walking with navigation speed in synthetic environment. The approach cannot be used while setting on chair. Chest expansion and Tilting of spine are treated as commands for navigation in [15]. The system needs a complex setup of sensors implementable only inside a controlled environment. Similarly, the foot-based interface for navigation suggested in [16] requires a cumbersome setup of waist-based magnetic trackers with conveyer belt for navigation.

Fiducial markers have also been effectively used for navigation. In the DeskCube system [17] different markers are glued on different faces of a cube. The system works fine when the cube is still but fails to cope with occlusion and blurring while moving the cube. Similarly navigation system based on QR codes necessitates the placement of navigation device closed to the marker [18].

Using the leap motion based system [19] for navigation, user needs to activate commands via buttons within a limited space. The fingers based locomotion; FWIP (Finger Walk In Place) [20] necessitates touch of the fingers on display screen for navigation. The system is applicable only for mobile based VEs. The pointing technique of [21] uses two fingers; one for viewing and the other for direction. Though it successfully avoids the mistakes of gaze-directed navigation but speed controlling is its main challenge.

As eyes tracking based interfaces ensure quick and effortless interactions [22], therefore eyes based interaction is becoming prominent in the domain of VR. The system of [23] traces gaze direction after pushing a button to reach to an object. The system needs HTC Vive with a Tobii Eye-tracker and is therefore applicable inside a lab environment. With EyeScout system [24] eye trackers are mounted over a rail system. The trackers are made aligned with user's lateral movement by a computational method. Based on 9-point polynomial algorithm, the HoloLens based system of [25] maps eyes position to a plane in which calibration points are displayed. The system needs a constrained setup containing wireless camera, eye tracker and IR (infrared) filter. The lazyNav system [26] supports other interactions besides navigation using HMD and trackers. Extensive research in the literature has been made to enhance efficiency of eyes tracking. The algorithm proposed by [27] for measuring focal distance in immersive CAVE (Cave Automated Virtual Environment) is to improve precision of eyes tracking. Moreover, as eyes-tracking based interaction is supposed to be the VR's next frontier [28], models for accurate detection of eyes using next generation devices have also been suggested [29].

III. EVEN-VE: The Proposed System

Interactions via gestures of eyes are becoming an effective alternative due to the availability of portable head-mounted eye trackers. Nevertheless, the clumsy setup of wires, as shown in Fig. 1 restricts the use of such systems only inside the premises of a lab.

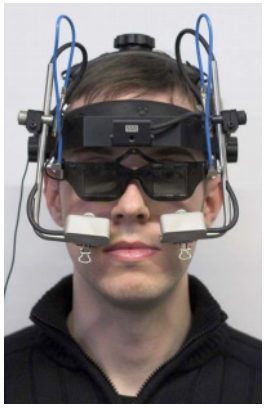


Fig. 1. The complex setup of eyes trackers [20].

Aim behind this research was to introduce a cheaper and easy-to-use interaction technique. The proposed system tracks eyes' positions and orientations through an ordinary camera without the use of any trackers

or bunch graphing [30]. A slightly longer than the normal flickering of eyes is assumed to start forward-navigation. If navigation is already activated then the same gesture is used to halt it. Backward navigation is activated by keep closing both the eyes for a longer time than the forward navigation. After detecting both the eyes, Mid-Point (MP) is calculated to get central point between the eyes [31]. MP is calculated from the horizontal and vertical positions of the eyes as,

$$MP.x = (LeftEye.x + RightEye.x)/2 \quad (1)$$

$$MP.y = (LeftEye.x + RightEye.x)/2 \quad (2)$$

Where *XEye* are point variables representing mid position of the eyes (eyeball with sclera). *X* is representing both *Left* (eye) and *Right* (eye).

A straight virtual ray starting from MP is supposed to represent user's direction. The ray's hit position is represented by a pointer in the designed 3D world, see Fig. 2. Before starting navigation, the pointer can be used to set position of the virtual camera. By default, navigation is performed straight inside the look vector.

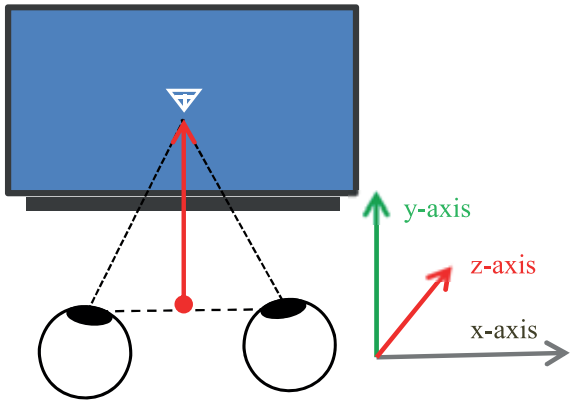


Fig. 2. Eyes with MP and virtual ray casted on screen.

Using 2D positions of the eyes in the initial five frames, Eyes Initial Average Area (*EIAA*) and Normal Speed Zone (*NSZ*) are defined. NSZ is a limited area where navigation with normal speed is performed whereas *EIAA* is calculated as,

$$EIAA = (\ LeftEye\_AA + \ RightEye\_AA) \ / \ 2 \quad (3)$$

Where *X_AA* is calculated from eyes' area in the initial five frames as,

$$X\_AA = \frac{\sum_{i=1}^{5} X\_Area_i}{5} \quad (4)$$

Area of an eye, as shown in Fig. 3, is calculated from the length and width of eye (eyeball including iris and sclera) as,

$$X\_Area = X.length \times X.width \quad (5)$$

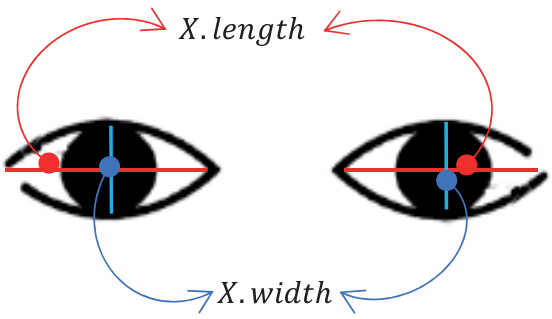


Fig. 3. Length and width of the eyes.

Navigation speed can be increased or decreased based on areas of the eyes. Moving head forward along z-axis results into increase in areas of eyes which in turn makes navigation speedier and vice versa.

Panning is performed by gentle tilts of head towards left or right along z-axis (*Rolling*). As conceivable, a slight bending towards left activates left panning whereas right panning is accomplished by bending head towards right. *Up panning* (flying) and *Down panning* (landing) are performed by *Pitching* gesture; raising and bowing head along y-axis respectively. With both *Rolling* and *Pitching* gestures of head, positions of eyes change. Calculating the dynamic positions of eyes against MP, horizontal or vertical panning is performed.

A. The System Architecture

The system designed for the implementation starts with a virtual world at front end with different 3D objects. At the backend, a region of interest (ROI) based on skin color is extracted from the scanned image. At the detection of eyes, MP is calculated and coordinates mapping is performed to set virtual camera position in the environment. As long as eyes are detected, the virtual pointer can move freely with the movement of eyes. If the eyes closing time extends the normal blinking duration, navigation is activated. Tilting of head, alters the eyes' vertical positions from the central MP, activates panning. Schematic of the proposed system is shown in Fig. 4.

B. Face Segmentation

In order to reduce processing cost and to avoid possibility of false detection first a ROI image (ROI-Image) is extracted. ROI-Image is supposed to contain both the eyes. This segmentation is carried out on the basis of skin color. As the YCbCr space thresholding provides best skin color segmentation [32], hence the YCbCr model is used for face extraction. Each scanned RGB frame is thus converted to YCbCr space (Frame_YCbCr) as,

$$\text{Frame\_YCbCr}\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.1 & 128 & 24.8 \\ -37.3 & -74 & 110 \\ 110 & -93.2 & -18.2 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (6)$$

From the binary image of Frame_YCbCr, see Fig. 5, *ROI-Image* is extracted with rows ‹*m*› and columns ‹*n*› using our designed algorithm [33] as,

$$\text{ROI\_Image}(m, n) = \left(\bigcup_{r=Dm}^{\text{Frame\_YCbCr.Row(0)}} (\text{Frame\_YCbCr}) \ , \ \bigcup_{c=Lm}^{\text{Frame\_YCbCr.Colum(Rm)}} (\text{Frame\_YCbCr}) \right) \quad (7)$$

Where *Lm, Rm* and *Dm* represents *Left-most*, *Right-most* and *Down-most* skin pixels.

Fig. 5. (a) RGB and (b) YCbCr models of a scanned image.

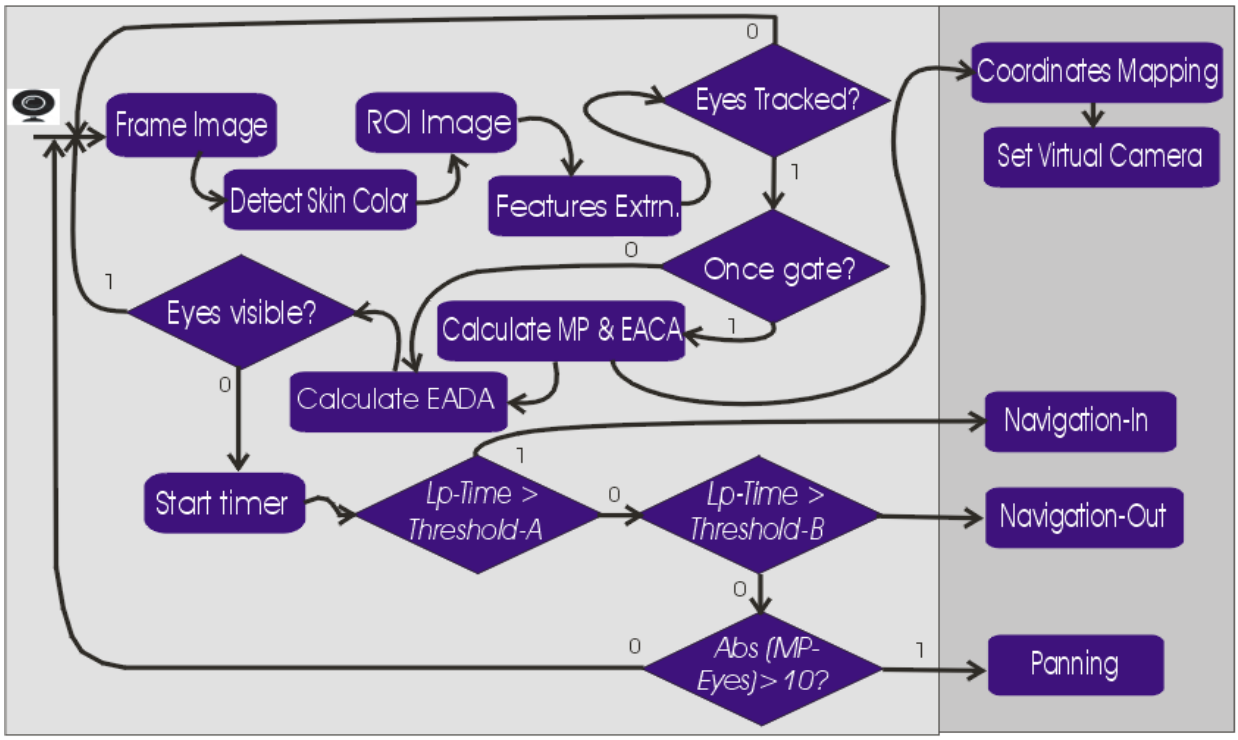


Fig. 4. Schematic of the proposed system.

C. Eyes Detection

The most accurate cascade classifier for eyes [34] is used for the detection and tracking of eyes. The algorithm detects eyes rapidly in live video and is more accurate than LBP [35]. With Haar-cascade, first the system is trained by several numbers of positive images and negative images and then rectangular features are selected by Ada-boost. Each time window of the eyes are moved over the scanned image and for each subsection, the Haar-like features are calculated. Greater the matching of positive features, higher will be the eyes recognition.

D. Coordinates Mapping

The notable challenge in implementation of the technique was to locate and move pointer of the virtual environment through pixels position of MP. The image frame of OpenCV starts from top left (0,0) while in OpenGL, (0,0) lies at the center of clipping area. To harmonize the dissimilar coordinate systems, we devise our four mapping functions m_1, m_2, m_3 and m_4 [36], as described in equations (8), (9), (10) and (11). The image frame is virtually split into four regions R_1 to R_4 as shown in Fig.6, where mapping for a region R_n is made by the corresponding m_n taking x and y of a pixel of R_n as independent variables.

$$m_1(x,y) = ((MP.x - (Tc/2)/(Tc/2), (Tr/2 - MP.y)/(Tr/2)) \quad (8)$$

$$m_2(x,y) = ((MP.x/Tc), (Tr/2 - MP.y)/(Tr/2)) \quad (9)$$

$$m_3(x,y) = ((MP.x - (Tc/2)/(Tc/2)), (MP.y/Tr)) \quad (10)$$

$$m_4(x,y) = ((MP.x - (Tc/2))/Tc/2), (MP.y/Tr)) \quad (11)$$

In the above functions, *MP.x* and *MP.y* represent x and y positions of the traced pixel. *Tr* and *Tc* represent total rows and total columns respectively.

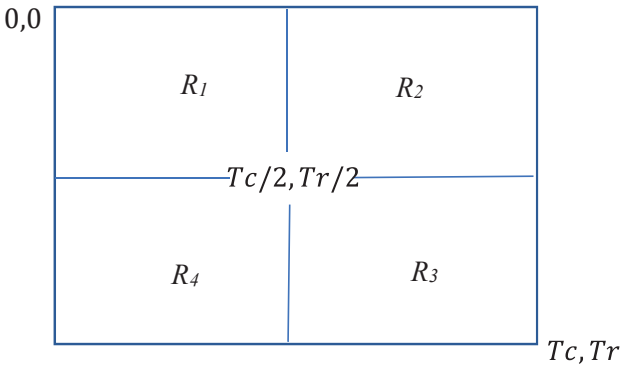


Fig. 6. Virtual division of the image frame.

E. Navigation

Navigation is the moving of a virtual camera in 3D VE towards or away from a look-at point. In the proposed system, navigation is activated/deactivated by a single prolonged blink of eyes. The normal blinking rate is approximately 1/3 of a second in which eyes remain closed for about 300-350 milliseconds. To avoid detection of normal unintentional blinking the closed state of eyes for less than 400 milliseconds are omitted. When navigation state is set enabled, user is informed by a beep and actual movement is started when eyes are opened. Keeping eyes closed after the first beep will activate reverse navigation. Backward navigation is distinguished by a lengthy beep sound. Textual information about each state is also displayed on the top of the virtual environment. By default navigation is performed in a straight line but user can change direction of navigation dynamically through eyes movement. The viewable scene is divided into 180 degrees as shown in Fig. 7. Moving the pointer thirty pixels towards right will result into a decrease of thirty degrees while moving thirty pixels towards left will make an increase of thirty degrees. The 30-pixels to 30-degress ratio can be decreased to divide the scene into further navigation paths. The virtual division of scene based on the position of the pointer (a 3D wire cone) is shown in Fig. 7.

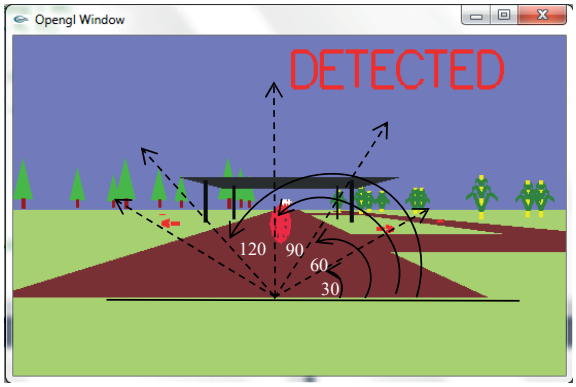


Fig. 7. Virtual division of the scene based on the pointer.

Navigation Speed

One of the main problems with VE navigation is speed control, the absence of which may either lead to disorientation or slow velocity [37]. Based on the quality of camera and distance of eye from it, various finite speed sectors '*Sn*' can be defined. Initially, the system was implemented and tested with the default three sectors, NSZ, one ahead and one behind the NSZ, see Fig. 8 and Fig. 9. Entrance into a particular sector *Sn* is determined from the Eyes Average Dynamic Area (*EADA*). Dynamic Areas (*DA*) of the eyes are calculated on the fly as,

$$LeftEye\_DA = LeftEye.Height * LeftEye.Width \quad (12)$$

$$RightEye\_DA = RightEye.Height * RightEye.Width \quad (13)$$

$$EADA = (LeftEye\_DA + RightEye\_DA) / 2 \quad (14)$$

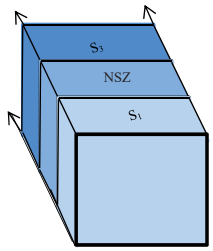


Fig. 8. First three different speed sectors for navigation.

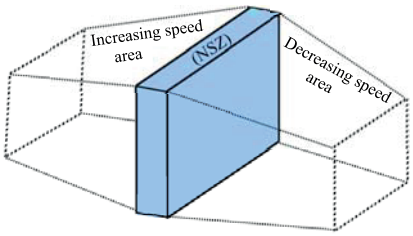


Fig. 9. NSZ and speeding areas over the look-vector.

With forward head's movement ahead of NSZ, *EADA* increases and hence speed is increased. Similarly, by moving head behind the NSZ, *EADA* decreases which results into decrease in speed. The '10' pixels extension in area is to avoid erroneous/unintentional increase or decrease. The pseudo-code for detecting variation in *EADA* and accordingly setting navigation speed is as follow,

$$if(EADA >= EIAA + 10)$$
$$Navigation\_Speed += 0.5$$
$$EIAA = EADA$$
$$if(EADA <= EIAA - 10)$$
$$Navigation\_Speed -= 0.5$$
$$EIAA = EADA$$

When the condition of forward or backward head's movement meets, current area of eyes (*EADA*) becomes previous area (*EIAA*); see second statement of the condition in the pseudo-code. With this, algorithm of the system supports 'n' sectors beyond and behind the *NSZ* provided that eyes are detectable in a sector $S_i \mid i \in \{1, 2, 3, ., ., n\}$. Eyes recognition and the extraction of area of eyes depend on the quality of camera and distance of user from camera, see Fig. 10.

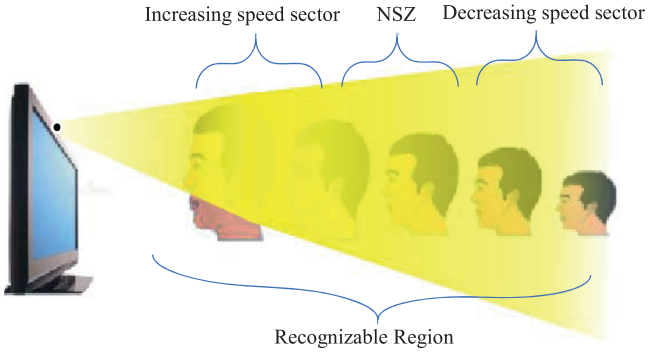


Fig. 10. The detectable region of camera with different speed sectors.

F. Panning

Panning is the translation of the camera eye and the look-at points horizontally or vertically. To avoid the possibility of disorientation, panning is enabled only inside NSZ. Furthermore, panning is performed with a constant moderate speed, independent from navigation's speed. Horizontal Panning is performed by slightly bending the head over the z-axis (Rolling gestures of head). With rolling gesture of head, one of the eye's position on y-axis increases whi.le that of the other decreases. Similarly, with vertical movement of head (pitching), x and y positions of both the eyes are changed. The changes in positions of eyes are traced dynamically to perform up or down panning by changing y-coordinate of virtual camera. All this is measured against the vertical and horizontal position of MP as shown in Fig. 11, Fig. 12 and Fig. 13.

The pseudo-code for left and right panning is given as,

$$if\ (abs(LeftEye.y - MP.y) > 10 \mid\mid abs(RightEye.y - MP.y) > 10)$$
$$if(LeftEye.y < MP.y \&\& RightEye.y > MP.y)$$
$$Panning\ along\ +ve\ x\_axis$$
$$if(LeftEye.y > MP.y \&\& RightEye.y < MP.y)$$
$$Panning\ along\ -ve\ x\_axis$$

$$if(abs(LeftEye.y - MP.y) > 20 \&\& abs(RightEye.y - MP.y) > 20$$
$$if(LeftEye.y > MP.y \&\& RightEye.y > MP.y)$$
$$Panning\ along\ +ve\ y\_axis$$
$$if(LeftEye.y < MP.y \&\& RightEye.y < MP.y)$$
$$Panning\ along\ -ve\ y\_axis$$

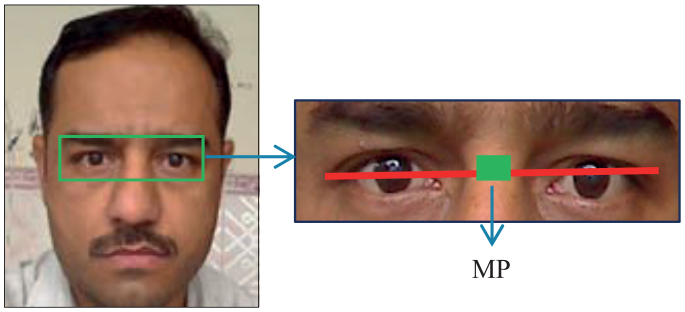


Fig. 11. Eyes tracking and setting of MP between the eyes.

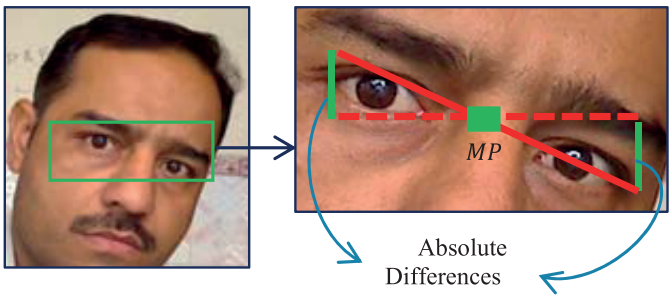


Fig. 12. Tilting of head towards Right (rolling) for horizontal panning.

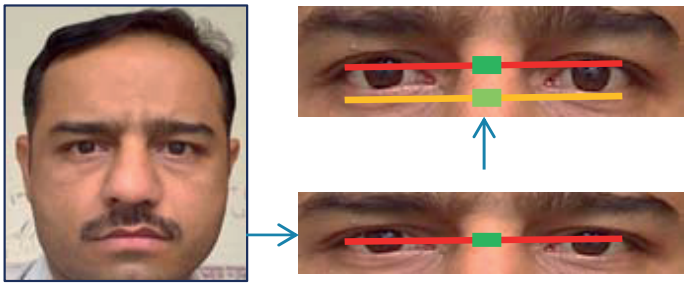


Fig. 13. Head's up gesture (pitching) for vertical panning.

IV. System Implementation and Evaluation

The proposed system was implemented in a Visual Studio project; EWI (Eyes' Wavering based Interaction). The frontend virtual scene of the EWI was designed in OpenGL. Activation of the system is conditioned to visibility of both the eyes. User is constantly informed about activation of the system by the text *"Activated"* displayed in upper part of the scene. Similarly, about left and right panning user is informed by the *"Turning"* text at the respective side of the scene as shown in Fig. 14. The system was tested by sixteen male participants. Two trials were performed by each participant for each of the four pre-defined tasks.

A. Testing Environment

The 3D environment designed for the evaluation contained four routes as shown in Fig. 13 where a robot made of cubes represents the user's position in the environment. Symbolizing finishing point of the scene is marked by a board with text "Stop". Each of the routes leads to the one Stop-board. To immerse users and to have a perception of navigation and panning, the scene contained different 3D objects at different positions.

Route-1: Straight pathway leading to Stop-board.

Route-2: Right-Straight-Left pathway.

Route-3: Left-Straight-Right pathway.

Route-4: Up-Straight-Down pathway to have a flying effect to navigate over the bandstand. Although, this route is not rendered in the scene, to avoid complication, but user can use it. A 2D model of all the routes is shown in Fig. 15.

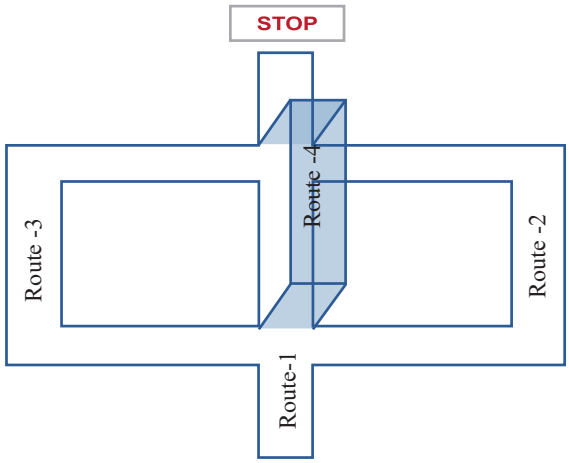


Fig. 15. 2D Model of the routes.

B. Evaluation Procedure

Each one of the participants was briefed about the environment and the interactions. They were guided by a demonstration about how to properly perform navigation and panning. Before actual trials, pre-trials were performed by each participant both for navigation and panning. For better detection of eyes, those wearing spectacles (two of the participants) were requested to remove glasses. Users were guided to hit the *'r'* keyboard key to restart a current incomplete task or to perform a new task. At the completion of a trial, *Enter* key was to be pressed to reset the entire system of EWI for a new participant. False detection and wrong turns were counted as errors. All the trials were performed inside the university IT lab in a moderate lighting condition.

C. Evaluation Tasks

Participants were asked to perform the following four interaction tasks in the designed 3D environment.

Task-1: Touching the Stop-board using Route-1 and then back to starting point following the same route.

Task-2: Touching the Stop-board using Route-2.

Task-3: Touching the Stop-board using Route-3.

Task-4: Touching the Stop-board using Route-4 and then back to starting point using the same route.

D. Results of the First Evaluation Session

Navigation-In is the forward while Navigation-Out is the backward movements inside the VE. Left/Right panning is turning, with shift of virtual camera towards the respective direction accordingly. Task-1 tests Navigation-In and Navigation-Out. Task-2 and Task-3 are to evaluate Navigation-In, Right panning and Left panning. Task-4 assesses Up panning, Down panning and Navigation. Missed detection or false detection of the system after posing the required gestures were counted as errors. With this setup, overall accuracy rate for all the 448 interactions, as shown in Table I, was 91% with average Standard Deviation (SD) 0.5.

TABLE I. STATISTICS OF THE INTERACTION TASKS

Interaction	Correct	Total	%age
Navigation-In	123	128	96
Navigation-Out	121	128	94
Panning Left	61	64	95
Panning Right	60	64	93
Panning Up	28	32	87
Panning Down	27	32	84
Total	420	448	**91**

Mean of the %age accuracy with standard deviation of the interactions are shown in Fig. 16. Though accuracy varies slightly for the first three interactions ($M_{Navigation\text{-}In}$ = 96, $M_{Navigation\text{-}Out}$ = 94.5 and $M_{Left\text{-}Panning}$ = 95.3), range of SD for all the interactions remained as low as 0.5. The less recognition of the last two interactions ($M_{Up\text{-}Panning}$ = 87.5 and $M_{Down\text{-}Panning}$ = 84.3) were due to the quicker move of eyes or of head. This implies that the system performance can be further improved with high efficiency camera.

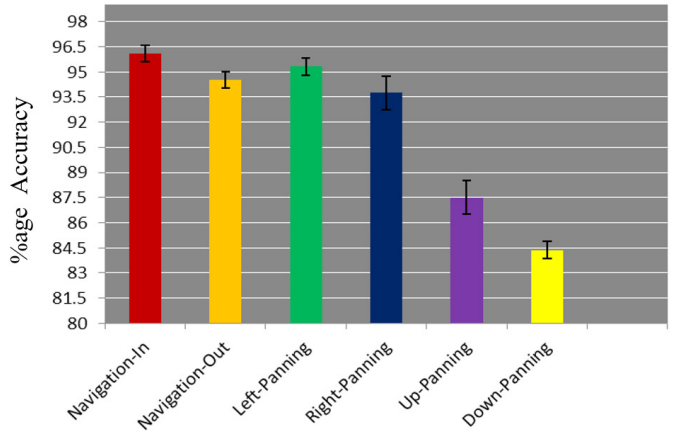


Fig. 16. %age accuracy with SD of the interactions.

E. Learning Effect

The learning effect was measured from the errors occurrence rate. Paired two sample *T*-test was used to analyze differences in means of the two trails. With null hypothesis (H_0) we assumed that mean difference (μ_d) is 0. The hypothesis was rejected as there was a significant difference between the outcomes of Trail-5 (M=34.5, 21.3) and Trail-6 (35.5,21.5) conditions; (t(5)=-3.8, p=0.0117). The graph indicating this vivid decrease in error is shown in Fig. 17.

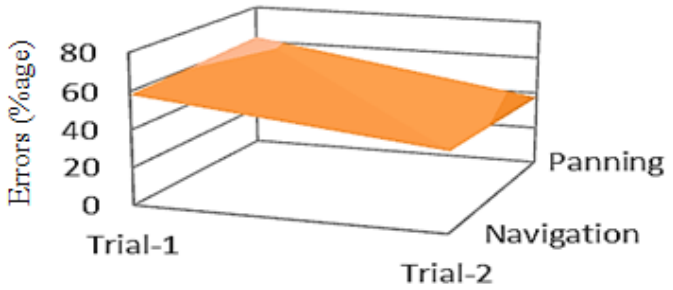


Fig. 17. The % age of errors occurred in the trials.

F. Assessment of Navigation Speed

In a separate evaluation session, performance of the system was assessed for different navigation speed in different sectors. Twelve participants, mean age 29 (SD=4.9, Range=17) familiar with image based HCI, were invited to the session. Two cameras, different in efficiency; HP Webcam and Logitech USB webcam, were used in the evaluation process. Participants were guided to perform the following tasks in a single trial using Route-1.

Task-5: Touch the Stop-board as quickly as possible.

Task-6: Get back to the starting point as slowly as possible.

The evaluation session was arranged in two phases with a short break in the middle. During the first phase, ordinary camera was used for tracking while in second phase the high quality, Logitech external camera. To properly evaluate individual trial, code of the system was updated to display number of the Highest Achieved Sector (*HAS*) and to stop detection when a trial completed. *HAS* is the highest sector for navigation in either direction and is obtained by incrementing the previous sector by 1.

$$HAS \in \{1, 2, 3, ., ., n\}$$

At the end of each trail, *HAS*, errors and task completion time were noted for statistical analysis. Results of the twelve trails of first phase are shown in Fig. 18.

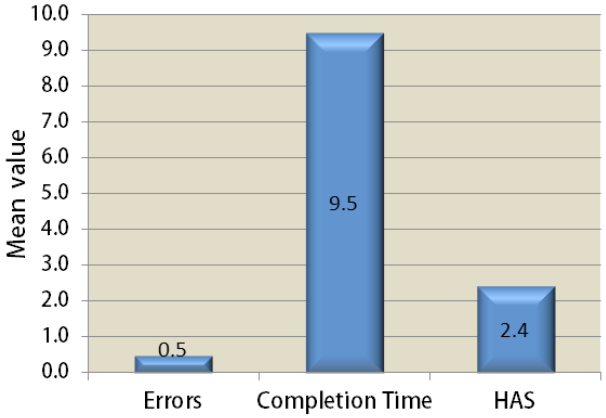


Fig. 18. Average of errors occurred, task completion time and *HAS* of the first phase trials.

By using the high quality camera, navigation in second phase was mostly performed in higher sectors (increase in *HAS*). Hence, the obtained average completion time was reduced 45%, see Fig. 19.

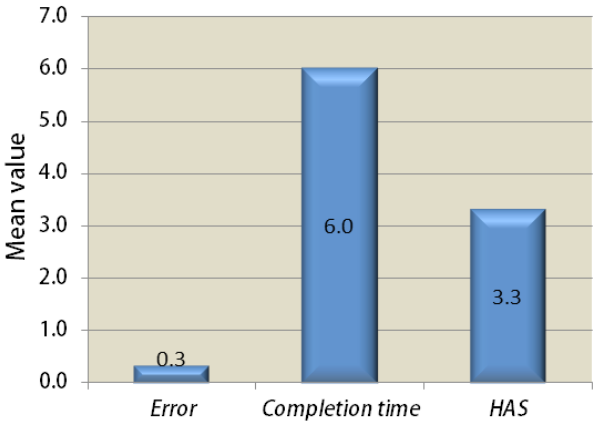


Fig. 19. Average of errors occurred, task completion time and *HAS* of the second phase trials.

As per the outcomes of test statistics, there exist a significant difference between mean completion times of tasks using ordinary camera (M=9.5,SD=S0.788) and quality camera (M=6.0, SD=S0.92), t(11)=11.1, p=0.0001. Similarly, a significant difference was reported between HAS achieved using ordinary Camera (M=2.14, SD=S0.45) and quality camera (M=3.3, SD=S0.6), t(11)=-4.7, p=0.0006.

Pearson's correlation (*r*) was used to measure the relationship between *Errors* and *HAS*. A moderate correlation was demonstrated while using ordinary camera (r=0.39), see Fig. 20. A downhill (negative) correlation was obtained by using the quality camera (r = -0.5), see Fig. 21. The diversity of *'r'* values therefore, testify the fact that error rate could be reduced with the use of a high quality camera. To avoid disorientation, turnings are made with a constant moderate speed independent of navigation speed.

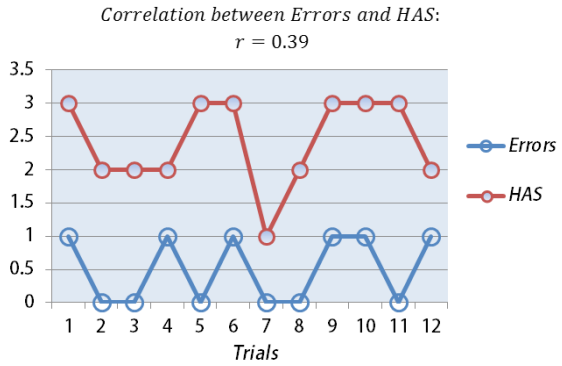


Fig. 20. The HAS and errors of the first phase trials.

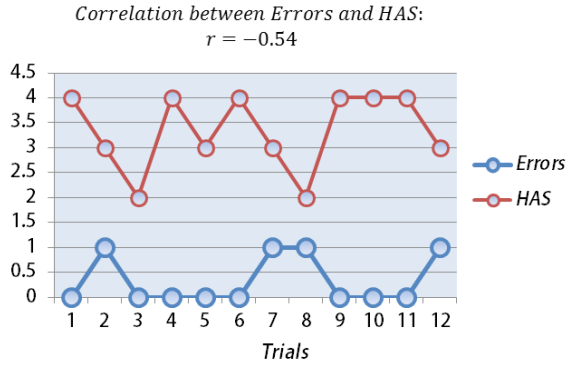


Fig. 21. The HAS and errors of the second phase trials.

G. Subjective Analysis

A questionnaire was presented to the user at the end of evaluation session to measure the four factors; Fatigue, Naturalism, Suitability in VEs and Ease of Use. The post-assessment questionnaire is shown in Table II.

TABLE II. The Post-assessment Questionnaire About the Four Factors

Question No.	Statement	Your response (tick one)?				
		Strongly agree	Agree	Indifferent	Disagree	Strongly disagree
1	I got tired after using the system					
2	The interactions performed were analogous to real world actions					
3	The technique is suitable for VR/AR applications.					
4	As a whole, the system was easy to use.					

Percentage of user's response to the four factors are shown in Fig. 22.

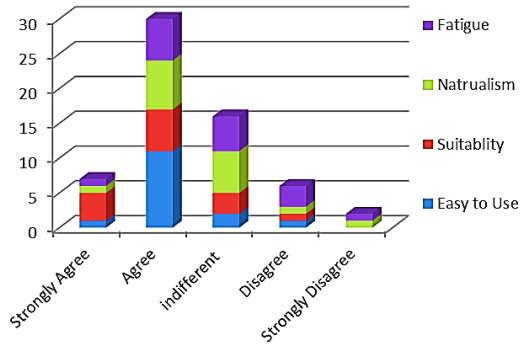


Fig. 22. Participants' response about the proposed system.

V. Comparative Study

To systematically compare the approach with other techniques, we follow the standards presented by [38] and [39] for navigation in VE. Based on the information obtained from the literature, we assign a score ('1') to a technique (1-9) if a standard (S_i) is obeyed. The final score is calculated as,

$$Score = \sum_{i=1}^{N} S_i/N \tag{15}$$

Where N is the total number of the standard rules for navigation, see Table III.

TABLE III. The Basic Five Standards of Navigation

Standards	Description
S_1	Keep user's hands free
S_2	Useable while sitting
S_3	Simple and inexpensive
S_4	Support unrestricted flying
S_5	Wireless/no connection

Including the proposed technique, a total of nine state-of-the-art gesture based navigation techniques were selected for comparison. Details of the techniques including *Speed control*, *Dynamic turning* and *Possible challenges* are presented in Table IV.

After extracting information from the research works about the five standards, final score was computed, see Table V.

TABLE V. Evaluation of the Techniques Based on the Standards

Technique No.	S_1	S_2	S_3	S_4	S_5	Score
1	0	1	0	1	0	**0.4**
2	1	1	0	1	0	**0.6**
3	1	0	1	1	0	**0.6**
4	1	1	0	1	0	**0.6**
5	1	1	1	1	0	**0.8**
6	1	1	1	1	1	**1**
7	1	1	0	0	1	**0.6**
8	0	1	1	0	1	**0.6**
9	1	1	1	1	1	**1**

TABLE IV. Details The Selected State-of-the-Art Navigation Techniques

Technique No.	Authors	Year	Method	Device / Sensor used	Title	Dynamic Speed control	Dynamic turning	Possible Challenge(s)
1	Linn, Andreas [23]	2017	Gaze based	HTC Vive with Tobii Eyetracker	Gaze Teleportation in Virtual Reality	Yes (based on thumb button)	Yes (arc turning)	Highly Sensitive to calibration error
2	Ferracani et al. [40]	2016	Hand's gesture based	HMD, Kinect and Leap motion controller	Locomotion by Natural Gestures for Immersive Virtual Environments	No	No	WIP and Swim needs continuous legs and arms movement which may lead to tiredness
3	Kammergruber et al. [41]	2012	Arm's gesture based	Microsoft Kinect	Navigation in virtual reality using Microsoft Kinect	No	Yes	Lighting condition and background may reduce accuracy
4	Stellmach et al. [42]	2012	Gaze based	Tobii T60 eye tracker	Designing Gaze-based User Interfaces for Steering in Virtual Environments	Yes	Yes	Continuous gazing at destination point is required which may lead to asthenopia
5	Vultur et al.[43]	2016	Gesture based (upper body)	Microsoft Kinect	Real-time Gestural Interface for Navigation in Virtual Environment	Yes	Yes	Imprecise segment length of motion may reduce accuracy
6	Muhammad et al.[36]	2018	Finger's gesture (colored marker based)	Ordinary Camera	VEN-3DVE: vision based egocentric navigation for 3D virtual environments	Yes	Yes	Lighting condition and similar background object may reduce accuracy
7	X. Tong et al.[44]	2016	Gesture based (upper body)	Microsoft Kinect	Exploring Embodied Experience of Flying in a Virtual Reality Game with Kinect	No	Yes	The outstretched arm's gesture may cause hurting and quick weariness
8	Muhammad et al. [45]	2015	Fingers' gesture (Fiducial marker based)	Ordinary Camera	Steps Via Fingers: A New Navigation Technique for 3D Virtual Environments	Yes	Yes (at a constant 45o)	Quick to and fro motion may lead to unavoidable occlusion
9	M. Raees and S. Ullah	2018	Eyes flickering based	Ordinary Camera	The proposed technique	Yes	Yes	Enough lighting is required to trace eyes

VI. Applicability in Virtual Reality Applications

An immersive VR system should offer exploration of 3D VE. Most of the techniques investigated in the literature require at least one hand [46] for navigation. Such systems restrict users to perform other 3D interactions with one hand. Keeping both hands free during navigation guarantees natural interactions [46]. Keeping in view this basic rule [38], goal behind this research was to keep user's hands free during navigation. By this way users would have the option to perform other interactions (selection, scaling and rotation) with any gesture. As the technique reserves no hand's gesture for navigation, hence can be easily mingled with any gesture based VR system. Furthermore, as no computation for pupil or iris analysis is performed, therefore the technique is suitable for current VR application where speed matters. Results of the subjective evaluation also suggest that the technique is suitable for VR application as 75% of the users picked the fact.

VII. Conclusion and Future Work

Navigation is often required in many 3D interactive virtual spaces. The emergence of virtual and augmented reality applications in different fields necessitates natural and simple way of navigation. With this contribution we propose a novel navigation technique which needs no extra device other than an ordinary camera. The intuitive eyes gestures for interaction not only ensure naturalism but also make it equally suitable for users with motor difficulties. Experimental results show that the proposed approach has reliable recognition and accuracy rates. Unlike other gesture based techniques [47], simple blinking of eyes are traced, hence the occlusion problems is minimized. Moreover, as the system remains indifferent to eyeglasses, therefore accuracy doesn't suffer from the size and color of eyeglasses. As the results suggest, errors due to false recognitions could possibly be reduced by using a high quality camera.

The proposed system is equally applicable in a wide spectrum of HCI including CAD, engineering, 3D gaming and simulation. In the emerging domain of Internet Of Things (IOT), the proposed system could be easily enhanced to interact with virtual object [48] and surgical navigation [49]. The work also covers the smooth integration of image processing and virtual environment which can lead to the design of more sophisticated virtual and augmented reality applications. This research work is a fraction of our aim of making possible interaction outside virtual reality labs. Although, we have succeeded for navigation, rotation and scaling are yet to be covered. In future, we are determined to enhance the system for rotation and scaling as well.

References

[1] Pfeiffer, Thies, M. E. Latoschik, and I. Wachsmuth, "Evaluation of binocular eye trackers and algorithms for 3D gaze interaction in virtual reality environments," *JVRB-Journal of Virtual Reality and Broadcasting,* vol. 5(16), (2008).

[2] Hidalgo, V. Larrumbe, L. Martin García, and M. T. Lorenzo, "Iris recognition using the JAVAVis Library," *International Journal of Interactive Multimedia and Artificial Intelligence,* 1.1: 43-48, (2008).

[3] L. Terziman, M. Marchal, M. Emily, F. Multon, B. Arnaldi, and A. L´ ecuyer, "Shake-Your-Head: Revisiting Walking-in-Place for Desktop Virtual Reality," Proc. ACM Symposium on Virtual Reality Software and Technology, pp. 27–34, (2010).

[4] D. S. Tan, G. G. Robertson, and M. Czerwinski, "Exploring 3D navigation: Combining speed-coupled flying with orbiting," In Proceedings of the ACMCHI 2001 Conference on Human Factors in Computing Systems, pp. 418–425, (2001).

[5] Malik, Shahzad, A. Ranjan, and R. Balakrishnan, "Interacting with large displays from a distance with vision-tracked multi-finger gestural input," In Proceedings of the 18th annual ACM symposium on User interface software and technology, pp. 43-52, (2005).

[6] Papadopoulos, Charilaos, D. Sugarman, and A. Kaufmant, "Nunav3d: A touch-less, body-driven interface for 3d navigation," In Virtual Reality Short Papers and Posters (VRW), pp. 67-68, (2012).

[7] Lee, Daeho, and S. Lee, "Vision-based finger action recognition by angle detection and contour analysis," *ETRI journal* 33, no. 3, pp. 415-422, (2011).

[8] Moerman, Clément, D. Marchal, and L. Grisoni, "Drag'n Go: Simple and fast navigation in virtual environment," In 3D User Interfaces (3DUI), IEEE Symposium, pp. 15-18, (2012).

[9] Tan, Desney S., George G. Robertson, and Mary Czerwinski. "Exploring 3D navigation: combining speed-coupled flying with orbiting." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 418-425. ACM, (2001).

[10] Fuhrmann, Anton, D. Schmalstieg, and M. Gervautz, "Strolling through cyberspace with your hands in your pockets: Head directed navigation in virtual environments," pp.216-225, (1998).

[11] Ajaj, Rami, F. Vernier, and C. Jacquemin, "Follow my finger navigation," *Human-Computer Interaction–INTERACT*, pp. 228-231.

[12] Mine, Mark, Frederick P. Brooks, and C. H. Sequin, "Moving objects in space: exploiting proprioception in virtual-environment interaction," In Proceedings of the 24th annual conference on Computer graphics and interactive techniques, pp.19-26, ACM Press/Addison-Wesley Publishing Co., (1997).

[13] Pierce, S. Jeffrey, Andrew, Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M.R. Mine, "Image plane interaction techniques in 3D immersive environments," In Proceedings of the 1997 symposium on Interactive 3D graphics, pp. 39-47, (1997).

[14] Fuhrmann, Anton, D. Schmalstieg, and M. Gervautz, "Strolling through Cyberspace with Your Hands in Your Pockets: Head Directed Navigation in Virtual Environments," In Virtual Environments Proceedings of the 4th EUROGRAPHICS Workshop on Virtual Environments, Springer-Verlag, pp.216-227, (1998).

[15] Davies, Char, and J.Harrison, "Towards Broadening the Aesthetics of Virtual Reality," *Compute Graphics* 30(4), pp. 25-28, (1996).

[16] LaViola, Joseph, D. A. Feliz, D. F. Keefe, and R. C. Zeleznik, "Hands-free multi-scale navigation in virtual environments," In Proceedings of the symposium on Interactive 3D graphics, pp. 9-15, (2001).

[17] Glueck, Michael, S. Anderson, and A. Khan, "DeskCube: using physical zones to select and control combinations of 3D navigation operations," In Proceedings of the 2010 Spring Simulation Multiconference, p. 200. Society for Computer Simulation International, (2010).

[18] Montañés, J.A. Puértolas, A.M. Rodríguez and I.S. Prieto, "Smart indoor positioning/location and navigation: A lightweight approach," *International Journal of Interactive Multimedia and Artificial Intelligence,* 2.2, pp. 43-50, (2013).

[19] Caputo, F. Marco, I. M. Ciortan, D. Corsi, M. De-Stefani, and A. Giachetti. "Gestural Interaction And Navigation Techniques For Virtual Museum Experiences," In *AVI CH*, pp.32-35, (2016).

[20] Kim, Ji-Sun, D. Gračanin, K. Matković, and F. Quek, "Finger Walking in Place (FWIP): A traveling technique in virtual environments," In International Symposium on Smart Graphics, pp. 58-69. Springer, Berlin, Heidelberg, (2008).

[21] Radkowski, Rafael, and C. Stritzke, "Interactive hand gesture-based assembly for augmented reality applications." The Fifth International Conference on Advances in Computer-Human Interactions, pp.303-308, (2012).

[22] Piumsomboon, Thammathip, G. Lee, R. W. Lindeman, and M. Billinghurst. "Exploring natural eye-gaze-based interaction for immersive virtual reality." Symposium on 3D User Interfaces (3DUI), pp. 36-39, (2017).

[23] L. Andreas, "Gaze Teleportation in Virtual Reality," (2017).

[24] Khamis, Mohamed, A. Hoesl, A. Klimczak, M. Reiss, F. Alt, and A. Bulling, "EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays." In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 155-166, (2017).

[25] Kaaman, Albert, and Kalle Bornemark. "Gaze-supported Interaction with Smart Objects through an Augmented Reality User Interface." (2017).

[26] Punpongsanon, Parinya, E. Guy, D. Iwai, K. Sato, and T. Boubekeur, "Extended LazyNav: Virtual 3D ground navigation for large displays and

head-mounted displays," IEEE transactions on visualization and computer graphics 23, no. 8, pp.1952-1963, (2017).
[27] Duguleană, Mihai, A. Nedelcu, and F. Bărbuceanu, "Measuring Eye Gaze Convergent Distance within Immersive Virtual Environments," Procedia Engineering 69, pp. 333-339, (2014).
[28] J. Karlén (2017, September 6). Eye tracking is virtual reality's next frontier. Available: https://venturebeat.com/2017/09/06/
[29] Padmanaban, Nitish, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein, "Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays," Proceedings of the National Academy of Sciences: 201617251, (2017).
[30] Dargham, J. Ahmad, A. Chekima and E.G. Moung, "Fusing facial features for face recognition," Distributed Computing and Artificial Intelligence. Springer, Berlin, Heidelberg, pp. 565-572, (2012).
[31] Magdin, Martin, and F. Prikler, "Real time facial expression recognition using webcam and sdk affective," International Journal of Interactive Multimedia and Artificial Intelligence, In Press (2017).
[32] Phung, S. Lam, A. Bouzerdoum, and D. Chai. "A novel skin color model in YCbCr color space and its application to human face detection," In Proceeding of International Conference on Image Processing, vol. 1, pp. I-I, (2002).
[33] M. Raees, S. Ullah, S. U. Rahman, and I. Rabbi, "Image based recognition of Pakistan sign language," Journal of Engineering Research 4, no. 1, pp. 21-41, (2016).
[34] Viola, Paul, and M. Jones, "Rapid object detection using a boosted cascade of simple features," In Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. I-I, (2001).
[35] Hansen, D. Witzner, and Qiang-Ji, "In the eye of the beholder: A survey of models for eyes and gaze," IEEE Tran. Patt. Analysis and Mach. Intelligence, 32 no. 3, pp.478-500, (2010).
[36] Raees, Muhammad, Sehat Ullah, and Sami Ur Rahman. "VEN-3DVE: vision based egocentric navigation for 3D virtual environments." International Journal on Interactive Design and Manufacturing (IJIDeM), pp.1-11, (2018).
[37] H. Sayers, "Desktop virtual environments: a study of navigation and age." Interacting with Computers, 16, no. 5, pp. 939-956, (2004).
[38] Zielasko, Daniel, S. Horn, S. Freitag, B. Weyers, and T. Kuhlen, "Evaluation of hands-free HMD-based navigation techniques for immersive data analysis." IEEE Symposium on 3D User Interfaces (3DUI), pp. 113-119, (2016).
[39] Tapu, Ruxandra, B. Mocanu, and E. Tapu, "A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments." 11th International Symposium on Electronics and Telecommunications (ISETC), pp. 1-4, (2014).
[40] Ferracani, Andrea, D. Pezzatini, J. Bianchini, G. Biscini, and A. D. Bimbo, "Locomotion by natural gestures for immersive virtual environments," In Proceedings of the 1st International Workshop on Multimedia Alternate Realities, pp. 21-24, (2016).
[41] Kammergruber, Florian, A. Ebner, and A. W. Günthner, "Navigation in virtual reality using Microsoft Kinect," In Proceedings of 12th International Conference on Construction Applications of Virtual Reality, (2012).
[42] Stellmach, Sophie, and R. Dachselt, "Designing gaze-based user interfaces for steering in virtual environments." In Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 131-138, (2012).
[43] Vultur, Oana-Mihaela, S. Pentiuc, and V. Lupu, "Real-time gestural interface for navigation in virtual environment," In International Conference on Development and Application Systems (*DAS*), pp. 303-307, (2016).
[44] Tong, Xin, A. Kitson, M. Salimi, D. Fracchia, D. Gromala, and B. E. Riecke, "Exploring embodied experience of flying in a virtual reality game with kinect," In IEEE International Workshop on Mixed Reality Art (*MRA*), pp. 5-6, (2016).
[45] M.Raees, S. ullah, and I. Rabbi, "Steps Via Fingers: A New Navigation Technique for 3D Virtual Environments," Mehran University Research Journal of Engineering & Technology, Pakistan, vol. 34, S1, pp.149-156, (2015).
[46] Hilsendeger, Andre, S. Brandauer, J. Tolksdorf, and C. Fröhlich, "Navigation in virtual reality with the wii balance board." In 6th Workshop on Virtual and Augmented Reality, (2009).
[47] T. Pfeiffer, Understanding Multimodal Deixis with Gaze and Gesture in Conversational Interfaces. PhD thesis, Bielefeld University, Bielefeld, Germany, (2010).
[48] Espada, J. Pascual, O.S. Martínez, B. Cristina, P.G. Bustelo and J. M. Lovelle, "Virtual objects on the internet of things," *International Journal of Interactive Multimedia and Artificial Intelligence,* 1.4: 23-29, (2011).
[49] Guiñales, Jorge, J.L. Carretero and M.B. García, "Virtual Planning and Intraoperative Navigation in Craniomaxillofacial Surgery," *International Journal of Interactive Multimedia and Artificial Intelligence,* 4.5: 65-67, (2017).

Muhammad Raees

Muhammad Raees is currently serving the department of Higher Education, KPK, Pakistan, as an Assistant Professor of Computer Science. He did M.Sc.(Computer Science) from Peshawar University, Pakistan in 2003. He completed M.Phil. with distinction in Virtual Reality Programming from University of Malakand in 2015. He is pursuing Ph.D. (Virtual Reality & Intelligent Systems) from University of Malakand. His research interests include Virtual Reality Programming, Computer Vision and Machine Learning.

Sehat Ullah

Sehat Ullah has been serving the Department of Computer Science and IT, University of Malakand Pakistan, since 2001. Currently he is Associate Professor of Computer Science and HOD of the Virtual Reality Department. He has Ph.D. degree in the field of VR and 3D Interactions from University of Évry, France. He has supervised more than eight Ph.D. scholars and has about 30 research publications to his credit.

Probabilistic Load Flow Solution Considering Optimal Allocation of SVC in Radial Distribution System

Walaa Ahmed[1], Ali Selim[1,2], Salah Kamel[1,3], Juan Yu[3], Francisco Jurado[2]*

[1] Department of Electrical Engineering, Faculty of Engineering, Aswan University, 81542 Aswan (Egypt)
[2] Department of Electrical Engineering, University of Jaén, 23700 EPS Linares, Jaén (Spain)
[3] State Key Laboratory of Power Transmission Equipment and System Security and New Technology, Chongqing University, Chongqing 400030 (China)

Received 26 July 2018 | Accepted 24 October 2018 | Published 5 November 2018

Abstract

This paper proposes a solution procedure for probabilistic load flow problem considering the optimal allocation of Static Var Compensator (SVC) in radial distribution systems. Pareto Envelope-based Selection Algorithm II (PESA-II) with fuzzy logic decision maker is developed to determine the optimal location and size of SVC based on the minimum total power losses and Voltage Deviation (VD). Combined cumulants and gram-chalier expansion are used for solving the probabilistic load flow problem. The proposed algorithm is tested on 33- bus and 69-bus distribution systems. The developed algorithm gives an acceptable solution with low number of iterations and less computation cost compared with the Monte Carlo method.

Keywords

Probabilistic Load Flow, SVC, Radial Distribution Systems, Multi-objective Particle Swarm Optimization.

DOI: 10.9781/ijimai.2018.11.001

I. Introduction

RADIAL structure of distribution systems along with the high ratio of current to voltage levels lead to an increase in the power loss and voltage deviation of distribution systems [1]. Therefore, real power loss and voltage deviation are considered one of the crucial problems in the restructured power system. Reactive power compensation devices would allow Static Var Compensator be the effective method to overcome these technical issues, whereby injecting a sufficient reactive power in suitable distribution system points which help in improving the voltage profile and decreasing the power loss.

Various VAR compensator types are utilized to enhance the steady-state transmittable power and control the voltage profile along the transmission line [2]. Capacitor banks, voltage regulators, Dynamic Voltage Restorer (DVR) and Distributed Flexible AC Transmission System devices (D-FACTS), are examples of such compensators. Recently, a reactive power compensation using the PV inverter is used in distribution systems to relax the voltage regulator [3]. However, D-FACTS devices are considered the best way to enhance voltage profile in power systems by increasing the capacity of the transmission line and controlling power flow in a flexible and fast way [4, 5]. There are different kinds of these devices such as the series compensator, shunt compensator, combined series–series compensator, and combined series–shunt compensator [6]. The Static Var Compensator (SVC) is considered a shunt FACTS device compensator. It has the ability to control the line power flow by injecting convenient reactive power into the system.

Decreasing the power losses can be achieved with integrated renewable energy resources. Recently, different methods have been presented using neural networks and machine learning to decrease the power and energy losses in the distribution system in existence of photovoltaic (PV) [7, 8]. In addtion, some methods have been presented to determine the power loss in smart grids [9, 10]. A steady-state security region-based chance-constrained model has been established to solve the power injection uncertainties of renewable energy resources [11].

Proper location and size of SVC decrease power losses and reduce Voltage Deviation (VD) to enhance the voltage profile. Many studies have addressed the optimal size and location of the SVC. Two different optimization methodologies have been investigated to solve the optimization problem, the first methodology tried to solve the optimization problem with a single objective function such as minimizing the total power loss while the second dealt with multi-objective function. However, both the total power loss and voltage deviation can be simultaneously minimized in distribution systems using Multi-Objective Optimization Problems (MOOP). A Pareto dominance concept can be used to classify the solutions of the MOOP as dominated or non-dominated solutions. An optimal location and setting of SVC using non-dominated sorting particle swarm optimization are introduced in [12]. Improvement by voltage profile using SVC in a distribution substation is presented in [13]

Many algorithms have been applied to solve MOOP such as; Pareto Archived Evolution Strategy (PAES) [14], Nondominated Sorting Genetic Algorithm (NSGA-II) [15], Strength Pareto Evolutionary Algorithm(SPEA) and improved version SPEA2 [16], Multi-objective Particle Swarm Optimization (MOPSO) [17], Pareto Envelope-based Selection Algorithm II (PESA-II)[18].

Nevertheless, all papers in the survey have solved the optimal SVC placement problem as a deterministic case neglecting load fluctuation [19]. The electric load can be affected by time and weather condition, however, there are random factors components depending on the consumers that cannot be modeled [20]. So, the deterministic load

* Corresponding author.
E-mail address: fjurado@ujaen.es

flow method does not satisfactory analyze the performance and the impact of SVC in the distribution system under load uncertainty. As a consequence, the method treats with specific values neglecting any uncertainties in the system, this can be done by using the Probabilistic Load Flow (PLF) [21]. PLF is presented in [22, 23] and further developed in [24, 25]. PLF can be solved by an analytical technique or Monte Carlo method [26]. Monte Carlo is a more accurate method, however, the analytical technique can give an acceptable solution with less computational cost compared to the Monte Carlo method [27]. There are different analytical methods to solve the PLF problem, like cumulants method [28, 29], or point estimate method [30]. These methods utilized the convolution properties to present the input and the output values of the systems as random variables [31, 32]. Moreover, the main feature of these methods is their computational efficiency in treating with random variables. A study of PLF by using a combined Cumulants and the Gram-Charlier expansion is presented in [31].

In this paper, a multi-objective function and probabilistic load flow are used to study the performance of including optimal SVC in the radial distribution system. Therefore, the topic discussed, and contribution of the work could be summarized as follows:

- Developing probabilistic load flow algorithm considering the optimal allocation of Static Var Compensator (SVC) in radial distribution systems;
- Pareto Envelope-based Selection Algorithm II (PESA-II) is used as a multi-objective optimization method to find the optimal location and size of SVC;
- Fuzzy logic decision maker is developed, and incorporated into PESA-II to find the best solution from the Pareto optimal set;
- Power loss and voltage deviation are used as multi-objective functions to be minimized.
- Combined cumulants and gram-chalier expansion are used for solving the probabilistic load flow problem;
- The developed algorithm is validated using 33- bus and 69-bus distribution systems;
- The number of iterations and computation time required for solving the probabilistic load flow problem are reduced compared with the Monte Carlo method.

The paper is organized as follows: Section II introduces the mathematical model of the probabilistic load flow. Section III describes the optimization process using PESA-II. Section IV gives the numerical results. Finally, the conclusions are presented in Section V.

II. Probabilistic Load Flow Analysis

The load flow is represented by a system of non-linear equations that reflect the balance at steady state in the network between the power consumed and the power produced, Generally speaking, probabilistic load flow calculations consist of two parts linearization of load flow equations and convolution calculations. The cumulant method is used to perform convolution computation of random variables and the Gram–Charlier series expansion, to compute their distributions.

$$P_i = V_i \sum_{k=1}^{n} V_k \left(G_{ik} \cos \theta_{ik} + B_{ik} \sin \theta_{ik}\right) \quad (1)$$

$$Q_i = V_i \sum_{k=1}^{n} V_k \left(G_{ik} \sin \theta_{ik} - B_{ik} \cos \theta_{ik}\right) \quad (2)$$

The active and reactive power P, Q at bus i are calculated using the voltage V_i and the values of the conductance G_{ik}, susceptance B_{ik}, and the angle θ_{ik} between the bus i and the connected bus k for all n buses of the system.

A. Linear Approximation

The linearization of load flow equations is performed around the solution obtained with a deterministic load flow, based on the expected values of the system. These expected values are obtained after solving the problem of the deterministic method of the Newton-Raphson load flow calculation. To illustrate this technique, two random variables X and Y are considered. At some point in the problem, these random variables are multiplied to give a third random variable Z.

$$Z = X.Y \quad (3)$$

If the deviations of X and Y are represented around their mean values $\bar{X}$ and $\bar{Y}$ by ΔX and ΔY, respectively, the following can be assumed.

$$X \cong \bar{X} + \Delta X \text{ . and } Y \cong \bar{Y} + \Delta Y \quad (4)$$

It is obtained, after neglecting the terms of the second order $\Delta X \quad \Delta Y$

$$Z \cong \overline{XY} + \bar{X}\Delta Y + \bar{Y}\Delta X = \bar{X}Y + \bar{Y}X + \overline{XY} \quad (5)$$

Therefore, if changes of random variables are small, the variable Z can be linearized since the expected values for X and Y are known. This technique can be applied to the angles and voltages in (6) of the load flow. Thus, the

$$P_i = V_i \sum_{k=1}^{n} (e'_{in} + f'_{in}\delta_i - f'_{in}\delta_n + g'_{in}V_i + h'_{in}V_n) \quad (6)$$

$$q_i = V_i \sum_{k=1}^{n} (e''_{in} + f''_{in}\delta_i - f''_{in}\delta_n + g''_{in}V_i + h''_{in}V_n) \quad (7)$$

Where e', f', g', h', e'', f'', g'' and h'' are calculated from system parameters and expected values for the variables.

B. Convolution Calculation

Suppose that there are two independent random variables X and Y, and their probability density functions $f_1(x)$ and $f_2(y)$, respectively, then $Z = X + Y$ is still a random variable. The probability density function of Z is:

$$\phi(Z) = \int_{-\infty}^{\infty} f(x, z-x)dx = \int_{-\infty}^{\infty} f_1(x) f_2(z-x)dx \quad (8)$$

Its distribution function is

$$F(Z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_1(x) f_2(z-x)dx\, dz \quad (9)$$

C. Moments and Cumulants

The convolution random variables can be replaced by the sum of their cumulants. The cumulants and moments of a random variable are the set of constants that reveal the properties of X and determine its distribution function [27]. However, cumulants have a number of properties that make their manipulation more useful. The cumulant method has low computational cost [28, 30]. It is also a flexible method that uses any random variable and not just normal distributions. Most statistical calculations using cumulants method are simpler than the corresponding calculation using moments.

When a random variable distribution is known, its moment of every order can then be obtained. Suppose the density function of a continuous random variable x is g(x), then its υ-order moment α_υ can be calculated by the following equation:

$$\alpha_\upsilon = \int_{-\infty}^{\infty} x^\upsilon \, g(x)dx \quad (10)$$

When υ= 1, the expectation of the random variable x,

$$\mu = \alpha_1 = \int_{-\infty}^{\infty} x \, g(x)dx \quad (11)$$

From the expectation μ, the central moment of every order M_υ

can be calculated. Then the central moment of every order M_υ can be solved by the expectation,

$$M_\upsilon = \int_{-\infty}^{\infty}(x-\mu)^\upsilon g(x)dx \tag{12}$$

The relationships between the cumulants and the moments of every order are given in [28, 30].

D. Gram–Charlier Expansion

The Gram–Charlier series expansion is mainly used in probabilistic production simulation [28]. These series represent the random variable distribution function by using the derivatives of the random variable. The coefficients of the series are formed by the random variable moments.

$$F(x) = \Phi(x) + \frac{c_1}{1!}\Phi'(x) + \frac{c_2}{2!}\Phi''(x) + \frac{c_3}{3!}\Phi^{(3)}(x) + \cdots \tag{13}$$

$$f(x) = \varphi(x) + \frac{c_1}{1!}\varphi'(x) + \frac{c_2}{2!}\varphi''(x) + \frac{c_3}{3!}\varphi^{(3)}(x) + \cdots \tag{14}$$

Where $\Phi(x)$ and $\varphi(x)$ represent the cumulative distribution function (CDF) and probabilistic density function (PDF) of a normal distribution with m=0 and and σ=1; c_υ are constant coefficients.

$$c_\upsilon = (-1)^\upsilon \int_{-\infty}^{\infty} H_\upsilon(x)\, f(x)\, dx \quad , \upsilon = 1,2,3,\ldots \tag{15}$$

Where $H_\upsilon(x)$ it is the Hermite polynomial of order υ. In practice coefficients Gram-expansion Charlier c_υ can be expressed in terms of moments central random variable expansion object. The first seven coefficients are

$$c_0 = 1, c_1 = c_2 = 0,\ c_3 = -\frac{\mu_3}{\sigma^3}, c_4 = \frac{\mu_4}{\sigma^4} - 3, c_5 = -\frac{\mu_5}{\sigma^5} + 10\frac{\mu_3}{\sigma^3},$$
$$c_6 = \frac{\mu_6}{\sigma^6} - 15\frac{\mu_4}{\sigma^4} + 30, c_7 = -\frac{\mu_7}{\sigma^7} - 21\frac{\mu_5}{\sigma^5} + 105\frac{\mu_3}{\sigma^3}$$

III. Optimization Model

This section presents the optimization methodologies which are used in this paper to determine the optimal size and location. The optimization model is based on PESA-II algorithm, two optimization approaches are presented in the following subsection:

A. Objective Functions and Constraints

The two objectives functions that should be minimized are P_{loss} and Voltage Deviation (VD)

$$F = \min\ (P_{loss}\ , VD) \tag{16}$$

Where the power loss in each branch z connected between two buses i and k is calculated as:

$$P_{loss\_z} = \mathrm{R}_{ik}\mathrm{I}_z{}^2 \tag{17}$$

The total power losses for all branches n_br in the distribution system can be expressed as follows:

$$P_{loss} = \sum_{z=1}^{n\_br} P_{loss\_z} \tag{18}$$

The voltage deviation at bus i in terms of the specified voltage V_{sep} can be calculated as follow:

$$VD_i = \left|\frac{V_{sep} - V_i}{V_{sep}}\right| \tag{19}$$

and

$$VD = \max(VD_i) \tag{20}$$

B. PESA-II Algorithm

PESA-II algorithm can be concluded in the following steps [18]:

Step 1: Initialize number of population and two population-based parameters, the size of the Internal Population (IP) and the maximum size of the archive, or External Population (EP)

Step 2: Evaluate the objective functions for each individual P_i chromosome in the population.

Step 3: Find the nondominated members and incorporate the non-dominated members of IP into EP.

Step 4: If a termination criterion has been reached, then stop, returning the set of chromosomes in EP as the result. Otherwise, delete the current contents of IP, and repeat the following until P_i new candidate solutions have been generated:

- With probability P_c, select two parents from EP, produce a single child via crossover, and mutate the child. With probability $(1 - P_c)$, select one parent and mutate it to produce a child.

Step 5: increment the iteration number, endif iteration number is greater than max iter, otherwise go to step 2.

C. Fuzzy Decision-Making

As shown in Fig. 1, a Pareto optimal set is determined using PESA-II, however, the main concern of the decision-maker is to find the best compromise solution from the Pareto set. A fuzzy decision-making approach is used in this paper to find the optimal Pareto point according to the decision-maker preferences. Fuzzy logic is considered a range decision in which the output is obtained by fuzzification of inputs and outputs, from the associated membership functions (MFs). Hence, the value of the objective function F of individual n can be normalized as follows:

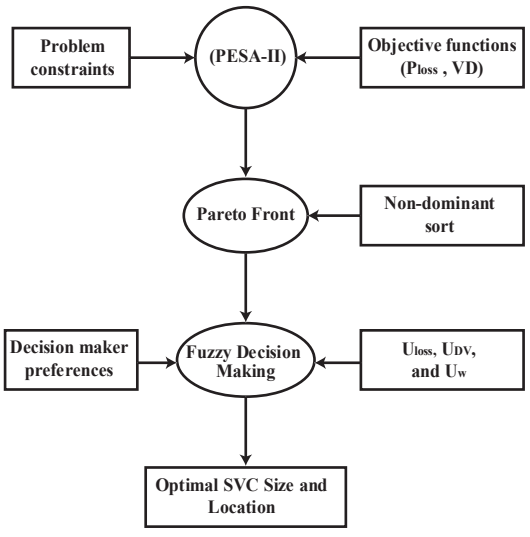


Fig.1. PESA-II with the fuzzy decision-making process.

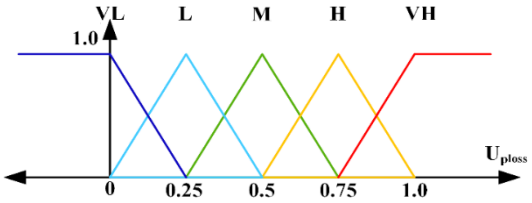
$$u_i^n = \begin{cases} 1 & F_i \le F_i^{min} \\ \dfrac{F_i^{max} - F_i}{F_i^{max} - F_i^{min}} & F_i^{max} \le F_i \le F_i^{min} \\ 0 & F_i \ge F_i^{max} \end{cases} \quad (21)$$

where, F_i^{min} and F_i^{max} are the minimum and maximum value of the ith objective function among all non-dominated solutions respectively.

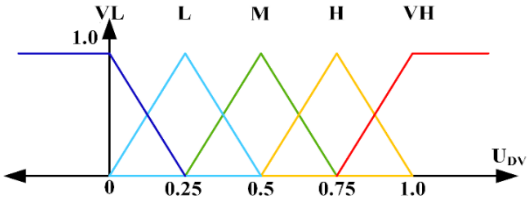
All normalized values are fuzzified using five triangular MFs for the two inputs power loss U_{loss}, voltage deviation U_{VD}, and the output weighting of the Pareto solution U_w, as shown in Fig. 2, the output U_w , can be calculated using the rules as shown Fig. 3 and given in Table I. After applying the fuzzy decision making on the Pareto set, the output U_w is a weighting factor between [0 ~ 1] for each solution in the Pareto set according to the fuzzy rules, hence the best solution will be the one that has the maximum weighting value.

TABLE I. RULES OF THE FUZZY CONTROLLER (VL =VERY LOW, L= LOW, M= MEDIUM, H= HIGH, AND VH= VERY HIGH)

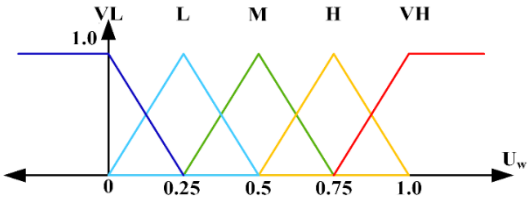
		U_{VD}				
		VL	**L**	**M**	**H**	**VH**
U_{Ploss}	**VL**	VL	VL	VL	VL	VL
	L	VL	L	L	L	L
	M	VL	L	M	M	M
	H	VL	L	M	H	H
	VH	VL	L	M	H	VH



a. Normalized power loss membership plot.



b. Normalized voltage deviation memberships plot.



c. Normalized weighting membership plot.

Fig. 2. Fuzzy membership plot for normalized objective functions values.

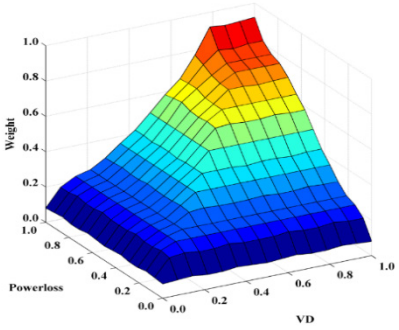


Fig. 3. Fuzzy rules.

IV. RESULTS

In this section, the obtained results of two standard test systems are presented and comprehensively discussed.

A. 33-Bus Radial System

The proposed approach is tested on the 33-bus radial distribution system. The single line diagram is displayed in Fig. 4. Statistical data of loads at all PQ buses are given in [33] and modeled as normal variables. The results of mean and standard deviation of voltages at PQ buss in the system without including SVC are given in Table II. The results in Table II show that the voltage obtained by the analytical technique is virtually identical to the results obtained with MCS. Also, it can be seen in Fig. 5 that CDFs for analytical technique and MCS at bus 18 and bus 33 are virtually identical.

There are two cases of connected SVC based on the SVC numbers: case 1 for one SVC and case 2 for two SVC. The optimal size and location for the two cases obtained by different multi-objective optimization algorithms are summarized in Table III and Table IV. It is obvious from the two tables that PESA-II gives the better solution for the optimal size and location for SVC compared to the other optimization algorithms.

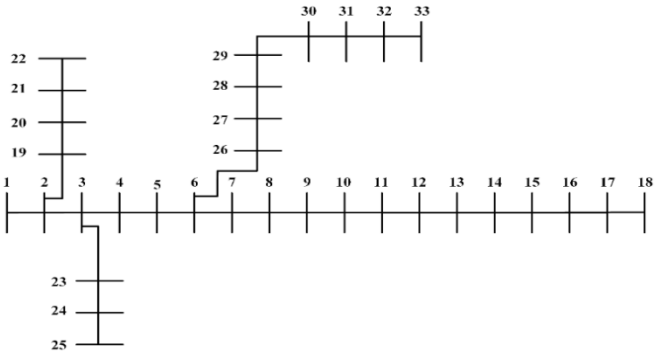
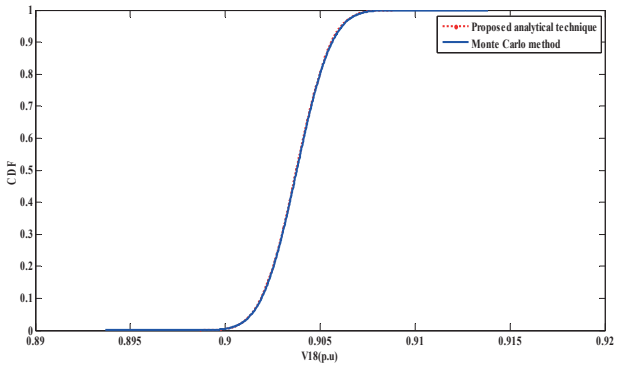


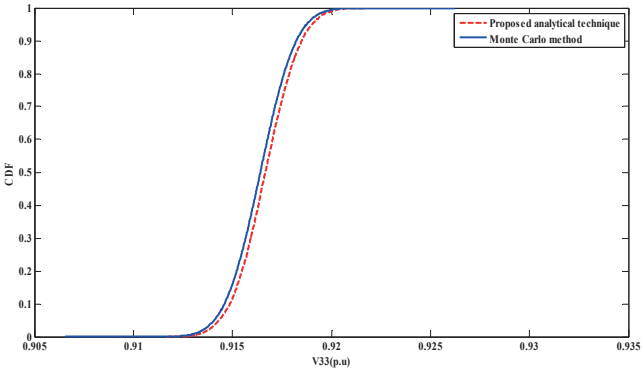
Fig.4. 33 bus single line diagram.

TABLE II.
VOLTAGES IN 33-BUS DISTRIBUTION SYSTEM WITHOUT INCLUDING SVC

Bus	Analytical technique		Monte Carlo	
	μ	σ	μ	Σ
2	0.997031	3.74E-05	0.997027	0.000038
3	0.982922	0.000232	0.982894	0.000236
4	0.975425	0.000318	0.975381	0.000326
5	0.968013	0.000416	0.967953	0.000426
6	0.949573	0.00068	0.949471	0.000698
7	0.946056	0.00071	0.945945	0.000729
8	0.932426	0.000886	0.932283	0.000906
9	0.926108	0.000985	0.92595	0.001006
10	0.920248	0.001089	0.920074	0.001111
11	0.919381	0.001105	0.919205	0.001129
12	0.91787	0.001136	0.91769	0.001160
13	0.91171	0.001265	0.911515	0.001290
14	0.909425	0.001318	0.909225	0.001344
15	0.908002	0.001342	0.907799	0.001368
16	0.906624	0.001371	0.906416	0.001396
17	0.904581	0.001402	0.904368	0.001429
18	0.903969	0.001413	0.903754	0.001440
19	0.996503	4.28E-05	0.996498	0.000043
20	0.992926	0.000145	0.992921	0.000143
21	0.992221	0.000173	0.992217	0.000172
22	0.991584	0.000211	0.991579	0.000210
23	0.979338	0.000326	0.979309	0.000331
24	0.972671	0.000547	0.972639	0.000553
25	0.969348	0.000655	0.969315	0.000664
26	0.947647	0.000715	0.94754	0.000734
27	0.945087	0.000765	0.944975	0.000786
28	0.933669	0.00102	0.93353	0.001048
29	0.925466	0.001218	0.925307	0.001251
30	0.921915	0.001304	0.921747	0.001339
31	0.917762	0.001366	0.917585	0.001403
32	0.916849	0.001381	0.916669	0.001418
33	0.916566	0.001383	0.916385	0.001421
	Base MVA=100		Base KV=12.66	



a. CDF of the voltage at bus 18.



b. CDF of the voltage at bus 33.

Fig. 5. CDF of the voltage at bus 18 and bus 33 without SVC.

TABLE III. OPTIMAL SIZE AND LOCATION OF SINGLE SVC USING DIFFERENT OPTIMIZATION ALGORITHMS FOR 33 BUS SYSTEM

Method	SVC Location	SVC Size (kVAR)	Power losses (kW)	VD (p.u)
MOPSO	7	2137.70	158.77	0.0632
NSGA II	7	2146.46	158.89	0.0631
PESA-II	7	1896.35	**156.38**	**0.0653**
SPEA2	7	2400.69	163.38	0.0608

TABLE IV. OPTIMAL SIZE AND LOCATION OF TWO SVC USING DIFFERENT OPTIMIZATION ALGORITHMS FOR 33 BUS SYSTEM

Method	SVC Location	SVC Size (kVAR)	Power losses (kW)	VD (p.u)
MOPSO	14	1071.60	194.14	0.0307
	30	1847.60		
NSGA II	14	1041.57	191.77	0.0308
	30	1851.95		
PESA-II	13	1225.18	**184.57**	**0.0367**
	30	1509.74		
SPEA2	14	1053.00	192.97	0.0307
	30	1854.70		

Fig. 6 shows the voltage profile of 33 bus with SVC control device in the mentioned two cases. Fig. 7 and Fig. 8 gives the Pareto Optimal Front for 33-bus for single SVC and two SVCs respectively. Fig. 9 gives PDF and CDF of voltages at bus 18 with and without SVC. Fig. 10 shows PDF and CDF of voltages at bus 33. Expected voltage profile of 33 bus system with single SVC in 24 hours is shown in Fig. 11, however Fig. 12 gives the SD of it in 24 hours. Expected voltage profile of 33 bus system with two SVC in 24 hours is shown in Fig. 13, however, Fig. 14 gives the SD of it in 24 hours. Table V gives a comparison between expected SD values with/without SVC, the expected value of Vmin and Vmax are improved in both cases. Also, Qlosses is decreased in case 1 and case 2 compared to the base case, however, the standard deviation is not improved in the two cases compared to without SVC. Computational times of two cases are presented in Table VI.

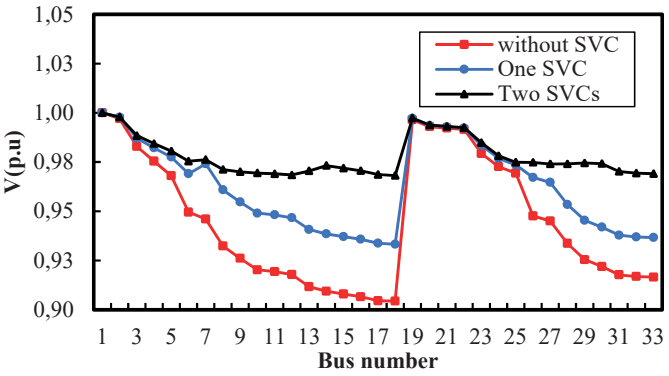


Fig. 6. Voltage profile of 33 bus with control device in different cases.

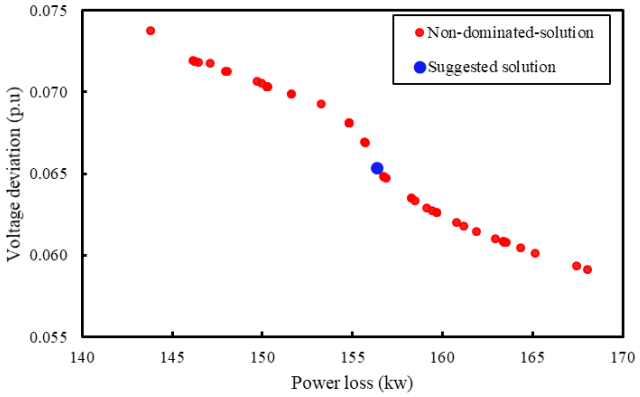


Fig. 7. Pareto Optimal Front for 33 bus with single SVC.

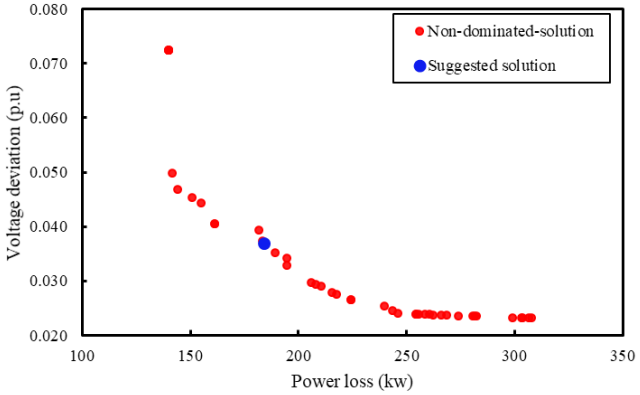
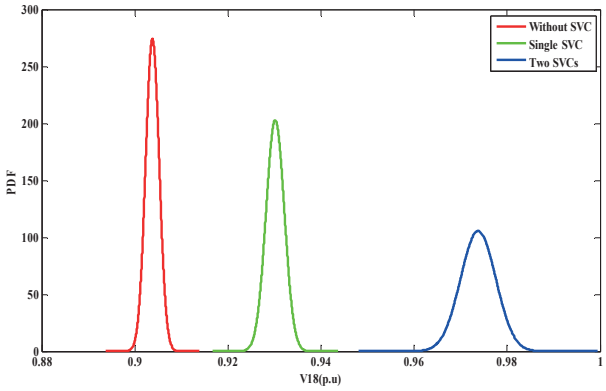
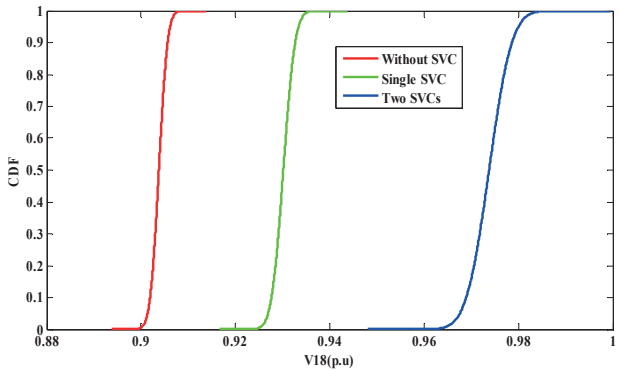


Fig. 8. Pareto Optimal Front for 33 bus with two SVC.

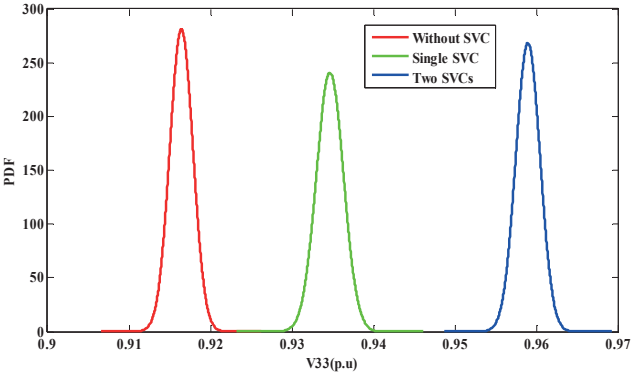


a. PDF

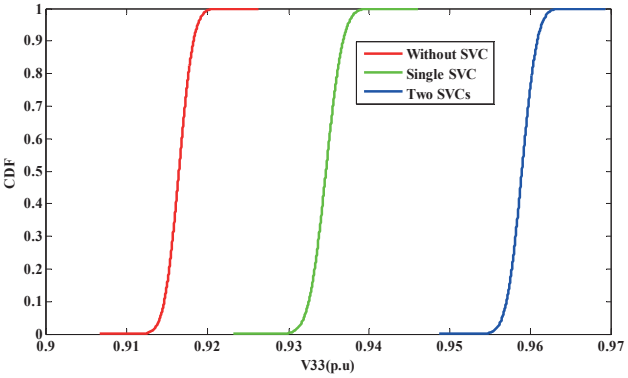


b. CDF

Fig. 9. PDF and CDF of bus 18.



a. PDF



b. CDF

Fig.10. PDF and CDF of 33 bus.

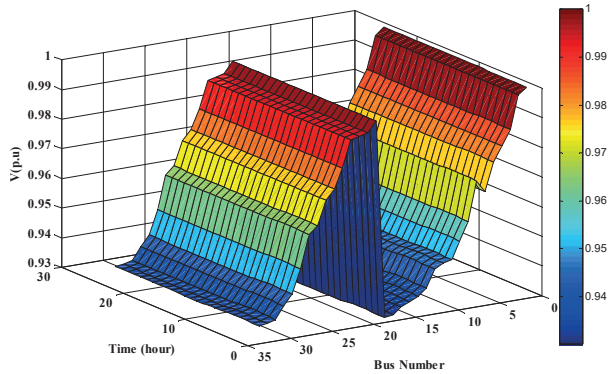


Fig. 11. Expected voltage profile of 33 bus system with single SVC.

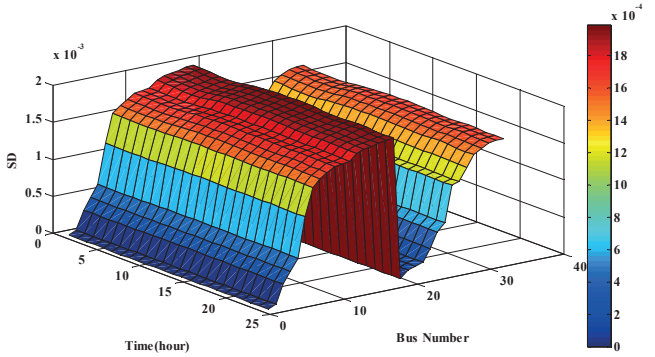


Fig. 12. Standard Deviation of 33 bus system with single SVC.

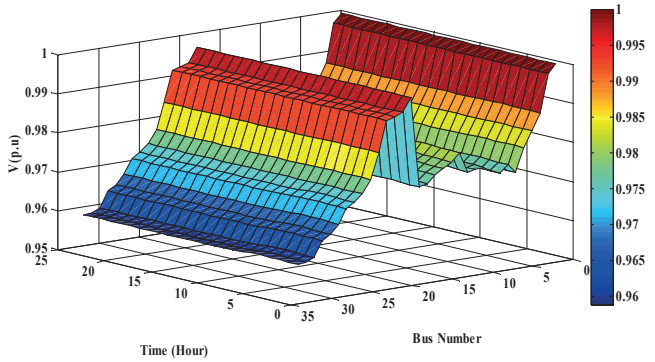


Fig.13. Expected voltage profile of 33 bus system with two SVC.

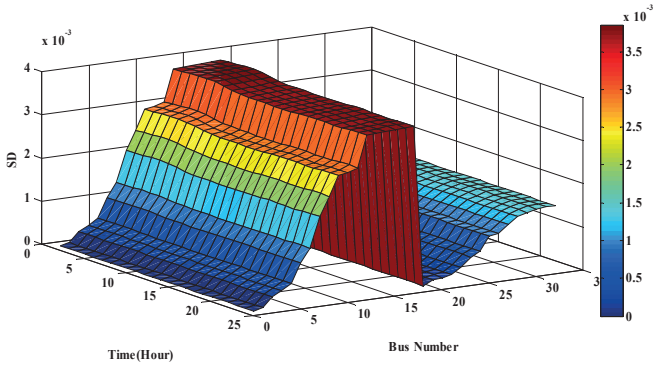


Fig. 14. Standard Deviation of 33 bus system with two SVC.

TABLE V. Stochastic Information of the System With/Without SVC for 33 Bus System

	Without SVC		One SVC		Two SVCs	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Vmax	0.9970	0.00005	0.9976	0.00004	0.9978	0.00004
Vmin	0.9131	0.00243	0.9307	0.00196	0.9588	0.00377
Ploss (kW)	202.67	0.00218	156.38	0.00205	184.57	0.00208
Qloss (kVAR)	135.14	0.00557	122.82	0.00457	112.12	0.00550
VD	0.0869	0.00025	0.0653	0.00023	0.0367	0.00019

TABLE VI. Computation Times

	Analytical technique (s)	MCS (s)
Without SVC	15.0	1611.8
With one SVC	69.08	10 516
With two SVCs	73.93	11 978

B. 69-Bus Test System

The proposed approach is also tested on 69-bus radial distribution system shown in Fig. 15 [34]. CDF of the voltage at bus 27 and bus 65 without SVC are shown in Fig. 16. Table VII and Table VIII show the values of optimal location and size of SVC in 69-bus system with different multi-objective optimization algorithms. From these tables the PESA-II still gives the better solution as in the previous test system. Fig. 17 gives PDF and CDF of voltages at bus 27 bus with and without SVC. Fig. 18 shows PDF and CDF of voltages at bus 65. Fig. 19 shows voltage profile of 69 bus in different cases. Pareto Optimal Front for 69 bus is shown in Fig. 20 and Fig. 21. Expected voltage profile of 33 bus system with single SVC in 24 hours is shown in Fig. 22; however, Fig. 23 gives the SD of it in 24 hours. Expected voltage profile of 69 bus system with two SVC in 24 hours is shown in Fig. 24, however, Fig. 25 gives the SD of it in 24 hours.

Table IX gives a comparison between expected and SD values with/without SVC, the expected value of V_{min} and V_{max} are improved in both cases also Q_{losses} is decreased in case 2 than case 1, however, the standard deviation is not improved in two cases compared to without SVC. Table X gives computation time of two cases.

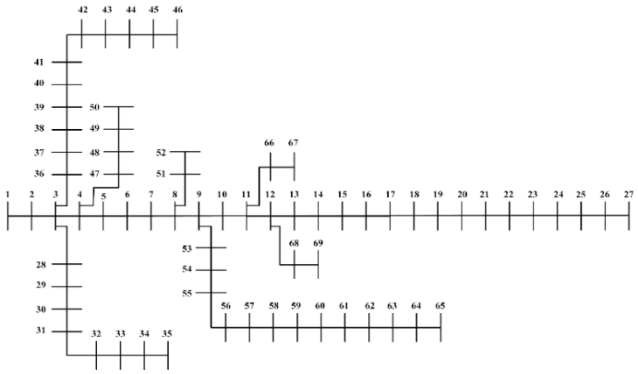


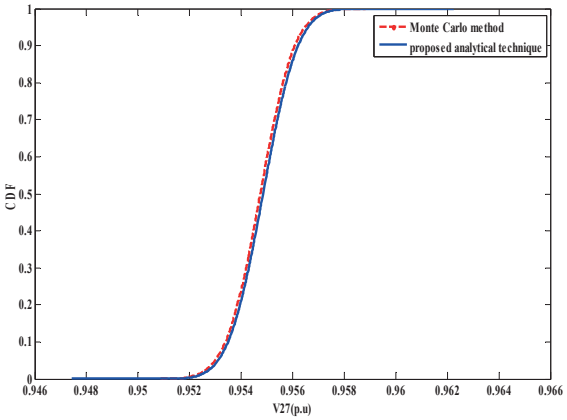
Fig. 15. 69 bus single line diagram.

TABLE VII. Optimal Size and Location of Single SVC Using Different Optimization Algorithms for 69 Bus System

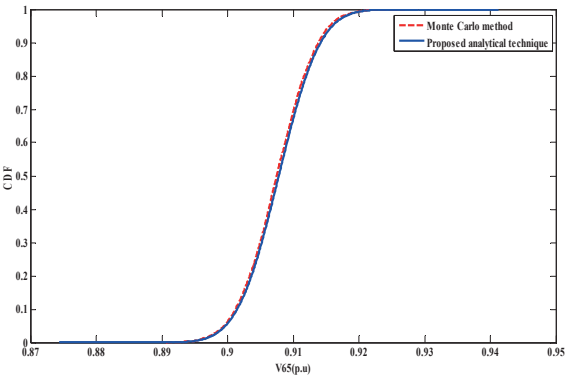
Method	SVC Location	SVC Size (kVAR)	Power losses (kW)	VD (p.u)
MOPSO	63	3055.60	278.17	0.0444
NSGA II	63	2306.92	196.64	0.0541
PESA-II	60	2245.33	**184.11**	**0.0603**
SPEA2	62	2370.93	196.80	0.0542

TABLE VIII. Optimal Size and Location of Two SVC Using Different Optimization Algorithms for 69 Bus System

Method	SVC Location	SVC Size (kVAR)	Power losses (kW)	VD (p.u)
MOPSO	55	2567.40	302.69	0.0398
	63	2387.40		
NSGA II	56	1627.94	297.87	0.0397
	63	2651.04		
PESA-II	56	1394.32	**265.40**	**0.0429**
	63	2508.19		
SPEA2	56	1791.39	298.17	0.0397
	63	2581.24		

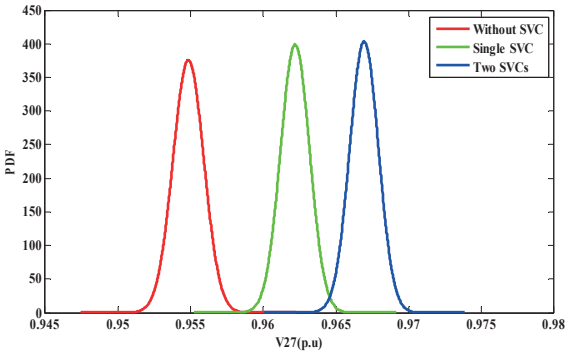


a. CDF of voltage at bus 27.

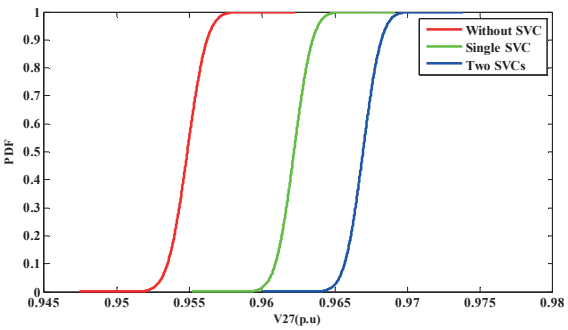


b. CDF of voltage at bus 65.

Fig. 16. CDF of voltage at bus 27 and bus 65 without SVC.

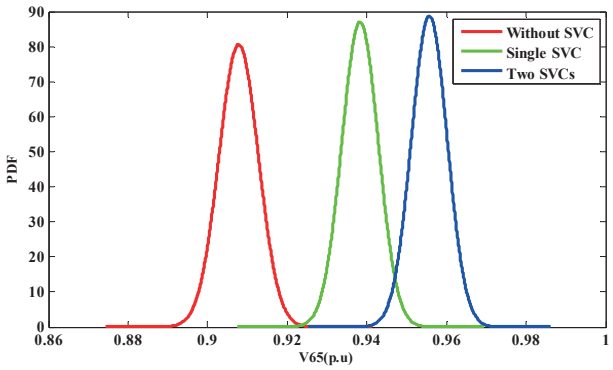


a. PDF

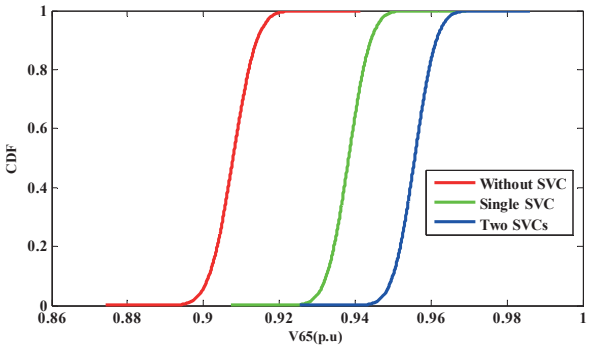


b. CDF

Fig. 17. PDF and CDF of without/with implementation SVC at bus 27.



a. PDF



b. CDF

Fig. 18. PDF and CDF of without/with implementation SVC at bus 65.

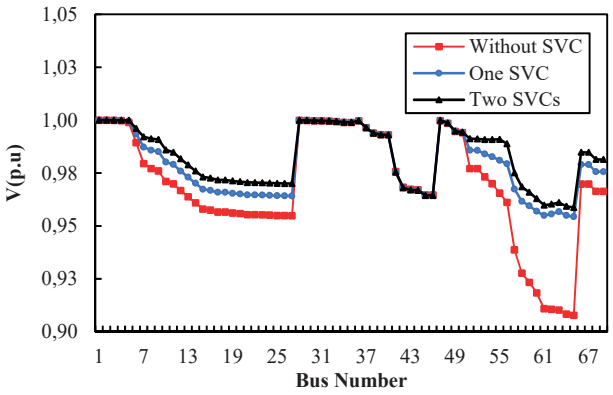


Fig. 19. Voltage profile of 69 bus with control device in different cases.

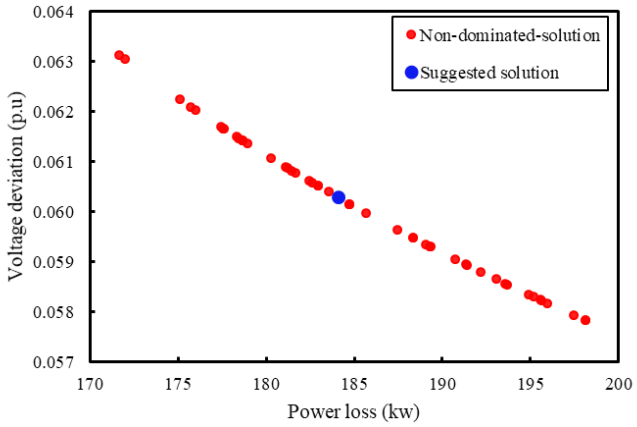


Fig. 20. Pareto Optimal Front for 69 bus with single SVC.

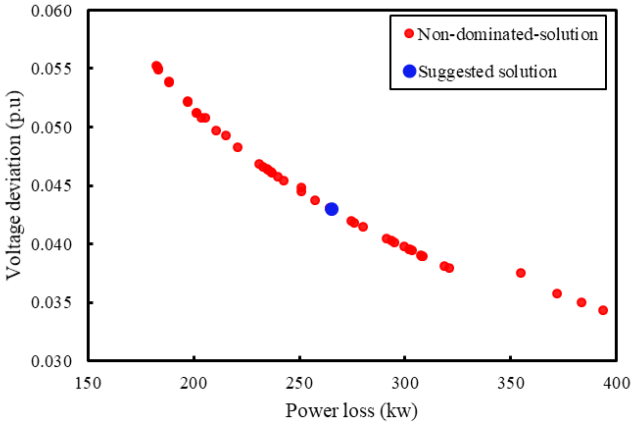


Fig. 21. Pareto Optimal Front for 69 bus with two SVC.

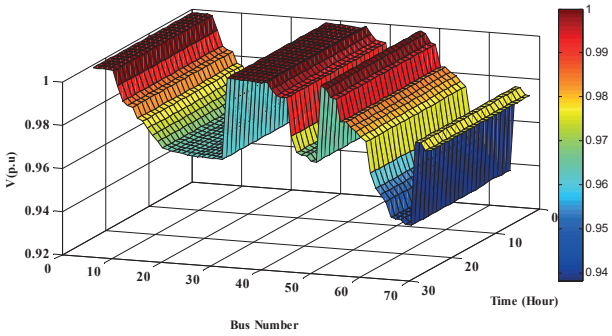


Fig. 22. Expected voltage profile of 69 bus system with single SVC.

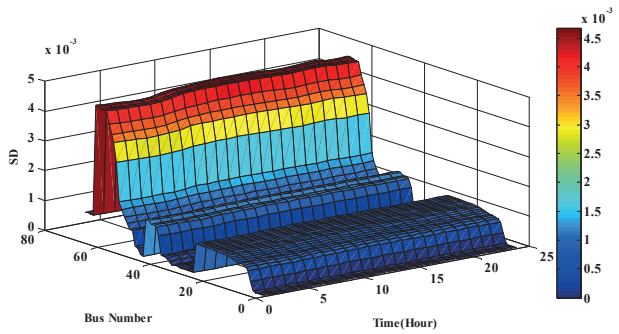


Fig. 23. Standard Deviation of 69 bus system with single SVC.

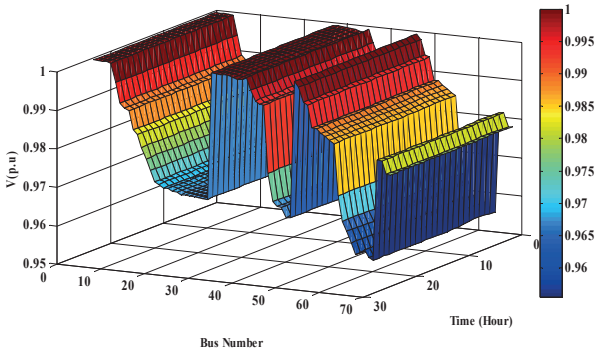


Fig. 24. Expected voltage profile of 69 bus system with two SVC.

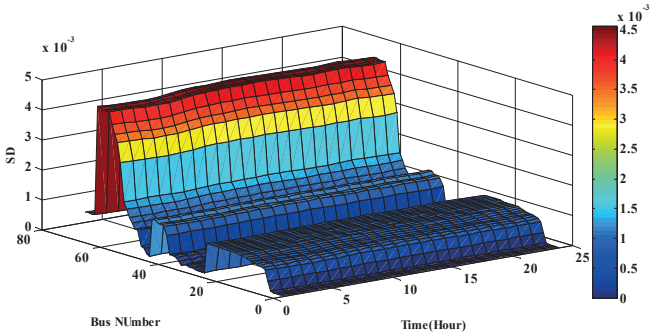


Fig. 25. Standard Deviation of 69 bus system with two SVC.

TABLE IX. Stochastic Information of the System With/Without SVC for 69 Bus System

	Without SVC		One SVC		Two SVCs	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Vmax	0.9999	1.13E-06	0.9999	1.13E-06	1.0001	1.12E-06
Vmin	0.9092	0.00124	0.9622	0.00109	0.9557	0.0045
Ploss (kW)	224.949	0.008723	184.11	0.007723	265.40	0.007686
Qloss (kVAR)	102.145	0.007653	129.582	0.007379	145.35	0.010197
VD	0.0908	0.00123	0.0603	0.000988	0.0429	0.00458

TABLE X. Computation Times

	Analytical technique (s)	MCS (s)
Without SVC	32.012	3611.8
With one SVC	80.57	12 709
With two SVCs	85.23	12 845

V. Conclusion

In this paper, the probabilistic load flow problem considering the optimal location and size of SVC in radial distribution system has been solved. Pareto Envelope-based Selection Algorithm II (PESA-II) has been utilized to achieve the target of this paper and minimize the total power loss and voltage deviation based on the optimal location and size of SVC. Combined cumulants and gram-chalier expansion have been used for solving probabilistic load flow problem. The proposed methodology has been validated using the standard 33- bus and 69-bus distribution systems. The results give an acceptable solution with a low number of iterations and less computation cost compared to the Monte Carlo method (MCS).

References

[1] M. Rahmani-andebili, "Reliability and economic-driven switchable capacitor placement in distribution network," *IET Generation, Transmission & Distribution,* vol. 9, pp. 1572-1579, 2015.

[2] E. Shahryari, H. Shayeghi, and M. Moradzadeh, "Probabilistic and Multi-Objective Placement of D-STATCOM in Distribution Systems Considering Load Uncertainty," *Electric Power Components and Systems,* vol. 46, pp. 27-42, 2018.

[3] A. Selim, M. Abdel-Akher, M. M. Aly, S. Kamel, and T. Senjyu, "Fast quasi-static time-series analysis and reactive power control of unbalanced distribution systems," *International Transactions on Electrical Energy Systems,* vol. 0, p. e2673, 2018.

[4] M. M. Farsangi, H. Nezamabadi-pour, Y.-H. Song, and K. Y. Lee, "Placement of SVCs and selection of stabilizing signals in power systems," *IEEE Transactions on Power Systems,* vol. 22, pp. 1061-1071, 2007.

[5] M. Saravanan, S. M. R. Slochanal, P. Venkatesh, and J. P. S. Abraham, "Application of particle swarm optimization technique for optimal location of FACTS devices considering cost of installation and system loadability," *Electric Power Systems Research,* vol. 77, pp. 276-283, 2007.

[6] S. Kamel, F. Jurado, and D. Vera, "A simple implementation of power mismatch STATCOM model into current injection Newton–Raphson power-flow method," *Electrical Engineering,* vol. 96, pp. 135-144, 2014.
[7] M. Abdel-Nasser, K. Mahmoud, and H. Kashef, "A Novel Smart Grid State Estimation Method Based on Neural Networks," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, 2018.
[8] K. Mahmoud and M. Abdel-Nasser, "Fast-yet-Accurate Energy Loss Assessment Approach for Analyzing/Sizing PV in Distribution Systems using Machine Learning," *IEEE Transactions on Sustainable Energy,* 2018.
[9] A.G.Azar, M.Afsharchi, M.Davoodi, and B.S. Bigham, "A multi-objective market-driven framework for power matching in the smart grid," Engineering Applications of Artificial Intelligence, 70, pp.199-215, 2018.
[10] E. Dall'Anese, and A. Simonetto, "Optimal power flow pursuit," IEEE Transactions on Smart Grid, 9(2), pp.942-952, 2018.
[11] T. Yang and Y. Yu, "Steady-state Security Region-based Voltage/Var Optimization Considering Power Injection Uncertainties in Distribution Grids," *IEEE Transactions on Smart Grid,* 2018.
[12] R. Benabid, M. Boudour, and M. Abido, "Optimal location and setting of SVC and TCSC devices using non-dominated sorting particle swarm optimization," *Electric Power Systems Research,* vol. 79, pp. 1668-1677, 2009.
[13] G. Kour, G. Brar, and J. Dhiman, "Improvement by voltage profile by static Var compensators in distribution substation," *International Journal of Instrumentation Science,* vol. 1, pp. 21-24, 2012.
[14] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using the Pareto archived evolution strategy," *Evolutionary computation,* vol. 8, pp. 149-172, 2000.
[15] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation,* vol. 6, pp. 182-197, 2002.
[16] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," *TIK-report,* vol. 103, 2001.
[17] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Transactions on evolutionary computation,* vol. 8, pp. 256-279, 2004.
[18] D. W. Corne, J. D. Knowles, and M. J. Oates, "The Pareto envelope-based selection algorithm for multiobjective optimization," in *International conference on parallel problem solving from nature*, 2000, pp. 839-848.
[19] G. T. Temur, "A novel multi attribute decision making approach for location decision under high uncertainty," *Applied Soft Computing,* vol. 40, pp. 674-682, 2016.
[20] F. Ruiz-Rodriguez, J. Hernandez, and F. Jurado, "Probabilistic load flow for photovoltaic distributed generation using the Cornish–Fisher expansion," *Electric Power Systems Research,* vol. 89, pp. 129-138, 2012.
[21] X.-F. Wang, Y. Song, and M. Irving, *Modern power systems analysis*: Springer Science & Business Media, 2010.
[22] B. Borkowska, "Probabilistic load flow," *IEEE Transactions on Power Apparatus and Systems,* pp. 752-759, 1974.
[23] B. B. a. C. H. G. R. N. Allan, "Probabilistic analysis of power flows," *Proceedings of the Institution of Electrical Engineers (London),* vol. 121, p. 12, Dec. 1974
[24] R. Allan and A. L. Da Silva, "Probabilistic load flow using multilinearisations," in *IEE Proceedings C (Generation, Transmission and Distribution)*, 1981, pp. 280-287.
[25] A. L. Da Silva, R. Allan, S. Soares, and V. Arienti, "Probabilistic load flow considering network outages," in *IEE Proceedings C (Generation, Transmission and Distribution)*, 1985, pp. 139-145.
[26] F. Ruiz-Rodriguez, J. Hernandez, and F. Jurado, "Probabilistic load flow for radial distribution networks with photovoltaic generators," *IET renewable power generation,* vol. 6, pp. 110-121, 2012.
[27] F. Ruiz-Rodriguez, M. Gomez-Gonzalez, and F. Jurado, "Reliability optimization of an electric power system by biomass fuelled gas engine," *International Journal of Electrical Power & Energy Systems,* vol. 61, pp. 81-89, 2014.
[28] G. J. Anders, *Probability Concepts in Electric Power Systems* New York: Wiley, January 1990.
[29] P. Amid, and C.Crawford, "A Cumulant-Tensor Based Probabilistic Load Flow Method," IEEE Transactions on Power Systems, 2018.
[30] W. Li, *Probabilistic transmission system planning* vol. 65: John Wiley & Sons, 2011.
[31] P. Zhang and S. T. Lee, "Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion," *IEEE transactions on power systems,* vol. 19, pp. 676-682, 2004.
[32] J. M. Morales and J. Perez-Ruiz, "Point estimate schemes to solve the probabilistic power flow," *IEEE Transactions on Power Systems,* vol. 22, pp. 1594-1601, 2007.
[33] J. Liu, M. Salama, and R. Mansour, "An efficient power flow algorithm for distribution systems with polynomial load," *International Journal of Electrical Engineering Education,* vol. 39, pp. 371-386, 2002.
[34] M. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *IEEE Transactions on power Delivery,* vol. 4, pp. 735-743, 1989.

Walaa Ahmed

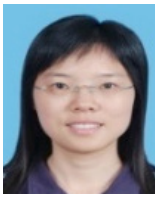
Walaa Ahmed received the B.Eng. from the Faculty of Engineering, South Valley University, Egypt in 2015. She is currently pursuing his MSc degree in Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University. Her research activities include power system modeling, analysis, and optimization.

Ali Selim

Ali Selim received his BSc and MSc degrees in Electrical Engineering from Aswan University, Egypt, in 2010 and 2016 respectively. He is currently a Ph.D. student in the Department of Electrical Engineering at University of Jaén (Universidad de Jaén), Spain. His research interests include mathematical optimization, planning and control of power systems, renewable energies, energy storage, and smart grids.

Salah Kamel

Salah Kamel received the international Ph.D. degree from University of Jaen, Spain (Main) and Aalborg University, Denmark (Host) in Jan. 2014. He is an Assistant Professor in the Electrical Engineering Department, Aswan University. Also, He is a Leader for a power systems research group in the Advanced Power Systems Research Laboratory (APSR Lab), Aswan, Egypt. He is currently a Senior Research Fellow in State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing, China. His research activities include power system modeling, analysis and simulation, and applications of power electronics to power systems and power quality.



Juan Yu

Juan Yu received the Ph.D. degree in electrical engineering in 2007 from Chongqing University, Chongqing, China, where she is currently a Full Professor. Her research interests include static equivalent, optimal reactive power flow, and risk assessment in power and energy systems.

Francisco Jurado

Francisco Jurado obtained the MSc and PhD degrees from the UNED, Madrid, Spain, in 1995 and 1999 respectively. He is Full Professor at the Department of Electrical Engineering of the University of Jaén, Spain. His research activities have focused on two topics: power systems and renewable energy.

Happiness and Technology: Special Consideration of Digital Technology and Internet

Francisco Mochón*

Universidad Nacional de Educación a Distancia - Uned (Spain)

Received 20 November 2018 | Accepted 27 November 2018 | Published 1 December 2018

Abstract

This research paper can be considered a survey about the impact of technology in happiness. The article points out that the scientific approach of happiness states that happiness can be measured and explanatory factors of well-being must be searched empirically. The analysis of technology impact on happiness starts with the opinion of philosophers and social thinkers, and then focus on the revision of empirical research works. The paper concludes highlighting that technology, being the motor of economic well-being, has positive and negative effects on the subjective well-being of individuals. Therefore it is essential to undertake an adequate regulation that promotes positive effects and mitigates the possible harm.

Keywords

Happiness, Subjective Well-Being, Technology, Digital Technology, Internet, Social Networks.

DOI: 10.9781/ijimai.2018.12.004

I. Introduction

Happiness is important because it constitutes a final goal for human beings. Happiness is something to which one aspires and its search motivates human action. An example of this is that a resolution of the United Nations of 2012 states that "the pursuit of happiness is a fundamental human objective" [1].

For centuries the study of happiness has been dominated by non-scientific traditions that are based on the idea that it is up to experts to judge the happiness of human beings [2]. Thus, the tradition of imputation is inspired by the work of philosophers, social thinkers and academics, and it is based on the fact that it is a third party who defines what is the good life [3] [4]. Experts propose criteria for making a judgment and make a list of observable attributes. Based on these observable attributes, the expert imputes the happiness - or well-being - of the people [5].

The tradition of presumption recognizes that happiness is something that people experience. However, instead of inquiring directly and asking people about their welfare state, the tradition uses theories about nature and human behavior. In this way, lists of factors that are presumed to be closely related to a satisfactory life experience are obtained. Within this tradition, happiness is associated with achieving a set of factors that are believed to be relevant to achieve happiness.

Both the tradition of imputation and that of presumption are based on measuring the well-being of people through a judgment made by a third person who considers variables or attributes that are observable. This has led to the conception of well-being as a list of attributes (possessions, deficiencies, and actions) and not as an experience of the people [6] [7].

* Corresponding author.

E-mail address: fmochon@cee.uned.es

II. The Scientific Study of Happiness

In the second half of the 20th century, the scientific study of happiness is born. Pioneering works arise from different disciplines: sociology [8] [9] [10], economics [11] [12]; [13] [14], psychology [15] [16] and political science [17]. In these works it is evident that it is possible to study happiness scientifically based on its direct measurement.

The scientific study of happiness is based on a conception of happiness as a human experience and on the measurement of happiness by asking directly those who experience it. Under the new approach, the information that people provide about their well-being experience is valuable both for knowing their welfare situation and for studying the importance that different personal and social environment factors have for their happiness [18].

Happiness is a human experience so it cannot be conceived in the absence of human beings who experience it. Happiness is neither an academic creation nor an invention of philosophers, but an experience that happens to human beings. Consequently, the work of academics should consist in investigating it in order to understand what their explanatory factors are. The starting point is that happiness refers to people's experience of well-being, and that each subject is the one who can best report this experience because he or she is who experience it. Logically, happiness is inherently subjective, since it is an experience of the subject and this experience cannot exist without the person [19].

Happiness research requires high-level techniques to deal with large information sets in order to extract the relevant information. In the study of happiness there are many observations –as many as persons in the world-, there are many variables, and there are many interrelations and synergies to take account of. In consequence, happiness research benefits from sophisticated models that allow for a better understanding of people's happiness. Without losing contact with what real human beings experience, it is important to use techniques that allow researchers to process all the information to reach valuable conclusions. With this purpose, Computer Science has joined the other disciplines providing its powerful calculation tools to advance the study of happiness [20].

III. Global Synthesis of Life: Life Satisfaction and Happiness

People have wellness experiences and can make a global synthesis of them. This synthesis constitutes a global appreciation of how happy they are. This global synthesis of life, as well as the essential experiences of well-being, constitutes the object of the scientific study of happiness. The overall synthesis of life is usually made with phrases such as "I am happy", "I am satisfied with my life", "my life is going well", and "I feel good about myself". The term happiness is used as a concept that refers to the overall synthesis of satisfaction with life. In other words, happiness refers to how the individual evaluates the overall quality of her or his life [21]. As such, the happiness of individuals will depend entirely on an individual perception and it will be linked to concepts of quality of life and well-being.

The happiness of human beings depends on many factors, some of personal nature and others related to the conditions of their physical and social environment. In this sense, happiness is conditioned by a wide group of variables, among which the following stand out; social relations (family, friends and colleagues), nature of work activities, parenting conditions, personality traits, availability and use of free time, the place where one lives, safety, the existence of children and their ages, couple's relationship, household income, macroeconomic environment, economic occupation, unemployment, health, values, expectations, and the possibility of participation in political decisions [22] [23] [24] [25] [26] [27] [28].

IV. The Measurement of Happiness: the Use of Social Networks as a Possibility

Ed Diener and his collaborators presented a method to measure happiness based on the idea that individuals can consistently identify their level of satisfaction with life on a scale and, as such, what must be done to ask people questions [21] [30] [31]. This way of measuring happiness is the one that justifies conducting surveys like the World Values Survey, and it is the most widely-used method. A numerical response scale is usually used (for example, in the range of 0 to 10), where 0 represents the lowest satisfaction (lowest happiness) and 10, the highest satisfaction (greatest happiness). In addition to the question about the global synthesis of life, one can also ask people about their satisfaction in different domains of life.

As a result of the growth of social networks, a new possibility has emerged to measure happiness. This new approach consists of inferring the feelings of social network users on the basis of a semantic analysis of the words used in their communications and messages. Likewise, a study done by the Vermont Complex Systems Center uses information from Twitter to infer how happy or unhappy people in different states of the United States feel. Specifically, the researchers Dodds and Danforth have developed a method that, by incorporating the direct human evaluation of words, allows to quantify levels of happiness on a continuous scale from a diverse collection of texts [32] [33]. In the study carried out by Dodds and Danforth, on the basisof ten million "tweets", a code for determining to what extent each analyzed message can be catalogued as happy or sad was developed. The study focused on certain key words that were deemed to be indicative. Following this approach, in the article by Mochón and Sanjuan, the happiness of a large group of Latin American countries is measured through the use of social networks. Specifically, the social network used is Twitter. The paper shows that it is possible to calculate, via objective and empirical means, factors that allow to measure happiness through the use of social networks [34].

V. Happines and Social Relations

The idea that the welfare of individuals is conditioned by their social relationships and social context is something generally accepted. It is argued that there is an important interaction between the social context and the attitude of individuals to their environment, which has a notable impact on the subjective satisfaction of people [35] [36] [37]. It has been pointed out that relationships that people cultivate in their lives, are some of the most valuable treasures a person can own. Given that several researchers have dealt with the impact of technology on social relationships and consequently on happiness, we should highlight the social component of subjective well-being.

In the literature on happiness a group of socio-economic determinants of the subjective well-being of individuals has been identified, among which is the network of social relationships; this is social capital and relational goods [37] [38] [39] [40] [41] [42] [43] [44] [45].

The network of social relationships is the result of situations as varied as family and marriage, relationships with friends and neighbors, relationships in the workplace, or the use of new technologies related to the Internet; email, social networks, sms, whatsapp ... The key is that this type of relationships affects happiness and also promotes integrity and trust in others [46].

New technologies have introduced a new way of relating among friends, family and co-workers. Social networks allow people to interact daily with all their friends by sending messages, photos and videos; which makes it easy to share experiences and keep the relationships alive. In this sense, it is worth highlighting the growing importance of Whatsapp as a communication tool. Thus, among the young, Whatsapp is used not only to exchange messages but as a tool to discuss the doubts of class and solve problems that may arise from the duties. In this sense it can be affirmed that social networks have contributed to create new links between people. In any case, its impact on happiness is a controversial issue that requires careful analysis, as we will see in the following epigraphs.

VI. Tecnology and Happiness

In a historical perspective, the relationship between technology and happiness has been a constant object of study by economists and social scientists since the advent of the Industrial Revolution. Generally, attention has focused on the relationship between material prosperity and well-being. In this sense, Gregg Easterbrook in his book The Progress Paradox said that although thanks to advances in technology almost all aspects of Western life have vastly improved in the past century– and in the present -, surprisingly most men and women feel less happy than in previous generations [47].

In any case, the key work to analyze the relationship between prosperity (caused by technological progress) and well-being was made by Professor Richard Easterlin [12] [48] [49] [50], who showed that in developed countries there was no real correlation between a nation's income level and its citizens' happiness. The results of Easterlin's work, known as The Easterlin´s Paradox, state that at a point in time happiness varies directly with income both among and within nations, but over time happiness does not trend upward as income continues to grow. The original conclusion for the United States was based on data from 1946 (when formal surveys of happiness started) to 1970; later evidence through 2014 confirmed the initial finding. The trend in United States happiness has been flat or even slightly negative over a roughly seven decades stretch in which real incomes more than tripled.

Therefore, from Easterlin's research work it is inferred that you could give people more income -and consequently more choice possibilities- and would not have much impact on their sense of well-being. In other words, it seems as if from a certain level of income, people get used to

high levels of income and value less and less the increases in income.

We find this same idea if we refocus the analysis on the relationship between happiness and technology and leave income aside. It seems as if people adapt very easily to the advantages that technology brings and no longer make them happy. So, let's imagine that at the end of the 19th century we asked anyone if they would be happier if they could have a vehicle that would allow them to travel in a day hundreds of kilometers, or if they could cross the Atlantic in a few hours or talk to a person who is located thousands of kilometers away. It is very likely that this person would say yes. However, few people today associate their happiness with having cars, traveling by plane or talking by phone with a relative who lives in another continent. The usefulness of advances in technology is recognized, but we quickly become accustomed to these advantages. Moreover, it is even considered that these advances can be a source of stress and frustration. Therefore, it is not clear that advances in technology make us happier [51].

This facility to adapt to the advantages of new technology coincides with one of the conclusions obtained by research in happiness; that people adapt very quickly to the good news. Thus, for example, it has been shown that if a person wins the lottery, at that time he will feel euphoric and very happy, but after a reasonable time he will return to his habitual levels of happiness [52].

The fact that we adapt very quickly to advances in technology does not mean that technology does not have positive or negative effects on our quality of life and consequently on happiness. The relevant thing is that its net impact is not always easy to determine. We will start with the positive effects of technology on happiness and later we will comment on the negative effects.

VII. Positive Effects of Technology on Happiness

The theory of economic growth has shown with clear clarity that the main driver of growth and improvement of living conditions has been technological progress. In this sense, it would seem logical to think that new technologies not only make people live better but also happier.

In this sense, new technologies, as consumers, have a positive effect at least during certain periods of time by providing us with a wide range of new products, such as cars or household appliances and by improving the quality of them.

Technology can also be used to communicate with one another. Thus, for example, the Internet or a mobile phone are communication tools that can be used to enrich social relationships. As a communication tool, technology can be used as a means to connect, to share knowledge or to empower people. In this sense, its impact on happiness is positive. But the relationship between happiness and technology, when it is used as a tool to communicate, is, as we will see, quite complex.

Technology has also radically changed the nature of work for most workers. This matters because the workplace is very important to people sense of well-being. With the industrial revolution, mechanization allowed workers to escape from agriculture. Although they were often thrown initially into hard industrial jobs, over time, and thanks to the significant increases in productivity, very substantial improvements in working conditions and wages have taken place. More recently, the appearance of the digital society, and the advent of knowledge-based businesses, means that workplaces have become less formal and more open, often creating a really nice work environment [53][51]. Thanks to technology, we have become globalized, becoming individuals without borders, overcoming the limitations of place and space. Some people work in offices while others do it from their homes, or even in a cafeteria. We move fluidly in and out of the hazy world of the internet-based "cloud" with part of our belongings in the physical world and other part in the virtual world.

In any case, where technology has had a more significant impact on the well-being of people is in the health field. An example of this is the considerable increase in life expectancy that has taken place in the vast majority of countries in the last hundred years. The highlight is that the majority of people are happy to be alive, and if they live longer they will feel happier.

VIII. Negative Effects of Technology on Happiness

The origin of criticism of technology has focused on what Heidegger's terminology is known as the question of technology – that is, the impact of technology on our humanity [54]. In this sense, it has been questioned people's ability to use technology to their own ends. Heidegger highlights the role of technology in bringing about the decline of human beings by constricting our experience of things as they are. He argues that we increasingly view human beings, only technologically — that is, we view people only as raw material for technical operations. We treat even human capabilities as though they were only means for technological procedures. People are mere human resources to be arranged, rearranged, and disposed of [55]. We tend to believe that technology is a means to our ends and a human activity under our control.

But in truth we now conceive of means, ends, and ourselves as fungible and manipulable. For these reasons, Heidegger denounces technology harmful effects and the view that technology is a neutral tool to be wielded either for good or evil.

Following the contribution of Heidegger, the two main criticisms of technology for its impact on happiness have a somewhat contradictory meaning. On the one hand, it is pointed that technological progress is leading to an ever more rigid, controlled, soulless society, in which it is easier for people to be manipulated and monitored. In this sense Jacques Ellul shows his concern for the emergence of a technological tyranny over humanity [56]. On the other hand, it has been criticized, referring especially to the role of television, how the most popular media of a time in history shapes the discourse of the world [57]. From a different perspective, Putman has pointed out that technology is contributing to the reduction in all the forms of in-person social intercourse. The consequence of this is a fragmented society, in which traditional relationships are harder to sustain, and a reduction of the social capital [58].

From these pioneering contributions, the idea that technology disrupts social relationships and fractures the community has gained followers and, as it will be seen later, has become central to the critique of the Internet. From this perspective, technology, and more specifically the Internet, supposedly isolates people from what critics always call the real world. One of the first times this criticism was pointed out was in a famous study conducted among the residents of the city of Pittsburgh (US), published in September 1998 [59]. This article points out that the Internet, being a communication tool, instead of allowing people to connect with a much wider set of potential friends and exposing them to information they might otherwise never have come across, the Internet makes people more depressed and lonely than they would otherwise have been. According to the authors of this work, the Internet could change the lives of average citizens as much as did the telephone in the early part of the 20th century and television in the 1950s and 1960s. For this reason, it is interesting to try to find out whether the Internet is improving or harming participation in community life and social relationships. According to the results of this research work the Internet was used extensively for communication. Nonetheless, greater use of the Internet was associated with declines in participants' communication with family members in the household, declines in the size of their social circle, and increases in their depression and loneliness. The authors described this result as

a paradox, since the Internet, as a communication tool, should improve the subjective well-being of individuals.

Although this research work had a great impact, its statistical support is not very solid, only 169 people from 73 households were interviewed. In fact, a few years later, some of the authors re-analyzed the issue and found that negative effects of Internet dissipated [60]. In the new research work the authors report that the people investigated generally experienced positive effects of using the Internet on communication, social involvement, and well-being. However, using the Internet predicted better outcomes for extroverts and those with more social support but worse outcomes for introverts and those with less social support.

The criticism of technology, and particularly the Internet, for its impact on social relations, is especially relevant from the perspective of happiness and deserves special attention. Keep in mind that one of the main conclusions of the scientific study of society is the existence of a high correlation between happiness and social relations. Logically a tool as broad and ubiquitous as the Internet will have a multitude of effects, some may be negative but others not. In addition, in essence, the Internet is a communications technology that, like the telephone, allows people to expand their affective and informational networks and this is something that people value positively. Obviously, the Internet is not the ideal place to establish all kinds of communications, but in any case it is a public communication area that works openly and without gatekeepers. Therefore, criticizing technology and, in particular, the Internet in a generalized manner due to its alleged negative effects on subjective well-being may be excessive [51].

A less controversial way in which technology can negatively affect people's happiness is in its relentless generation of newness [51]. One of the implications of studies on happiness is that people have a hard time being happy with what they have when they know that others have more or have better things [61] [62]. Nowadays technological change takes place so quickly that if we buy any technological product (a mobile phone, a computer, a television,...), we know that in a few months there is going to be a better, faster version of the products. We will be left with obsolete products while other people will have new and more technologically advanced products, which will negatively affect our well-being. There is no way to avoid this feeling that is in the heart of the modern consumer.

And then there are the worries about AI [artificial intelligence] and the technological displacement of labor. Simply by focusing on robotics, it has the potential to transform lives and work practices. Its impact will be increasing, as the interactions between robots and people multiply. Although there is no consensus on the effects that this will have on employment, what is indisputable is that its impact will be very important and difficult doubts arise. How should the benefits of robotics be distributed? The universal basic income will no longer be a possibility and will become an obligation and, given the important effect that employment has on subjective well-being, how will all this affect the happiness of the individual? [63].

IX. Digital Technology and Happiness

We are going to focus the analysis on the incidence of digital technology and, in particular, of Internet. In a recent research carried out in Spain [37], the incidence of social networks on happiness is analyzed. It is observed that individuals, regardless of their age, who use social networks have, on average, a greater life satisfaction than those who do not use them. The results of the survey show that, in addition, those with more than 65 years of age who use social networks feel more satisfied even than those of mature age. It seems that social networks can be a good way to combat loneliness. The feeling of being communicated at any time of the day with your friends and family and being able to share images, videos, etc. with them, makes individuals more satisfied.

In some other research works it has also been found that virtual relationships can be as intimate as in-person relationships [64]. In fact, Bargh and colleagues found that online relationships are sometimes more intimate [65]. This can be especially true for those individuals who are more socially anxious and lonely—such individuals who are more likely to turn to the Internet to find new and meaningful relationships [66] [67]. In other words, these research works suggest that for people who have a hard time meeting and maintaining relationships, due to shyness, anxiety, or lack of face-to-face social skills, the Internet can offer a safe, nonthreatening place to develop and maintain relationships. Likewise, some researchers have shown that young people are using digital technology and online social media within their everyday lives to enrich their social relationships [68].

In any case, the effects are not always positive; depending on how the Internet is used and, in particular, the social networks, these can be beneficial or harmful [69] [70] [71]. In this sense, one reason why Internet technology can have negative effects on happiness is due to the corporate and governmental power to surveil users (attendant loss of privacy and security). To this we must add the effect of the addictive technologies that have captured the attention and mindspace of the youngest generation [72].

Thus, although until recently social networks were presented as an instrument of socialization because they allow sharing ideas, connecting with friends and alleviating the isolation that the Internet could generate, and even promoting social change and the empowerment of citizens, in recent dates doubts have grown. Especially since 2017, criticism of the networks has proliferated, largely due to the scandals related to Facebook [73]. It has been argued that the platforms are designed to hook the users and get them to spend as much time as possible in them creating addiction, which tightens the debate as they filter the information showing only a view of the facts and contaminate it with false information, and that even they can be a tool to manipulate democratic electoral processes.

X. The Expert's Opinion

As a final balance on the impact of digital technology on happiness, we will analyze the results of research that adopts a similar approach as that used by researchers to measure happiness: ask the interested parties [72]. In this sense, Pew Research Center and Elon University's Imagining the Internet Center decided queried 1,150 technology experts, scholars and health specialists on the following question: Over the next decade, how will changes in digital life impact people's overall well-being physically and mentally?

The conclusions of this investigation can be summarized by saying that 47% of those queried predict that individuals' well-being will be more helped than harmed by digital life in the next decade, while 32% say people's well-being will be more harmed than helped, and the remaining 21% predict there will not be much change in people's well-being compared to now [72].

As a general comment, it can be said that many of those who argue that human well-being will be harmed also acknowledge that digital tools will continue to enhance various aspects of life. They also note that there is no turning back in the sense that new technologies are here to stay. At the same time, hundreds of them suggested interventions in the coming years that they feel could mitigate the problems and emphasize the benefits. Moreover, many of the hopeful respondents also agree that some harm will arise in the future, especially to those who are vulnerable.

To analyze the answers of the interviewees in a systematic way,

these can be classified into three categories: 1) The positive effects of digital technology. 2) The negative effects of digital technology. 3) Remedies to mitigate the possible negative effects.

1) The Positive Effects Of Digital Technology

The benefits of digital life on happiness are analyzed in terms of the following four factors [72]:

- *Connection*. Digital life links people to people, contributing to spread the knowledge, facilitating education and suplying entertainment anywhere globally at any time in an affordable manner. People need to be connected and the Internet is a communication tool par excellence. In subjects specific to society, science, education or politics, the Internet connects people by facilitating rewarding information and relationships.
- *Commerce, government and society*. Digital life revolutionizes civic, business, consumer and personal logistics, opening up a world of opportunity and options. To show the advantages of a hyperconnected society, let's think about the massive benefits to life from access to finance, to online shopping, to limitless free research opportunities, to keeping in touch with loved ones in far-away places.
- *Crucial intelligence*. Digital life is essential to tapping into an ever-widening array of health, safety, and science resources, tools and services in real time. Advances in computer science have meant that information is increasingly distributed globally and openly. For example the relatively recent trends towards openness in scientific publications, scientific data and educational resources are likely to make people across the world better off by expanding individuals' access to a broad set of useful information, by decreasing barriers to education and by enhancing scientific progress.
- *Contentment*. Digital life empowers people to improve, advance or reinvent their lives, allowing them to self-actualize and meet soul mates. The internet helps to break down barriers and supports people in their ambitions and objectives. Internet helps people achieve their desire to improve their education, to communicate with others, to share their experiences, to create networks of enterprise, commerce, culture, sports... All these are supported by digital technologies.
- *Continuation toward quality*. Emerging tools will continue to expand the quality and focus of digital life; the big-picture results will continue to be a plus overall for humanity. The future artificial intelligence (AI) will enhance human well-being. Throughout history it has been shown that human beings need tools and want improvements, and AI is facilitating them and will continue to do so. And as the saying goes 'First we make our tools, then our tools form us.'

2) The NegativeEffects Of Digital Technology

The negative impact of digital technology on happiness is analyzed in terms of the following five factors [72]:

- *Digital deficits*. People's cognitive capabilities will be challenged in multiple ways, including their capacity for analytical thinking, memory, focus, creativity, reflection and mental resilience. The digital society is characterized by an intrusive connectivity that has harmful cognitive and emotional consequences.
- *Digital addiction*. Internet businesses are organized around dopamine-dosing tools designed to hook the public. The current generation of tools for consuming attention is very effective and can cause addictive effects. Network effects and economies of scale have placed control of these tools in a very small number of very powerful companies.
- Digital distrust/divisiveness. Personal agency will be reduced and emotions such as shock, fear, indignation and outrage will be strengthened. Although technologies are created with a sincere desire to advance understanding of mood, cognition, etc., or with the pretension of facilitating the control of our response, the actual implementation of these techniques and devices is likely to be quite different. It is possible that they may finally be used to reduce well-being because a population in a state of fear and anxiety is far more malleable and profitable.
- Digital duress. Information overload + declines in trust and face-to-face skills + poor interface design = rises in stress, anxiety, depression, inactivity and sleeplessness. There are organizations that are actively vying people´s attention, distracting them with smartphone notifications, highly personalized news, addictive games, Buzzfeed-style headlines and fake news.
- *Digital dangers*. The structure of the internet and pace of digital change invite ever-evolving threats to human interaction, security, democracy, jobs, privacy,... In addition, many people are unable to adapt to the behaviors and needs that digital technology requires.

3) Remedies to Mitigate the Possible Negative Effects

Five possible lines of action are presented to combat the possible problems that digital technologies may cause [72]:

- *Reimagine systems*. Societies can revise both tech arrangements and the structure of human institutions – including their composition, design, goals and processes. The challenge to be overcome is neither more nor less than simply learning to call what we have created what it really is, and then regulate and manage it accordingly.
- *Reinvent tech*. Things can change by reconfiguring hardware and software to improve artificial intelligence (AI), virtual reality (VR), augmented reality (AR) and mixed reality (MR). We can resort to human-centered technology design to improve our experiences and outcomes, to better serve us.
- *Regulate*. Governments and/or industries should create reforms through agreement on standards, guidelines, codes of conduct, and passage of laws and rules. Security and privacy cause great concern for what is necessary to come to some kind of detente.
- *Redesign media literacy*. Formally educate people of all ages about the impacts of digital life on well-being and the way tech systems function, as well as encourage appropriate, healthy uses. The primary change needs to come in education. From a very early age, people need to understand how to interact with networked, digital technologies.
- *Recalibrate expectations*. People must gradually evolve and adjust to digital changes. People must learn how to reign over the pitfalls, threats, bad guys and ill-meaning uses.

VII. Final Reflections

Does technology make us less happy or happier? This is the question we have tried to answer throughout this article. From the analysis made in previous pages, it is inferred that the most objective analysis is not the one made by social thinkers or philosophers. The studies of this type of authors are interesting to become aware of trends and anticipate possible future issues. However, they are not usually the most appropriate way to obtain the specific response to an issue, as in our case, to know the incidence of technology on the subjective well-being of individuals.

The most reliable results are obtained when research work is carried out on the impact of technology in specific cases. From them it is evident that technology tends to have a positive impact on the subjective wellbeing of individuals but it can also generate negative

effects. On the other hand, it should not be forgotten that technology, in many cases, offers tools; and the impact of these on happiness to a large extent will depend on how we use them. Technology can be a very important source of well-being, although it is essential to learn to ration its use. You have to know how to discriminate between its possible uses and discern those platforms that are worth getting involved with and those in which we should not enter.

A similar conclusion is reached when analyzing the opinion of the technology experts, scholars and health specialists. They affirm that technology will continue to improve many aspects of our life but in certain aspects it may harm the subjective well-being of the individuals, especially those who are vulnerable.

In any case, and given that progress and technological innovation is essential for the advancement of society, it is necessary to pay special attention to regulation. Only through proper regulation we can mitigate the possible damages derived from technology and emphasize its benefits.

References

[1] Rojas, M. and Martínez, I. (coords.) (2012) La Medición, Investigación e Incorporación en Política Pública del Bienestar Subjetivo: América Latina. Reporte de la Comisión para el Estudio y Promoción del Bienestar en América Latina, Foro Consultivo Científico y Tecnológico, México.

[2] Rojas, M. (2014) El Estudio Científico de la Felicidad, Fondo de Cultura Económica.

[3] McMahon, D. (2006) Una Historia de la Felicidad, Taurus.

[4] Tatarkiewicz, W. (1976) Analysis of Happiness, Martinus Nijhoff.

[5] Rojas, M. (2017) El estudio de la felicidad. In La felicidad de los españoles. Iglesias, J. and De Juan, R. (coords.) Tecnos.

[6] Rojas, M. (2006) Life Satisfaction and Satisfaction in Domains of Life: Is it a Simple Relationship? Journal of Happiness Studies, 7(4), 467-497.

[7] Rojas, M. (2007) The Complexity of Well-Being: A Life-Satisfaction Conception and a Domains-of-Life Approach. En I. Gough y A. McGregor (eds.). Researching Well-Being in Developing Countries. Cambridge University Press.

[8] Veenhoven, R. (1984) Conditions of Happiness. Kluwer Academic.

[9] Campbell, A. (1976) Subjective Measures of Well-Being, American Psychologist, 31, 117-124.

[10] Campbell, A. Converse, P.E. and Rodgers, W.L. (1976) The Quality of American Life. Perceptions, Evaluations and Satisfactions, Russel Sage Foundation.

[11] Easterlin, R. (1973) Does Money Buy Happiness? The Public Interest, 30, 3-10.

[12] Easterlin, R. (1974) Does Economic Growth Improve the Human Lot? Some Empirical Evidence, in David, P.A. and Reder, M.W. (eds.), Nations and Households in Economic Growth, Academic Press, 89-125.

[13] Praag, BMS van (1968) Individual Welfare Functions and Consumer Behavior, North Holland.

[14] Praag, BMS van (1971) The Welfare Function of Income in Belgium: An Empirical Investigation, European Economic Review, 2, 337-369.

[15] Diener, E. (1984) Subjective well-being, Psychological Bulletin, 95, 542-575.

[16] Andrews, F.M. and Withey, S.B. (1976) Social Indicators of Well-Being. Plenum Press, New York.

[17] Lane, R. (1991) The Market Experience, Cambridge University Press.

[18] Layard, R. (2006) Happiness: Lessons from a New Science. Penguin.

[19] Ferrer-i-Carbonell, A. (2002) Subjective Questions to Measure Welfare and Well-Being, Discussion paper TI 2002-020/3. Tinbergen Institute.

[20] Mochón, F. and, Rojas, M. (2014) Editor's Note. International Journal of Interactive Multimedia and Artificial Intelligence. (Special Issue on AI Techniques to Evaluate Economics and Happines), 2(5).

[21] Diener, E. Emmons, R.A. Larsem, R.J. and Griffin, S.A. (1985) The Satisfaction With Life Scale. Journal of Personality Assessment, 49(1), 71-75.

[22] Frey, B. and Stutzer, (2001) Happiness and Economics: How the Economy and Institutions Affect Human Well-Being. Princeton University Press.

[23] Ahn, N. and Mochón, F. (2010) La felicidad de los españoles: Factores explicativos. Revista de Economía Aplicada, XVIII, 54, 5-31.

[24] Praag, BMS van and Ferrer-i-Carbonell, A. (2004) Happiness Quantified: A Satisfaction Calculus Approach. Oxford University Press.

[25] Rojas, M. (ed.) (2016) Handbook of Happiness Research in Latin America, Springer.

[26] De Juan, R. Mochón, F. and Rojas, M. (2014) Expectations and Happiness: Evidence from Spain. Journal of Social Research & Policy, 5(2), pp. 89-102.

[27] Ahn, N. Mochón, F. and De Juan, R. (2012). La felicidad de los jóvenes. Papers Revista de Sociología. 97/2, 407-430.

[28] Argyle, M. (2002) The Psychology of Happiness. Routledge.

[29] Cummins, R. A. (1996) The Domains of Life Satisfaction: An Attempt to Order Chaos. Social Indicators Research, 38, 303-332.

[30] Crooker, K. y Near, J. (1998) Happiness and Satisfaction: Measures of Affect and Cognition? Social Indicators Research, 44, 195-224.

[31] Grinde, B. (2002) Happiness in the Perspective of Evolutionary Psychology. Journal of Happiness Studies, 3, 331-354.

[32] Dodds, P. and Danforth, C. (2009) Measuring the Happiness of LargeScale Written Expression: Songs, Blogs, and Presidents. Journal of Happiness Studies, 11(4), 441-456.

[33] Frank, M. Mitchell, L. Dodds P, and Danforth, C. (2013) Happiness and the Patterns of Life: A Study of Geolocated Tweets. Scientific Reports 2013. Vol. 3, No: 2625.

[34] Mochón, F. and Sanjuán, O. (2014) A First approach to the implicit measurement of happiness in Latin America through the use of social networks. International Journal of Interactive Multimedia and Artificial Intelligence, 2(5), 17-23.

[35] Diener, E. Seligman, M. Choi, H. and Oishi, S. (2018) Happiest People Revisited. Perspectives on Psychological Science. March 29. https://journals.sagepub.com/doi/abs/10.1177/1745691617697077

[36] Mochon, F. and De Juan, R. (2015) Happiness and Social Capital: Evidence from Latin American Countries. Handbook of Happiness Research in Latin America. Editorial Springer.

[37] Mochón, F. y R. de Juan. (2017) Capital social y bienes relacionales. In La felicidad de los españoles. Iglesias, J. and De Juan, R. (coords.) Tecnos.

[38] Frey, B. and Stutzer, A. (2000) Happiness, Economy and Institutions. Economic Journal, 110, 918-938.

[39] Frey, B. and Stutzer, A. (2002) What can Economists Learn from Happiness Research? Journal of Economic Literature, 40, 402-435.

[40] Layard, R. (2003) Happiness: Has social science a clue? Three Lectures. First lecture: what is Happiness?; second lecture: income and Happiness: rethinking economic Policy; third lecture: what would make a Happier society. Lionel Robbins memorial lectures. Londres: London school of economic.

[41] Veenhoven, R. (1993) Happiness in Nations: Appreciation of Life in 56 Nations. Rotterdam: Erasmus.

[42] Veenhoven, R. (2000) The Four Qualities of Life. Ordering Concepts and Measures of the Good Life. Journal of Happiness Studies, 1, 1-39.

[43] Veenhoven, R. (2001) What we do Know about Happiness? Working Paper. Rotterdam: Erasmus University.

[44] Iglesias, E. Pena, A and Sánchez, JM. (2013) Bienestar subjetivo, renta y bienes relacionales. Los determinantes de la felicidad en España. Revista Internacional de Sociología, 71(3), 567-592.

[45] Gui y Sugden, 2005). Gui, B. and Sugden, R. (2005) Why Interpersonal Relations Matter for Economics. In Gui, B. and Sugden, R. (Eds.) Economics and Social Interaction Accounting for Interpersonal Relations. Cambridge (Mass.): Cambridge University Press, 1-22.

[46] Nussbaum, M. C. and Sen, A. K. (1993) The Quality of Life. Oxford: Clarendon Press.

[47] Easterbrook, G. (2004) The Progress Paradox: How Life Gets Better While People Feel Worse. Random House.

[48] Easterlin, R. (2017) Paradox Lost?. Review of Behavioral Economics. 4 (4): 311-339.

[49] Easterlin, R. A. and Angelescu, L. (2009) Happiness and Growth the World Over: Time Series Evidence on the Happiness-Income Paradox, IZA Discussion Paper, num. 4.060.

[50] Easterlin, R. A., Angelescu, L., Switek, M., Sawangfa, O. y Zweig, J.S. (2010) The Happiness-Income Paradox Revisited. Proceedings of the National Academy of Sciences of the United States of America (PNAS).

[51] Surowiecki, J. (2005) Technology and Happiness. MIT Tecnology Review.

January 1.https://www.technologyreview.com/s/403558/technology-and-happiness/
[52] Rojas, M. (2009) Consideraciones sobre el Concepto de Progreso. In M. Rojas (coord.). Midiendo el progreso de las sociedades: reflexiones desde México. México, Foro Consultivo Científico y Tecnológico, 71-78.
[53] Hochschild, R. A. (1997) The time bind: when work becomes home and home becomes work. Metropolitan Books.
[54] Heidegger, M. (1977) The Question Concerning Technology, and Other Essays. Garland Publishing,
[55] Blitz, M. (2014) Understanding Heidegger on Technology. The New Atlantis. Number 41, Winter 2014, pp. 63-80. https://www.thenewatlantis.com/publications/understanding-heidegger-on-technology
[56] Ellul, J. (1964) The Technological Society. John Wilkinson.
[57] Postman, N. (1985) Amusing Ourselves to Death: Public Discourse in the Age of Show Business. Penguin.
[58] Putnam, Robert D (1995) Bowling Alone: America's Declining Social Capital. Journal of Democracy. 6 (1): 65–78.
[59] Kraut R, Patterson M, Lundmark V, Kiesler S, Mukopadhyay T and Scherlis W. (1998) Internet paradox. A social technology that reduces social involvement and psychological well-being? American Psychologist, Vol 53(9), Sep, 1017-1031.
[60] Kraut, R, Kiesler, S. Boneva, B. Cummings, J. Helgeson, V. and Crawford, A. (2001) Internet Paradox Revisited. Journal of Social Issues. October 12. Version 16.2 http://kraut.hciresearch.org/sites/kraut.hciresearch.org/files/articles/kraut02-paradox-revisited-16-20-2.pdf
[61] Buss, D. (2000) The Evolution of Happiness. American Psychologist, 55(1), 15-23.
[62] Praag, BMS van. Frijters, P. and Ferrer-i-Carbonell, A. (2003) The Anatomy of Subjective Well-being. Journal of Economic Behavior and Organization, 51, 29-49.
[63] Mercader, J.R. (2017). Robotización mecanización pérdida de empleo. Trabajo y Derecho, 27, 13-24.
[64] Brannan, D. and Mohr, C. D. (2018) Love, friendship, and social support. In R. Biswas-Diener & E. Diener (Eds), Noba textbook series: Psychology. Champaign, IL: DEF publishers.
[65] Bargh, J. A, McKenna, K. Y. A. and Fitsimons, G. G. (2002) Can you see the real me? Activation and expression of the true self on the Internet. Journal of Social Issues, 58, 33–48.
[66] McKenna, K. Y. A. Green, A. S. and Gleason, M. E. J. (2002) Relationship formation on the Internet: What's the big attraction? Journal of Social Issues, 58, 9–31.
[67] McKenna, K. A. (2008) MySpace or your place: Relationship initiation and development in the wired and wireless world. In S. Sprecher, A. Wenzel, & J. Harvey (Eds.), Handbook of relationship initiation (pp. 235–247). New York, NY: Psychology Press.
[68] Hynan, A. Murray, J. and Goldbart, J. (2014) Happy and excited': Perceptions of using digital technology and social media by young people who use augmentative and alternative communication. Child Language Teaching and Therapy. January 27. https://journals.sagepub.com/doi/abs/10.1177/0265659013519258
[69] Pénard, T. Poussing, N. and Suire, R. (2013) Does the Internet make people happier?. Journal of Socio-Economics, Elsevier, 46, 105-116.
[70] Holsten, H. (2018) How does internet use affect well-being? February 26. https://phys.org/news/2018-02-internet-affect-well-being.html
[71] Zhan, G. and Zhou, Z. (2018) Mobile internet and consumer happiness: the role of risk, Internet Research, 28(3), 785-803. https://doi.org/10.1108/IntR-11-2016-0340
[72] Anderson, J. and Rainie, L. (2018) The Future of Well-Being in a Tech-Saturated World. Pew Research Center. Internet & Technology. April 17. http://www.pewinternet.org/2018/04/17/the-future-of-well-being-in-a-tech-saturated-world/
[73] Carmody, T. (2018) Facebook is in a Trust Crisis. Adweek. January 23. https://www.adweek.com/digital/facebook-is-in-a-trust-crisis/

Francisco Mochón Morcillo

Ph.D. in economics from the Autonomous University of Madrid and from Indiana University (U.S.A.). Professor of Economic Analysis at the UNED. He has been Advisor to the Ministry of Economy and Finance of Spain, General Director for Financial Policy of the Government of Andalusia, CEO of the consulting and research firm ESECA and Chief Financial Officer (CFO) of Telefónica Group (1988-2001). He has been the Chairman of the Social Board of the University of Málaga. Member of ; the expert group in Digital Economy of the European Economic and Social Commitee (EESC); the advisory committee of U-TAD; and the advisory committee of the Futures Market of Olive Oil (MFAO). He has published numerous research articles and is the author of more than fifty books on economics, finance and business. Currently his research interests are Digital Economy and Economics of Happiness. He currently directs a book collection on Big Bata (Garcia Maroto Editores) and another on the digital transformation of organizations (Alfaomega).

LA UNIVERSIDAD
EN INTERNET

Rectorado
Avenida de la Paz, 137
26006 Logroño (La Rioja)
t (+34) 941 21 02 11

www.unir.net