sciendo

International Journal of Advanced Statistics and IT&C for Economics and Life Sciences

Volume 10
Number 1
December 2020

ISSN 2559 - 365X
eISSN 2067 - 354X

International Journal of Advanced Statistics and IT&C
for
Economics and Life Sciences
Vol. 10, Nr. 1, 2020

CONTENTS

ISSN 2067 – 354X
eISSN 2067 - 354X

sciendo

DOI: 10.2478/ijasitels-2020-0001

READING IN RESEARCH AND INFORMATION CENTRES. CASE STUDY of LBUS

Volovici Rodica [1], *Capatana Andreea* [1]

[1]*Lucian Blaga University of Sibiu, Romania*
rodica.volovici@ulbsibiu.ro

Abstract

The paper focuses of to open the minds of pupils and turn them to reading, teacher librarians must prepare pupils, bring them face to face with reading, situations, themes from each reading instance, so that they can analyze, describe, freely express themselves, using the right words to convey their thoughts, feelings, emotions, and compare the information they came across while reading with their own experiences, the situations they've lived through, thus anchoring them in daily reality.

Keywords: Teacher, Librarians, Information Centers, Research

1 Research and Information Centers - Spaces Dedicated to Reading

"When reading a book, I feel like a different person."
(Henri James)

The major changes that have occurred in the information society, which most people see as a revolution, have given rise to a new concept, i.e. the Research and Information Centre, a place especially dedicated to documentation.

Therefore, according to article 1 of the Regulation regarding the Organizing and Functioning of the Research and Information Centre, the RIC is a multidisciplinary resource center providing pupils, teachers and the local community with a space dedicated to training, communication and information, a laboratory to experiment with new educational technologies, a space for culture, openness, meeting and integration.[14]

In other words, Research and Information Centers meet pupils, teachers, and the whole local community's needs for information, offering a work tool that supports their personal and professional development, the acquiring of new competences, such as: autonomy in research, comparing

sources and various supports, organizing their research work by selecting, exploiting and communicating information; besides, for teachers, the RIC makes available diverse documentary materials enabling them to adapt their work methods and techniques to the learning needs of 21st-century pupils.

Article 2 of the Order of the Ministry of Education and Research no. 3328/08.03.2002 on allowing Teaching Corps Houses to set up Research and Information Centre defines these structures as follows: multidisciplinary resource centers that provide pupils, teachers and communities information in various formats (book funds, magazines, audio and video tapes, CD-ROMs, computers, etc.), implement cultural animation projects, unfold pedagogical activities, as well as life-long learning actions for teachers and auxiliary teaching staff in the pre-university education system. [25]

Thus, RICs are considered a landmark of innovative processes in Romanian schools, addressing the entire educational community, which includes not only the pupils and the teachers in those schools that have RICs, but also public segments connected to the school: parents, members of the local community, pupil's associations and clubs, etc.

Françoise Chapron, honorary president of FABDEN (Fédération des associations des professeurs documentalistes de l'Éducation nationale, French Ministry of National Education)[26] and associate professor of Rouen University, claims that an important role performed by the RIC is that, through the activities organized here, users (especially pupils) are prepared to assimilate reading codes and strategies to use the new information available in the RIC, which turns this research center into a learning laboratory fit to train them to use other cultural structures, such as: municipal libraries, specialized research centers, communication networks.[27]

2 The role of teacher librarians in training reading for pleasure

As shown above, teacher librarians contribute to the pupil's cultural training, acting not only as mediators between books and readers, but also as animators of cultural life and information efforts.

From my point of view, the role of teacher librarians is to undertake, develop, implement new ways to bring pupils closer to books.

Teacher librarians have to adapt to the real needs of the pupils in today's generation, and especially to learn and get to know how to approach these needs, so that they can support pupils through interesting activities that draw them to books. In order to ensure that the activities conducted by teacher librarians are more attractive, these should include more than mere reading, but blend harmoniously with dance, drawing exhibitions, watching films.

Alongside specialized teachers, teacher librarians should guide pupils to develop competences, such as:

Identifying various types of resources available in a RIC;
Selecting the appropriate documents for their specific research topic;
Becoming familiar with and identifying various search tools (works of reference, clarification systems, indexes, etc.);
Understanding and extracting digital search information, etc.;
Communicating the results of their research in an organized manner.

Teacher librarians are reading consultants, the binder between school subjects and a psychopedagogue, always ready to answer readers' questions. That is why, teacher librarians must infer or even be familiar with the reader's knowledge level, their perception level and their capacity to assimilate information. The activity of teacher librarians revolves around the needs of RIC users.
Therefore, I believe that teacher librarians should be guided by the Latin proverb *timeo hominem unius libri*, meaning that, even those who have only read one book, but have gotten to know it thoroughly, are competent users and opponents to fear in a debate.
Professional and psycho-pedagogical training, communication skills, empathy, good humor, courteousness and availability, the talent to be close to pupils, a warm, welcoming smile are features of a true teacher librarian.
Teacher librarians will not show disapproval of or criticize pupils, for any child can succeed, the essential thing is to not push them to fail through exaggerated or premature requests[28], on the contrary, aim to stimulate pupils to achieve the best results they can, and guide them to secure results that can raise their self-confidence and creativity level.
At the same time, the role of teacher librarians is to organize activities in the RIC to develop pupil's competences, inform and train them, meet some of their curiosities, cultivate their taste for reading, educate them as citizens, develop them as autodidacts (searching, selecting, treating, communicating information), support them to interact better with those around them (by expressing and defending their opinions, accepting the opinions of others), etc.
Considering all the above-mentioned features of teacher librarians, it's safe to say that their role is to train pupils and help them develop certain competences, skills and abilities.
While in the education system, pupils' interaction with books cannot be replaced, and so it's very important for teachers (librarians or specialized) to join hands to consciously bring pupils closer to the world of books, so they can experience the joy of reading, as each book is a journey through souls, thoughts and beauties [4] (Mihail Sadoveanu).
Fundamentally, I believe that teacher librarians should be fascinating, as, according to Augusto Cury [29], *the highest skill of a fascinating teacher is to be able to permeate the minds of their pupils and to find special answers, different from those they are used to. Fascinating teachers turn information into knowledge and knowledge into experience.*

3 Case Study - Interest for Reading

This study aimed at determining the level of interest for reading.
The methodology employed consisted of a survey (data gathering instrument) applied at the "Lucian Blaga" University of Sibiu. The survey included a set of 20 questions. Following the distribution and filling in of the survey, the data was processed and interpreted. We interpreted the results of each question, as well as established correlations between questions. The survey was filled in by a total number of 38 readers.
Most respondents (19 people), i.e. half of the participants in the study, are 25 years old or under.

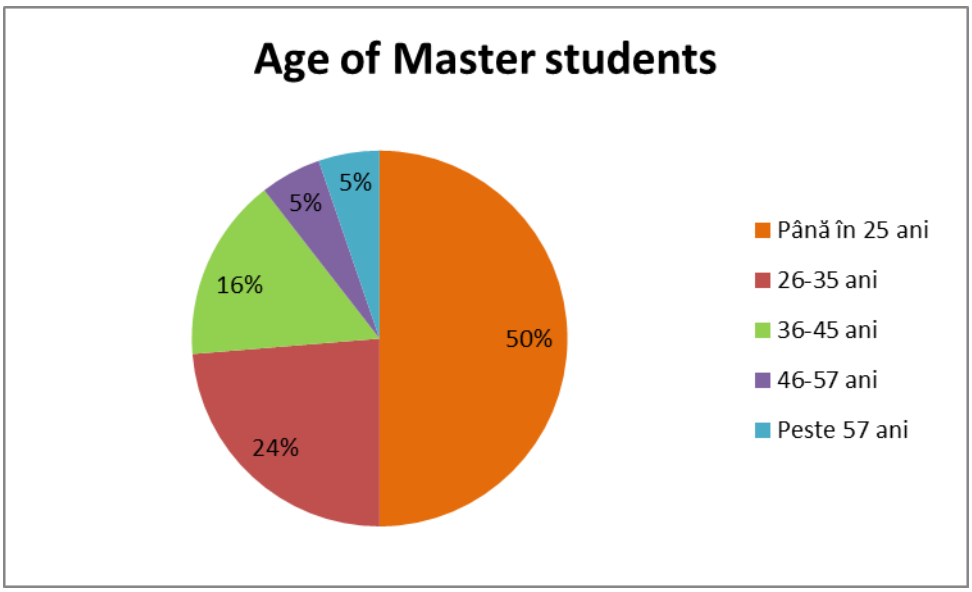


Figure 5 - Age of Master students

As to their gender, 27 are female, while only 11 are male. Moreover, 24 live in the urban environment, while the remaining 14 live in rural environments.

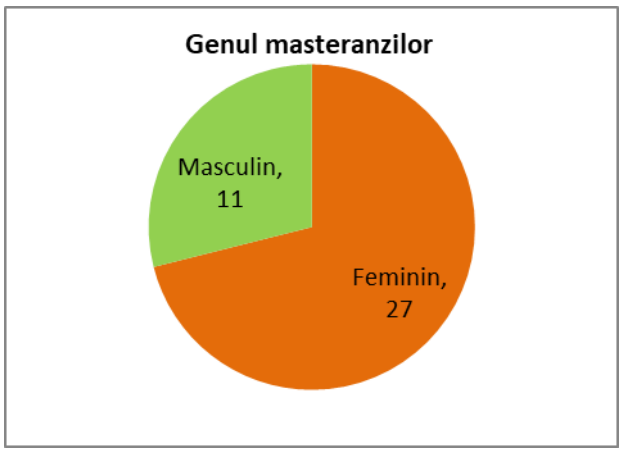


Figure 6 - Gender of Master students

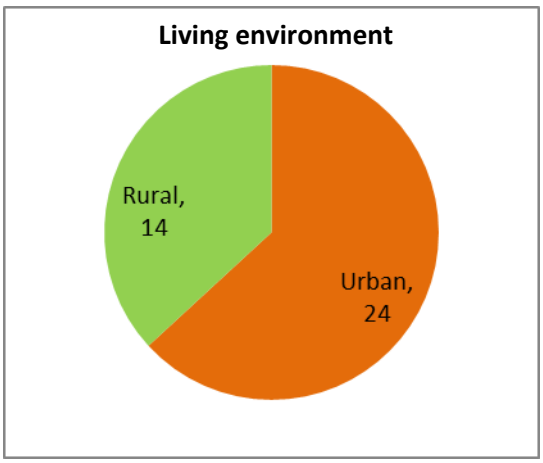


***Figure 7* – Living environment**

As mentioned above and according to the table below, the survey was distributed to Master students from all the Faculties within the "Lucian Blaga" University of Sibiu. Most of the Master students (11) are enrolled in a master program at the Faculty of Engineering, 10 studying Quality management, and 1 the Management of Industrial Affairs. A similar number of respondents (11) are enrolled in a master program at the Faculty of Letters and Arts, 8 majoring as teacher librarians and 3 in Information and Research Sciences (Library sciences). The number of respondents from other Faculties was lower, as this study mainly aimed to compare the two Faculties that have very different fields of activity: Engineering - sciences, and Letters and Arts - focusing on humanities, with a view to discovering whether there was a difference between the two in what regards the interest for reading, how often they read, as well as other opinions on reading.

***Table 7*– Respondents' Master Major**

Faculty of Engineering	Quality management	10
	Management of Industrial Affairs	1
Faculty of Letters and Arts	Teacher librarian	8
	Library sciences	3
Faculty of Economics	Accounting expertise and audit	3
	Finance	1
Faculty of Law	Human rights	1
	European Law	1
Faculty of Sciences	Finance	1
	Applied Biology	1
Faculty of Social Sciences	Educational management	2
Faculty of theology	Practical theology	1
	Historical theology	1
Faculty of Medicine	Management of occupational healthcare	1
	Sanitary Management	1
Faculty of Agricultural Sciences, Food Industry and Environmental Protection	Management of modern food processing	1
	Quality assurance and food safety	1

Regarding the question "How often do you read?", most (14) replied every day, followed by 12 people who indicated a weekly frequency, which is great. From the 14 who read daily, 9 go to the Faculty of Letters and Arts, 3 to the Faculty of Engineering, 1 to the Faculty of Sciences, and 1 to the Faculty of Social Sciences. Only 1 person answered that they did not read and it's astonishing that they go to the Faculty of Letters and Arts.

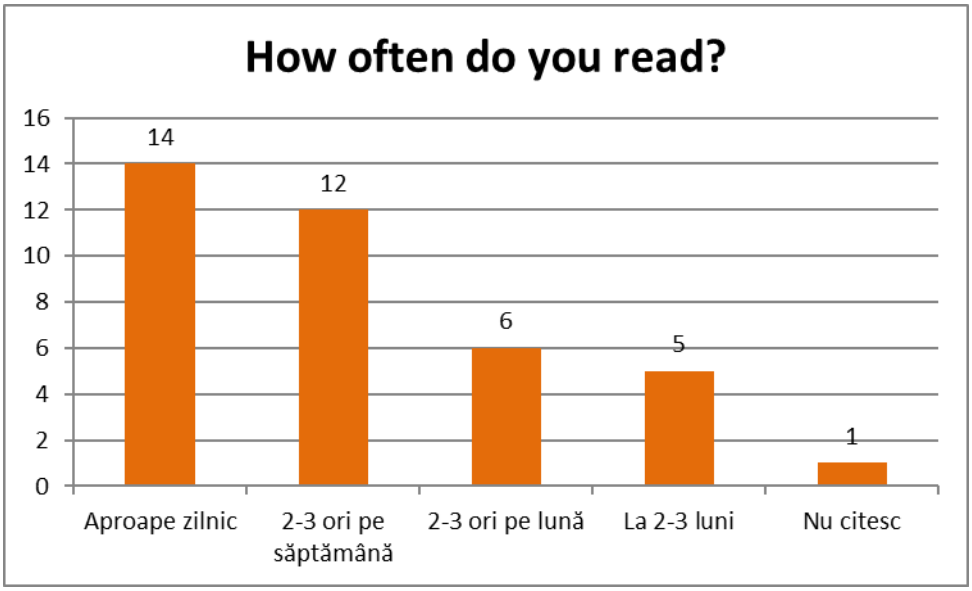


Figure 8 – How often do you read?

Regarding the last time they read a book, 17 people said that they had read during the week when they filled in the survey. To this question too, one person replied they had never read a book.

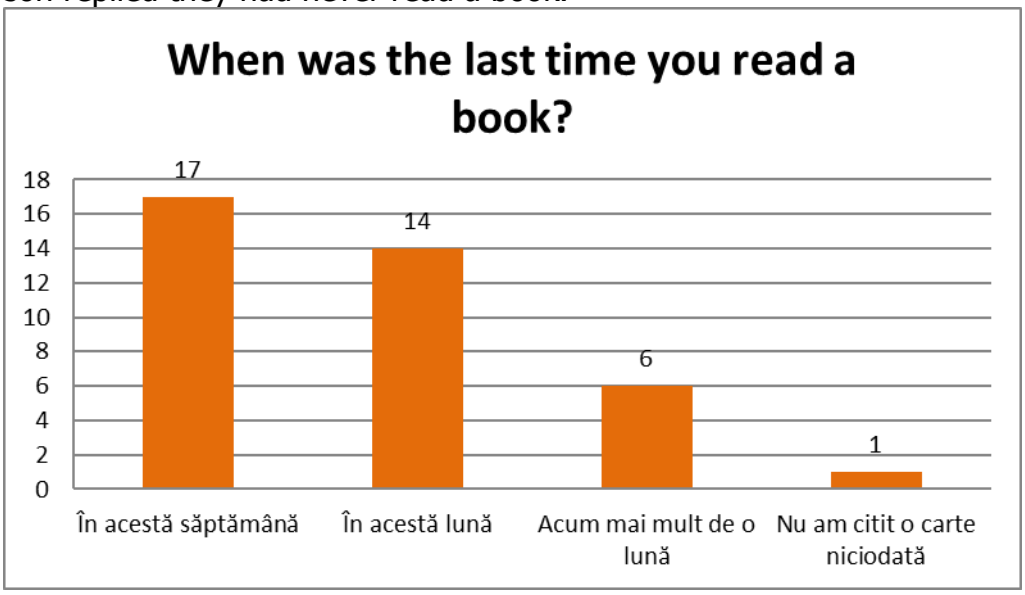


Figure 9 – When was the last time you read a book?

The following question asked respondents to rank the importance of reading from 1 to 5, with 1 being of little importance, and 5 very important. Twenty-nine MA students, i.e. 76.31%, chose 5. Nobody chose 1 and 2, which is gladdening, considering the times we are living.

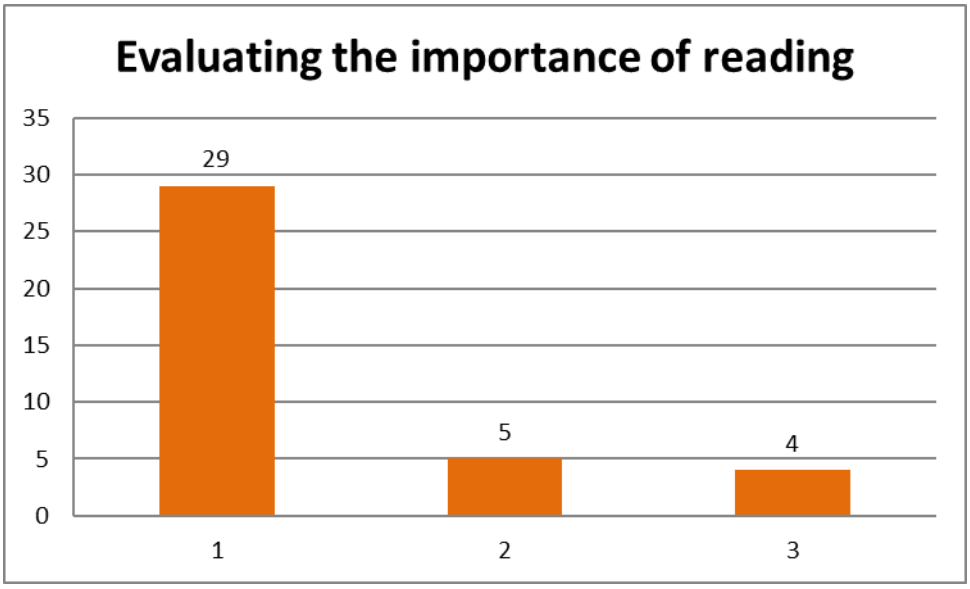


Figure 10 – Evaluating the importance of reading

In the next question, students chose the meaning of reading out of several options. This was a multiple-choice question, whose results are presented in the table below.

***Table 8*– Respondents' Master Major**

Is a way to forget one's problems	20
Is essential in the life of all people	21
Is a means to discover new things	26
Is an ideal way to spend one's free time	24
Has no benefits	0

The majority think that it is a way to acquire new information, followed by "it's an ideal past time"; we should note that nobody chose the answer "reading has no benefits".

This was followed by an open question, where respondents listed at least three reasons why reading was important. Their answers include:

It's relaxing, is a means to develop one's vocabulary, books are friends that never betray;

Personal development, finding inner peace, language improvement;

Develops one's memory, improves one's vocabulary and is a way to experience new things alongside each character;

Pleasure, balance, develops imagination;

Relaxation, personal development, initiation;

Improves general knowledge, develops vocabulary, relaxation;

Relaxes the mind, develops one's imagination, helps keep one's brain active, enriches the vocabulary of the reader.

Then, the MA students were asked whether reading "made one a richer person" or was "a waste of time", and luckily the first option was picked by 36 of the 38 students, while the second by only two of them. Considering the availability of online libraries, MA students were asked to choose between this and the classic library. Thirty of them prefer the classic library, compared to eight who prefer the online.

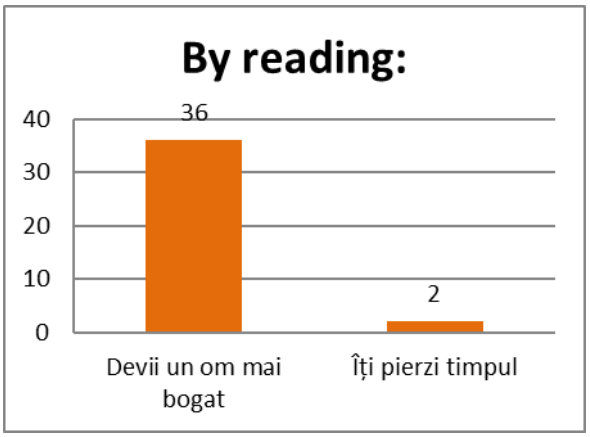


***Figure 11* – By reading..**

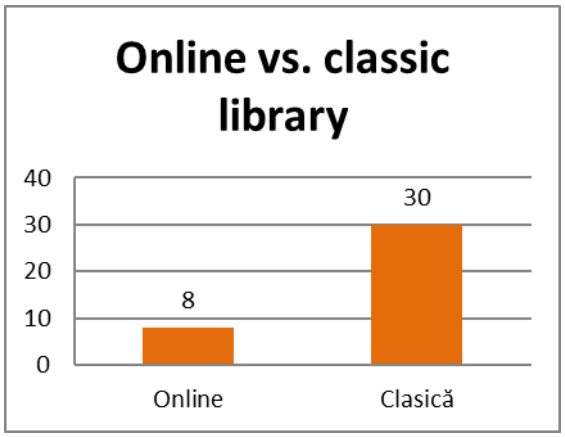


***Figure 12* – Online vs. classic library**

Regarding the importance of having a library at the work place, the results show that 19 MA students (50%) think it is important to have a library available at work, while 13 (34.3%) think it's very important. Nobody selected the option "not at all important", and two people (5.3%) said this aspect was of little importance.

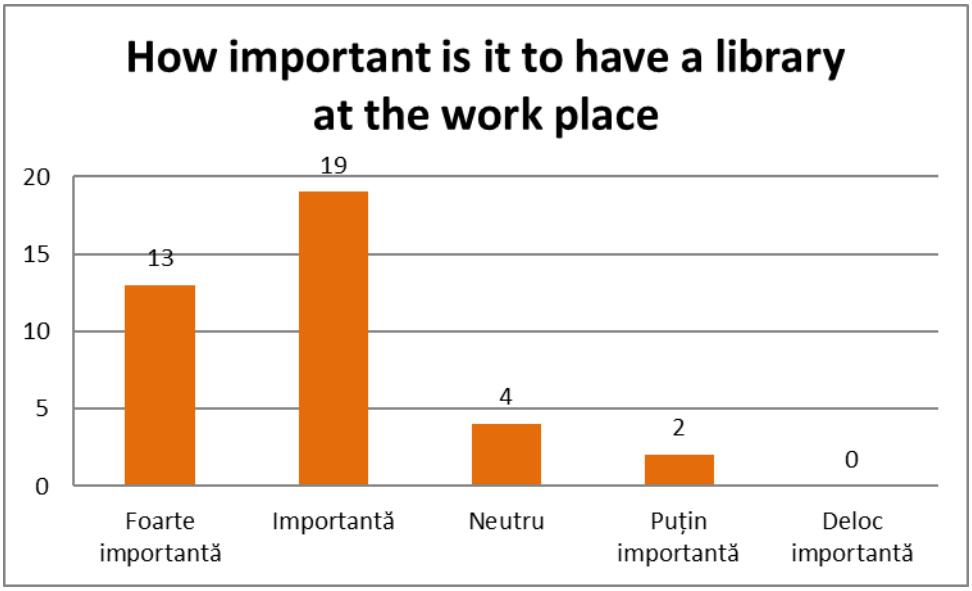


***Figure 13* – How important is it to have a library at the work place**

Moving on, the participants to the survey evaluated the relevance of six questions. The results illustrated in the table below show that most of them agreed with all the statements. These referred to the role of reading in developing one's vocabulary, to the time in one's life when their taste for reading develops, or to the preference for reading at night, before going to sleep.

***Table 9*– The benefits of reading**

	Not at all important	Less impor-tant	Neutral	Important	Very impor-tant
Reading is important for developing one's vocabulary	0	1	1	7	29
Reading is a means of relaxation	0	2	3	9	24
Reading stories to children at an early age sparks their interest for reading	0	1	2	7	28
Children who have a bookcase in their home will have a bigger passion for reading	0	2	7	10	19
One's taste for reading starts developing in their childhood	0	3	3	9	23
I like to read at night, before falling asleep	2	1	5	12	18

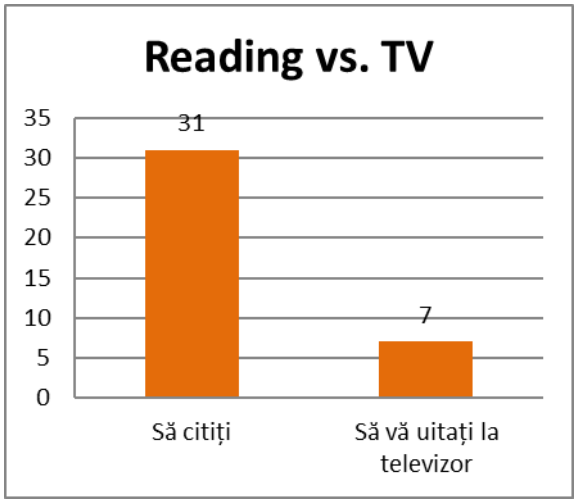
The option Not at all important was only chosen by two students in connection to the statement: "I like to read at night, before falling asleep".
Later on, the students were asked to note true or false the four statements by C.S. Lewis (British university professor and writer, whose works included fiction novels, works on medieval literature, etc.), V. Nabokov (who, among others, wrote "King, Queen, Knave", "The Eye", "Pnin". "Pale Fire", etc.), E. Hemingway (who wrote novels, short stories, prose and worked as war reporter), and G. Keillor (American author, humorist, story-teller, radio personality).

***Table 10*– True or false**

	True	False
"We read to know that we are not alone." C.S. Lewis	27	11
"Knowing you have something good to read before bed is among the most pleasurable of sensations." V. Nabokov	34	4
"There is no friend as loyal as a book" E. Hemingway	30	8
"A book is a gift you can open again and again." G. Keillor	37	1

Almost all students (37 out of 38) agreed with G. Keillor's statement: "A book is a gift you can open again and again." - G. Keillor. In the case of C.S. Lewis' statement, "We read to know that we are not alone", we find the biggest difference between answers (27 - true and 11 - false).

Between reading and watching TV, most respondents, 31 (81.6%), choose reading, and only seven watching TV. Moreover, 28 out of 38 MA students think reading a book is more absorbing than watching a movie about a book.



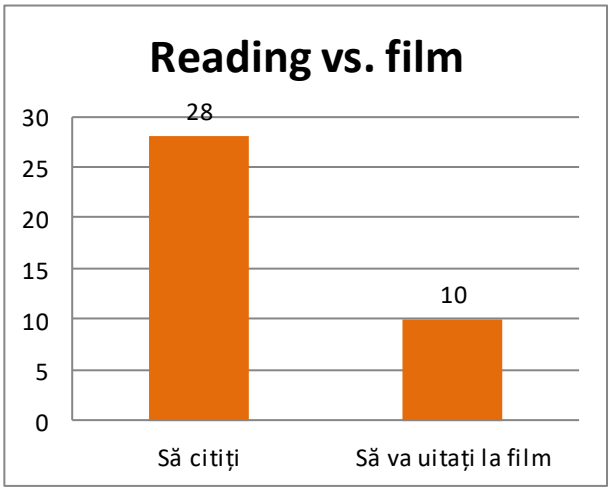


***Figure 14* – Reading vs. TV** ***Figure 15* – Reading vs. Film**

Regarding the need to include a reading class in the school curriculum, most respondents (36) agreed. The two MA students who didn't are going to a master's program at the Faculty of Engineering, i.e. Faculty of Agricultural Sciences, Food Industry and Environmental Protection.

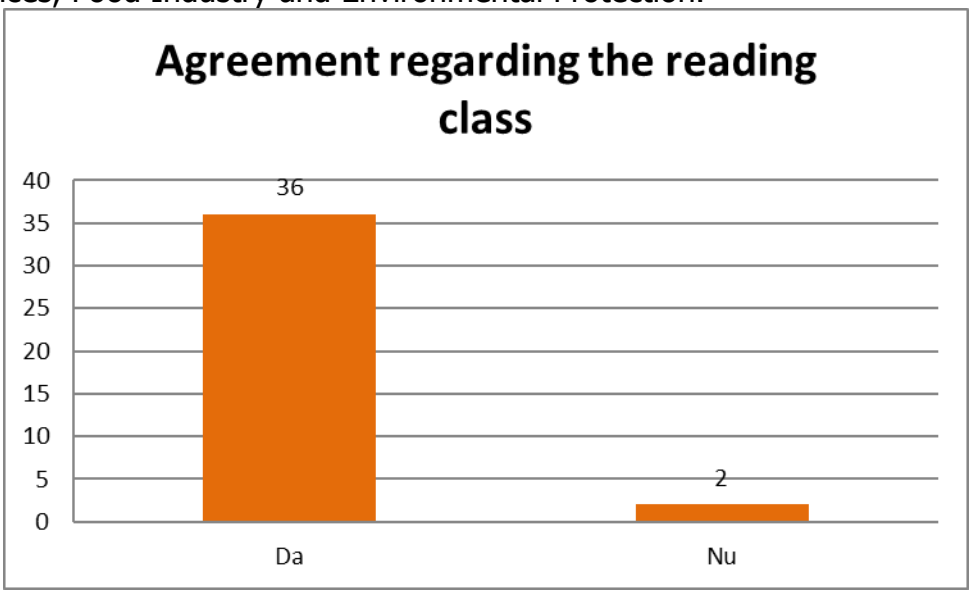


***Figure 16* – Introducing a reading class in the school curriculum**

Moreover, 24 students think it is very useful, and 10 that it is useful for children to take part in a reading circle. Three students are neutral, one thinks it is of little importance, and none thinks it's not at all useful.

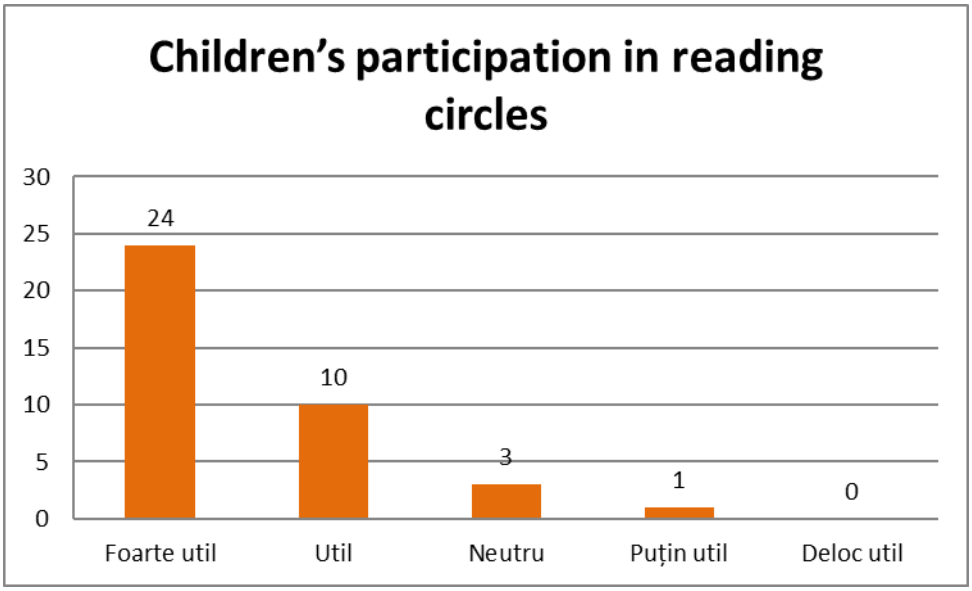


Figure 17 – Children's participation in reading circles

The last question invited students to present their opinions on stimulating children to read more. A complex answer was: "Offering interesting books for each age, offering books/stories topical for the times we're living and that are understandable at their age, and interpreting the knowledge and the symbols in that book/story in their own way, which also results in their desire to read more than one story/book", with which I completely agree. Another interesting answer was: "Through school challenges presented as games. They are assigned a book and the child who tells the most interesting things from the book receives an honorary title/a high mark/some kind of attention, like a diploma for instance. Or the teacher can assign words/key syntagms from that book and they are challenged to find them. This way, they will see reading like a journey." Other answers included:
Organizing attractive literary circles, with various relaxation and information activities;
Rewarding games;
See their parents read, too;
Introducing a reading programme in schools.
To conclude, it's clear that, regardless of the master's programme they go to, whether it's a humanities or a technical school, students still read and see reading as a useful, beneficial act. Despite the fact that, nowadays, technology is extremely advanced, that there are a plethora of ways to spend one's free time, reading is still preferred by young people, which confers hope that future generations will read too, whether classic or electronic books.

4 CONCLUSIONS

The impact of new technologies on the contemporary society has created a need to adapt to the changes that have engulfed all the fields of knowledge, especially reading.

In this sense, in conducting this work, we thought it was useful to approach this topic, in an attempt to dissipate, or at least mitigate one of the major problems facing the new generation, i.e. the decline in their interest in reading, to exploit the major importance of the Research and Information Centre in promoting reading and, at the same time, prove that it's absolutely necessary for all educational factors, in general, and for teachers, especially, to try to continuously look for and develop new methods to instill the joy of reading to youngsters.

We thus captured the importance of reading, noticed that pupil's declining interest in reading can be worrisome, yet at the same time, the situation can improve if efforts in this sense are organized with a sense of professionalism, responsibility, and enthusiasm.

Consequently, we believe that our objective, i.e. revealing *how* and *if* it is possible to educate pupil's reading taste, was accomplished, both by approaching certain necessary aspects in this sense, and by showing the activities within the targeted RICs.

We tend to believe that the information in our work will be useful to all the readers with an interest in our topic, and the activities discussed will be a real, positive example of the influence of reading on pupils.

We hope that, in the future, reading will hold a higher and higher share in school curriculums, regardless of the age level, as reading must be, for both present and future generations, the foundation on which, in time, they will build further options of methods to expand their cultural horizon.

References

[1] ANCUȚA, Partenie. Modalitați de stimulare a lecturii la elevi. Timișoara, Excelsior Publishing House Art, 2007.
[2] CORNEA, Paul. Introducere în teoria lecturii. Second Edition, Polirom Publishing House, 1998.
[3] COSTEA, Octavia. Didactica lecturii. O abordare funcțională. Iași, European Institute, 2006.
[4] HAREA, Gabriela Colaborator bibliotecar-cadre didactice: factor de stimulare a lecturii elevilor. In: Biblioteca. Revista de bibliologie și știința informării, year XIX, new series, 2008, no. 11-12.
[5] MURĂRAȘU, D., SIMONESCU, Dan. Lectura particulară și biblioteca școlară. Bucharest, 1939.
[6] NICU, Adriana. Strategii de formare a gândirii critice. Bucharest, Didactică și pedagogică Publishing House, R.A., 2007.
[7] POPESCU, Cristina. Cititor – bibliotecar – bibliotecă. In: Biblioteca. Revista de bibliologie și știința informării, year XIV (LV), new series, no. 8/2003.
[8] POPTĂMAȘ, Dimitrie. Reflecții despre carte, bibliotecă și lectură. Târgu-Mureș, Nico Publishing House, 2009.
[9] REGNEALĂ, Mircea. Vocabular de biblioteconomie și Știința informării, vol. I-II: Bucharest, ABIR, 1995-1996.
[10] SÂRGHIE, Octavian, Mihai. Lectura – o experiență fara sfârșit. In: Biblioteca. Revista de bibliologie și știința informării, year XXIII (LXIV), NEW SERIES, no. 1/2012.
[11] ȘINCAN, Eugenia. Îndrumător pentru învățători, părinți și copii. Craiova, Scrierea pentru dezvoltarea gândirii critice Publishing House, Volume I, 1998.
[12] STEELE, Jeannie L., MEREDITH, Kurtis S., TEMPLE, Charles. Lectura și scrierea pentru dezvoltarea gândirii critice, Volume I, 1998.

[13] STEELE, Jeannie L., MEREDITH, Kurtis S., TEMPLE, Charles. Lectura și scrierea pentru dezvoltarea gândirii critice, Volume II, 1998.
[14] Ministry of Education, Research, Youth and Sports (MECTS): www.edu.ro/cdi.
[15] Rules for Teacher librarians, available on the official website of MECTS www.edu.ro/cdi.
[16] ARGATU, Daniela. Dezvoltarea competenței de lectură în biblioteca școlară. Article available at: https://www.ccd-suceava.ro/pdf/biblioteca/competenta%20de%20lectura.pdf.
[17] BĂDESCU, Cezar-Paul. Pastila de cultură: de ce să citim literatură? Article available at:https://adevarul.ro/pastila_de_weekend/Pastila_de_ciltura_De_ce_sa_citim_literatura_0_576542919.html.
[18] CÂNTEC, Oltița. De ce să nu mai citim. Article available at:https://evenimentul.ro/articol/de-ce-sa-nu-mai-citim.html.
[19] CIUCĂ, Cristina. Strategii pentru educarea gustului pentru lectură în rândul elevilor. Article available at: https://www.asociatia-profesorilor.ro/Educarea-gustului-pentru-lectura.html.
[20] FEȘTILĂ, Constantin Bogdan. Povestea înființării unui Centru de documentare și informare. Article available at: https://sites.google.com/site/cdifalticeni/proiectulcdi.
[21] MOLDOVAN, Adela. Tipuri de lectură. Article available at: https://lectura.pasiunea.eu/tipuri-de-lectura_articol_2713.html.
[22] PĂICUȚ, Laura, VIDICAN, Delia, BUHA, Lavinia. Documentare – informare profesională. Revista profesorilor documentariști. Article available at: https://www.scribd.com/document/31534609/Numarul-2-Aprilie-2010-REVISTA-CDI.
[23] RADU, Corina. Despre lectură și rolul cititorului în devenirea unei opere literare. Article available at: https://lectura.pasiunea.eu/despre-lectura-si-rolul-cititorului-in-devenirea-unei-opere-literare_articol_49.html.
[24] SEBASTIAN, Alexandru. Tehnici de documentare și informare – CDS – profilul bibliotecar-documentarist. Article available at: https://bibliotecari.blogspot.com/2006/10/centre-de-documentare-si-informare.html.
[25] http://lewebpedagogique.com/flegbadau/initiere-in-cercetarea-documentara-partea-i/2-centrul-de-documentare-si-informare-cdi/22-organizarea-si-functionarea-unui-cdi/221-regulamentul-de-organizare-si-functionare/,
[26] Fédération des associations des professeurs documentalistes de l'Éducation nationale, French Ministry of National Education https://data.bnf.fr/
[27] Françoise Chapron, Les CDI (Centre de Documentation et d'Information) des Lycees et College en France: Quelle Contribution a la Formation des Citoyen?, available at: http://archive.ifla.org/IV/ifla63/63chaf.htm
[28] Dorina Dolha, Principiile pedagogiei Freinet, available at: www.aiarad.ro/pdf/publicatii/quovadis_nr4_2004.pdf,
[29] Cury A., Brilliant Parents, Fascinating Teachers, Bucharest, For You Publishing House, 2005, p.20

DOI: 10.2478/ijasitels-2020-0006

Searching, Learning, Gaming - Engaging students with Europeana's digital archives

Cristina Ioana ROIU [1],

[1]*Library of the Romanian Academy of Sciences, Romania*
crissroiu@gmail.com

Abstract

Mass digitisation of the collections held by cultural institutions -galleries, libraries, museums, archives, have made available a huge amount of historical, cultural, informational resources in digital format, which are more and more used in educational activities. This paper describes some innovative non- formal educational activities developed around the Europeana1914-1918 archive between 2014-2018, where searching in big digital archives, gaming and improving the digital skills were key aspects.

Keywords: digital archives, non-formal education, gamification, digital skills

1 Education in big digital archives times -general context

In the last decades GLAM institutions worldwide have massively digitised their' collections to increase access to these collections, to offer multiple services and to engage audiences with this digital content.

Therefore, at present time a plethora of cultural digital resources is available to be used by cultural lovers, professionals, researchers, people from creative sectors or education.

From the perspective of students and teachers, it is well known that cultural heritage has a crucial role to play in education as digital collections can provide multiple perspectives on historical, scientific**,** and cultural developments across Europe and beyond.

In order to bring these sources of knowledge to every learner, educator or educational organisation and to explore the pedagogical potential of the digital cultural collections, GLAM institutions(galleries, libraries, archives, museums) have been continuously developing educational activities in collaboration with the educational sector.

The main challenges when dealing with such a huge number of digital items are usually associated with:

- What collections should be used?
- How to use them in education in order to maximize the impact of educational activities?
- Which new patterns and models in education are needed?

2 Europeana and the educational uses of its' collections

With its' more than 57 million digital items, Europeana - Europe's digital platform for cultural heritage is one of the major players within the digital cultural heritage sector worldwide.

Trying to achieve its' main aim -to empower the cultural heritage sector in its digital transformation, Europeana develops expertise, tools and policies to embrace digital change and encourage partnerships that foster innovation.

But in the first place, Europeana's activities and projects make it easier for people to use digital cultural heritage for education, research, creation and recreation.

During the last 10 years, Europeana's projects and digital collections involved multiple innovative educational activities and enabled successful collaborations between sectors like GLAMs and education (or creative industries and education).

Educators in Europe and worldwide frequently use Europeana's digital collections as they come from trustful sources and many items are free to reuse for educational purposes.

3 The Europeana 1914-1918 project and its' educational activities in Romania

Prior to the centenary of the outbreak of the Great War and all the way through the end of 2018 we have witnessed a great interest and a rediscovery of the First World War times everywhere in Europe and worldwide.
Media and editorial productions, exhibitions and thousands of cultural and historical events and projects were dedicated to describing this historical event that changed the whole world in a dramatic way.

Within this plethora of projects and events dedicated to WWI the Europeana 1914-1918 (1) project was special, as it brought a different approach to describing the Great War times and to cultural history, linking people's own stories to the official histories of the war.

Europeana 1914-1918, which has become with the years the most successful Europeana project in terms of public engagement - has been launched in 2011 with the first Community Collecting Day held in Germany.

Originated in a 2006 University of Oxford's idea, the project aimed at collecting letters, family objects, photos and other memorabilia dating from WWI time privately held by Europeans in order to be digitised and then added to the Digital Archive Europeana 1914-1918.

During the time, this WWI Europeans'Archive gathered more than 10,000 interesting family stories most of them unpublished before and more than 100,000 digitised pages in text, image or audio format.

By bringing together in January 2014 this dataset with two other digital archives built during the Europeana 1914-1918 Collections Project (400 000 WWI items digitized from the collections of some of the most important European libraries) and Europeana EFG 1914 (collecting in digital format 660 hours of films and audio-visual materials from WWI times) it has been created the greatest digital Archive dedicated to WWI worldwide).

3.1 Europeana 1914-1918 and the 'School Otherwise Week'Workshops 2014

Even from the beginning of the Europeana 1914-1918 project major institutions like British Library tried to create educational materials using this impressive digital archive. In this way, the British Library's educational website (2) designed in co-operation with history teachers offers very interesting learning resources dedicated to WWI .

The team who coordinated the first Collection Days in Romania has also organised in April 2014 in Bucharest a series of workshops and meetings regarding World War One and Europeana 1914-1918 Archive during the 'School Otherwise Week's non- formal educational program.

At these activities a lot of students and pupils took part, where the 'Ion Neculce', 'Gheorghe Lazăr', 'Ion Luca Caragiale' National Colleges and the 'Titu Maiorescu' Secondary School have been involved [3].

Searching together with their' teachers through the Romanian documents of Europeana1914-1918 archive ,but using also the English and French digitized

memorabilia, the students from 9 –12 grades have tried to recreate the atmosphere of those times and select some stories or characters that impressed them.

They have written essays about these stories or characters and produced Power Point presentations which were presented to their' colleagues during the workshops.

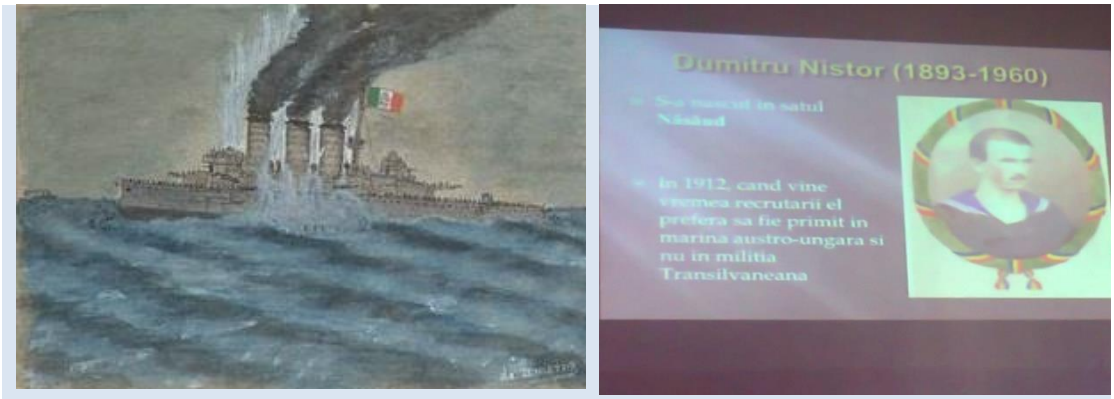


Figure 1. Images from the ppt presentation of Dumitru Nistor's life

The students tried to describe not only the historical perspective on the events but also to emphasize the human reactions to those events. In this way, the subject of their essays focused on some important historical characters, personal stories or some interesting topics (4) :

- The Story of Dumitru Nistor
- Queen Maria of Romania and the First World War
- Artists in wartime
- Children in wartime
- The moral support of the home families for the soldiers
- Medals and medals recipients in The Great War
- The story of Iuliu Bologa
- A story about the Romanian Scouts

Some conclusions after these workshops:
The students loved to learn about WWI from personal diaries or documents: these memorabilia have the power of a testimony, there is a visual impact and also the storyteller's personal feelings' impact.

The students improved their ability to search in a big digital archive looking for multiple digital formats: text, image, audio, audio-visual and multilinguistic material.

They have also improved their English and/ or French knowledges and learned how to make a good Power-Point presentation.

The students' abilities to create links between different stories and topics and to place an event within a global context have also been improved

From the teachers' point of view, the Europeana 1914-1918 archive contains first hand resources of exceptional documentary value, because they come from the descendants of the people directly involved in the war like the officers, the soldiers and their families.
A History teacher from the 'Ion Neculce' National College described the Europeana 1914-1918 Archive from the historian but also from the teacher point of view:
'It is all about a lived history, where scientific information goes hand in hand with sensitivity, the sense with the sensibility.
Every piece of historical information is filtered by the personal experience of the storyteller so that objective and subjective, general and particular are melting together in a multifaceted view.
The perspectives on the events are dramatic, heroic and tragic. War has created special and unusual situations at which people reacted in unexpected ways. Values and attitudes like humanitarianism, dignity, tolerance, critical thinking are to be remembered above all.'

3.2 Europeana 1914-1918 and Gaming-type activities for students- 2018

Four years after the above described workshops, the Romanian Europeana 1914-1918 team and students and teachers from the 'Ion Neculce' College in Bucharest have developed together other interesting activities , in the special context of the 2018 year: marking 100 years since the end of WWI and the Great Union of Romania.

The series of workshops were intended to present important facts about WWI and the importance of the year 1918 for Romania in innovative ways using mainly game type activities.
In the first workshop students learned how to use the Kahoot application on their phones and then watched a series of videos highlighting important WWI events and context.
A competition tested their First World War related knowledge using the Kahoot application: questions and answers alternatives were displayed on a shared screen while students answered on their devices.
The students firstly divided into three work groups answered the questions related to the WWI theme during the time allocated by the Kahoot application. At the next stage they competed individually, the whole event being extremely appreciated by the students.

The presentation of the Europeana1914-1918 digital Archive -mainly the Romanian held items has concluded the workshop.

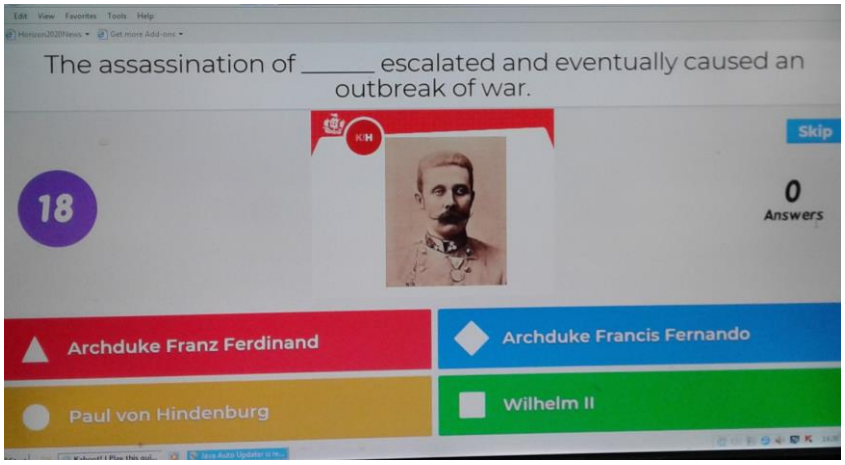


Figure 2. A Kahoot Question from the competition

It is well known that Kahoot! is now used by many teachers to present the content in a more engaging way, to deepen understanding, increase class participation and improve how much information students actually remember. By combining slides with quiz and poll questions in a kahoot, one can provide additional explanation and context to a topic or a new piece of material.

All in all students enjoyed this interactive session where the already held WWI knowledge was completed by new data about Romania's situation at the end of the War, the situation on the front, the movement of the troops and the situation of the families left at home. The materials used - diary pages, photographs, stories and family documents as well as films from the First World War time had a special impact on students.

The Second Workshop –Searching Europeana1914-1918 archive and Creating GIFs
With the launch of the famous GIF IT UP competition in October 2018 the students and teachers of the 'Ion Neculce' College and the Romanian Europeana team had the opportunity to collaborate again.

GIF IT UP - an annual online competition, organised by Europeana, Digital Public Library of America, Trove and DigitalNZ, challenges people to use digitised cultural heritage material to produce unique GIFs (animated images) and share them online.
GIF IT UP is also supported by GIPHY , the world's largest library of animated GIFs.

Participants must use the openly licensed paintings, photography and texts digitised in Europeana Collections and the digital libraries of the other organisers to create, submit, and share striking new GIFs. While the competition has an appeal to all culture lovers, encouraging students to make GIFs in educational contexts offers young people an opportunity to animate and engage with cultural heritage.

The 2018 GIF IT UP edition had a special section dedicated to World War One to mark the end of The Great War.

In this context, during the second workshop students learned to produce GIFs and also learned about copyright and what images can be used from the digital collections of the organizing institutions.

The practical exercises offered the opportunity to search again in Europeana 1914-1918 image archive and select interesting images to produce GIFs.
In this complex process of browsing the digital collections, students also learned a lot about historical events and the general WWI context and have improved their knowledge of foreign languages while also having fun.

Many of them submitted their creations online and felt the responsibility and pride of taking part in a famous online competition.
One of the students received the public choice award for her GIF in the 2018 competition.

Figure 3. Anamaria Şune's winning GIF of the GIF IT UP 2018 People's Choice Award
https://giphy.com/gifs/gifitup-gifitup2018-XHKvHPrpfCcn4x5az3

Following this workshop, these kind of activities were considered an excellent form of engaging young people with the digital cultural and historical content of Europeana.

4 Conclusions

The potential for the use of digital cultural heritage in education is widely acknowledged, but it doesn't happen just by providing access to the material. Sources need to be selected, contextualised, and (crucially) become part of learning activities for students.
Big archives like Europeana 1914-1918 being pan-European, can offer a global image on the events allowing the students to see them in a larger context and to better understand the traumatic experiences faced by all Europeans during World War One no matter which side their governments were on during the conflict.
The informational resources are trustful and comprehensive offering a multilingual, multiformat and complex digital archive where particularly the image and audio-visual resources are very attractive for students.
The activities associated with students' personal searching in Europeana1914-1918 are a good example of integration of exploratory approaches to learning, strengthening students creativity and retention.

New technologies offer multiple opportunities to engage student learning in new ways, one of these potentials for motivating students to learn being gamification (or game-based learning).
The growing popularity of gamification in education resides in its potential to foster motivation, behavioural changes, friendly competition and collaboration in different contexts [5].
Gamification in the above presented cases is associated more with gamefully experiences rather than introduction of game design elements in a systematic design of learning processes.
Raising interest for such an important topic like the First World War, followed by an engaged attitude from the students was one of the major successes of these activities.

During the above described activities students learned nor only important historical aspects, but they understood what copyright is, have improved their digital skills and the ability to successfully search in big archives.
Adaptability and interdisciplinary approaches to learning are keys in the new eLearning environment and GLAMs with their (now) impressive digital collections have an important role to play in shaping new patterns and new models in education.

References

[1] Pro Europeana.Europeana1914-1918. 2012 [Online]. [Accessed 29 March 2019]. Available at: http://pro.europeana.eu/project/europeana1914-1918
[2] British Library..World War One. [Online]. [Accessed 29March 2019]. Available at: http://www.bl.uk/world-war-one/teaching-resources

[3] Pro Europeana, Romanian Students engage with Europeana WWI content (Online]. [Accessed 29March 2019]. https://pro.europeana.eu/post/romanian-students-engage-with-europeana-wwi-content
[4] Biblioteca Academiei Romane .Buletin Informativ nr.10 , An III. [Online]. [Accessed 29 March 2019]. Available at: http://www.biblacad.ro/buletine%20informative/2014/BINF10.pdf
[5] Dichev, C., Dicheva, D. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. Int J Educ Technol High Educ 14, 9 (2017).

DOI: 10.2478/ijasitels-2020-0004

Studying big data using virtual escape rooms

Solomovich Lior [1,2]

[1] *"Kaye College", [2] "Sapir College", Education in the Age of Information Technology Department, Israel*

solomovi@gmail.com

Abstract

The Corona period created great challenges in the teaching and learning process. This presentation describes a research based on studying the subject of big data using virtual escape rooms. Due to the period it was decided that the final task in the Big Data course would be to build a virtual escape room that deals with a topic from the world of Big Data. The aim of the study was to examine how the understanding of various topics related to Big Data is sharpened when they are taught through virtual escape rooms. Each group of students chose a topic from the world of big data and built a virtual escape room around it. Peer learning was then performed, and each group edited rooms of other groups, so that each group learned about a variety of topics related to the big data world. By monitoring peer learning it was possible to assess the nature of the learning process experienced by the group. This unique learning allows for flexibility in terms of time and place in holding meetings. The interactions between the various group members contribute to enrichment the knowledge and the development of creative ideas.

According to the results of the study, this type of learning sharpened different skills among the participants - social, communicative and thinking. It also developed a deep understanding when the escape room practically demonstrates the topic of big data in an optimal way through a variety of representations. During the activity in the escape room there is created an interesting and original connection between the materials and the ideas, open to a variety of interpretations. Digital tools are used effectively and creatively that contribute to understanding the message (Kemp, 2018).

Keywords: big data, virtual escape room, 21st century skills

1 Introduction.

As part of the Big Data course taught by students in the technology marketing at Sapir College in Israel, the students developed virtual escape rooms. First the students were divided into groups of four and then they chose the topic they wanted to explore, defined the puzzles for the escape room and designed the virtual environment. Evaluation of the escape rooms was done through peer learning when each group was asked to evaluate the escape rooms of the other

members. In this way all the students learned about other topics related to the world of big data.
The results were made through the qualitative approach. In order to understand the experience, 15 interviews were conducted with students who participated in the course. Through interviews, data were collected with the aim of finding out how the students who participated felt about the process they experienced and whether the original way in which they learned contributed to them understanding the subject of big data.
Online teaching requires to think about teaching and evaluation and plan the learning sequence, develop e-learning materials and provide a technical, pedagogical and administrative support system and design the learners' unique learning environment [2].
The game "Escape Room on the subject of Big Data" can be used as an innovative pedagogical tool that allows a deep understanding of the subject of Big Data. First, I gave some theoretical background on big data and then I focused on helping students using virtual escape rooms to study the subject. The themes of the escape rooms are very diverse, and the challenge levels are varied. and the activity includes a frame story that the participants become acquainted with the subject before entering the room.
The idea of learning through play is not new, but it gets better over the years and even expands into higher education. Incorporating play into learning contributes to the motivation, enjoyment and interest of learners as well as of teachers. games help to develop communication skills and social rules like patience and perseverance. When people play physical or online games, they get involved, get excited, take responsibility and initiative, solve problems, risk asking questions and mistakes and get instant feedback and a chance to correct.
There are very few research literatures on escape room as a teaching pedagogy. From studies conducted so far among medical students [3] and students of physics [9], escape rooms were found to be an effective teaching pedagogy both in the academic aspect and in the motivational aspect. In addition, university faculty members in the UK who have been exposed to this pedagogy have reported that escape room activity is innovative, evokes interest and enjoyment and has educational value. They estimated that they applied this pedagogy in their teaching [4].

2 Escape room

2.1 What is escape room?

Escape Room is a game where a group of participants is locked in a room and has to find their way out by solving puzzles and clues scattered around the room. Escape rooms, as a challenging leisure activity, were inspired by computer games, and in less than a decade, have become popular in many countries [6]. The game is characterized by a frame story that has a dimension

of mystery, a defined goal and there are time / space constraints for each encounter. There is need for teamwork to succeed in getting out. A wide range of challenges of various kinds, hints that allowed to advance the game when needed [7]

2.2 Why to teach with escape room

Escape rooms are an expression of innovative, active, collaborative and constructivist teaching approaches that can shape learning in the best way, much more than conventional teaching. They help learners understand the value of observing problems from different perspectives, exposing them to collaborative teamwork, promoting engagement and perseverance in the task, strengthening social ties, encouraging teamwork and deepening learning through group discussion. The students' need to work together for succeeding in a task under the pressure of time allows them to develop communication skills. Moreover, escape rooms allow for the integration of technology in the classroom, as they make it possible to easily integrate websites, videos or other interactive digital material within the various puzzles of escape activities. However, the process of creating puzzles that focus on learning objectives and obligate students to engage with content from the curriculum instead of just looking for clues requires time and thought. Therefore, it is important to create quality games for the long term. Thus, after the development of the game, it will be possible to use it again and again in subsequent years. Before any application of an escape room study it is important to perform a test in order to assess the time required to complete the game and identify any error that may prevent the success of the participants [12; 6]

2.3 Virtual escape room

In virtual escape rooms we can play on computers or as apps on mobile devices. In the digital escape room, there may be various changes and adjustments that contain various elements taken from the physical escape room such as locks, magnets, cryptography and keys. In order to enhance the experience, there is used augmented reality to. The games in the virtual escape room can be available to a single player or to a team [10].
When developing a learning-oriented virtual escape room there are four issues to consider:

- Learning objectives: These objectives aim to process the topic and evaluate the student's learning experience and achievements and find out accordingly which indicators of change that must be made in order to improve.
- Single or multidisciplinary theme: single domain or multiple disciplines presented as part of the game experience.
- Soft Skills: Interactive action games that can help develop soft skills such as communication and teamwork.
- Troubleshooting: Developing troubleshooting challenges to make the game experience interesting for players.

2.4 How can the learning experience with virtual escape room can contribute to the 21st century skills?

Traditional learning supplies very few opportunities for the development of 21st century skills that include, among others, Communication, Collaboration, Critical Thinking and Creativity [1]. Learning in the "escape rooms" method makes it possible to cultivate each of these four skills through the activity.

An escape room implements the innovative teaching style characterized by a lecturer serving as an active learner's guide as opposed to a teacher who imparts knowledge to the passive learner; thus the lecturer role changes and he also become fertile cushion for game-based learning and 21st century learning skills.

2.5 Why to study "big data" with escape room

The term "big data" consists of a wide range of topics around them students created the virtual escape room. Topics such as military and police intelligence, biology (Outbreaks of pandemic), commerce, cyber warfare, digital marketing, indexing and textual information search, gambling, sports, industry. Students have made escape room about all this topic. The connection between a virtual escape room and the study of various topics related to the world of big data has allowed students a large space of exploration and creativity.

2.6 Representative example of Virtual escape room

Here is an example from the various rooms and topics the students made. The main theme of the group the was prediction of pandemic outbreaks.

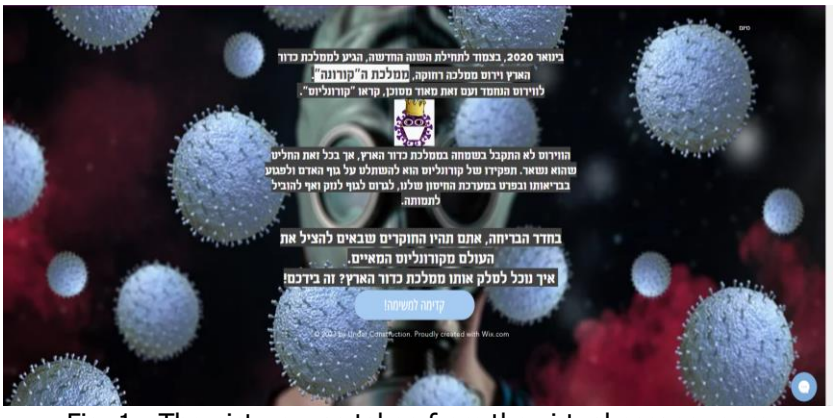


Fig. 1 - The picture was taken from the virtual escape room

The students learned about the connection between Big Data and pandemics - how huge data can allow us to predict where the epidemic is about to break out. After giving tasks, reading articles, researching and delving deeper into the subject. they prepared an escape room based on graphs, tables, pictures and widgets. The participants in the escape room are not expected to do analyzes

but rather to answer questions in the escape room aim to test understanding and insights.

In contrast to the "tasks" given in the lesson, the challenge in an escape room is to create a sequence between the clues. Each clue has two functions: one is to lead to the next riddle. The other, which is no less important, is to lead to a stage that promotes the learner to achieve the goal of the lesson.

The frame story was: you are a secret agent who sent abroad to find the vaccine against a virus that is spreading in the world. It is known that the vaccine already exists. You must go through different countries, where you will be referred by codes and clues in order to reach the vaccine.

Here are few examples for some of the riddles written by the group members.
1. You have just landed in Barcelona; you receive a message from the commander directing you to a file with 2 tickets - to Iceland and Bangladesh. And in addition, statistics on these countries. Which place will you not fly to? According to the number of tests per 1000 people and based on the graph, where is it safe not to fly?
Hint: need to check how many people live in Iceland and Bangladesh and how many tests they do through https://ourworldindata.org/grapher/population-density

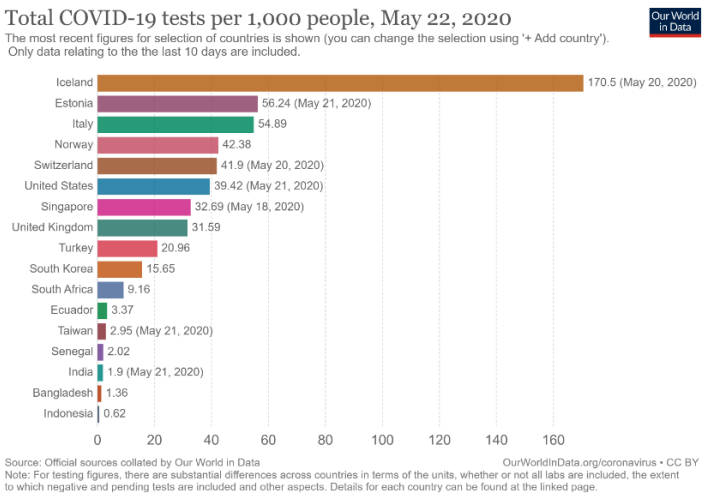


Fig. 2 - The graph was taken from the virtual escape room

2. Invented the vaccine at university with a reputation score of 71.3. https://www.statista.com/statistics/226686/world-university-rankings-by-reputation/
What country are you currently flying to, to get the vaccine?

3. On the day that the number of people who recovered reaches 1,500,000, you will be able to return to Israel. please use the website https://hgis.uw.edu/virus/
Hint: Use the cursor on the statistics site to cross-reference data

3 Results

Each group of students included 4 students. All in total were 15 groups. After they finished, they participated in other students escape room and gave an anonymous feedback to their friends. To carry out the research the process was constantly reshaped thorough a brain storming with the students.

The methodological approach was Qualitative. When they finished participating in the escape room, I had Semi-structured interview containing 20 questions. After that I had 15 interviews with representatives of each group of the students which describe the experience they had.

The analysis of the data was collected from the interviews indicated on 3 themes which were very dominant sense of involvement, innovation, collaboration through the learning process.

The student's respondents during the interviews were very positive toward the process they all had. One of the responds said "The tool is very special, different, innovative and interesting. preparing the escape room was very challenging for me and I enjoyed every moment. Throughout the room preparation I learned about the subject, after that I thought about how I could get the participants who would come into my room to study The subject of autonomous cars in a creative way" another respond was "All the tasks I built for my escape room on big data could have also appeared as questions in the book but the multimedia options that exist on the net gave me the opportunity to be creative and have the learning process more interesting and fun" another responds said "In my opinion the escape room contributes greatly to the motivation of the students. It made us learn willingly, eagerly and with pleasure". One of the resonates emphasize the difficulties she overcame thanks to the group support "I did not just learn how to make the room. I worked with the trial and error method. There are still things I would like to do, working in a group helped and supported".

The application of an escape room, as part of the teaching big data, is an advanced tool adapted to learning in the 21st century and promotes important skills for this period, such as: collaboration in a teamwork, creative thinking, digital literacy, interpersonal communication and higher order thinking. Exposure to an escape room as a pedagogical tool managed to invoke a new wave of teamwork, different thinking, high motivation and great enjoyment from the preparation and implementation of the escape rooms. The sequence of leading puzzles from one to the other, the clock running on the screen and the competition between the groups motivated all the students to work hard and learn without realizing like they had done it. Looking at this activity, it seems that an escape room as a pedagogical method of teaching incorporated a central element of "Game Based Learning", a kind of learning that is characterized by the accessibility of the study material to the game. This learning has some basic components, for example fun driving motivation,

experiential and interactive learning that is rewarded with immediate feedback and independence of the learner but also includes teamwork [12].

Despite the students' concerns, the design and performance of the escape rooms were accompanied by great creativity and enthusiasm. Their ideas had been realized and in a challenging and joint process they set up creative, high-quality escape rooms that lead to learning on a wide range of topics. Along with the much planning and thought in preparing the room, all the students testified the great satisfaction and enjoyment they felt in the way of this teaching and in the great enjoyment of their students. They saw both man and women students interested, enthusiastic, learning while playing and cooperating. The variety of clues and puzzles allowed all students to experience a unique and meaningful experience and learn about different aspects of big data in a creative way. Nevertheless, this teaching method is still in its primary form. There is still a long way in front of us to test and evaluate the effectiveness of the method among students by more research.

4 Conclusions

In conclusion, creating a game by the students creates a sense of involvement in their learning. It enables students to discover their independence, produce sketches and test themselves, and fosters collaborations. It also invites students to learn the field of visual literacy through experience.

References

[1] Ananiadou, Katerina, and Magdalean Claro. "21st century skills and competences for new millennium learners in OECD countries." (2009).
[2] Boettcher, J. V., & Conrad, R. M. (2016). "The online teaching survival guide: Simple and practical pedagogical tips". John Wiley & Sons.
[3] Boysen-Osborn, Megan, Sara Paradise, and Jeffrey Suchard. "The toxiscape hunt: An escape room-scavenger hunt for toxicology education." *Journal of Education and Teaching in Emergency Medicine* 3.1 (2018).
[4] Clarke, S., Peel, D. J., Arnab, S., Morini, L., Keegan, H., & Wood, O. (2017). escapED: a framework for creating educational escape rooms and Interactive Games For Higher/Further Education. International Journal of Serious Games, 4(3), 73-86.
[5] Dwyer, C. P., M. J. Hogan, and I. Stewart. "An integrated critical thinking framework for the 21st century. Thinking Skills and Creativity, 12, 43–52." (2014).
[6] Abdul Jabbar, Azita Iliya, and Patrick Felicia. "Gameplay engagement and learning in game-based learning: A systematic review." *Review of educational research* 85.4 (2015): 740-779.
[7] Nicholson, Scott. "Creating engaging escape rooms for the classroom." *Childhood Education* 94.1 (2018): 44-49.
[8] Qian, Meihua, and Karen R. Clark. "Game-based Learning and 21st century skills: A review of recent research." *Computers in Human Behavior* 63 (2016): 50-58.
[9] Vörös, Alpár István Vita, and Zsuzsa Sárközi. "Physics escape room as an educational tool." *AIP Conference Proceedings*. Vol. 1916. No. 1. AIP Publishing LLC, 2017.

[10] Ouariachi, Tania, and Elving JL Wim. "Escape rooms as tools for climate change education: an exploration of initiatives." *Environmental Education Research* (2020): 1-14.
[10] Kemp, Henry D., et al. "Virtual Reality Escape Room." (2018).
[11] Zhang, Xiyuan, Jing-Qi Lu, and Kyoungju Park. "Interaction Design for Room Escape Virtual Reality Games." *TECHART: Journal of Arts and Imaging Science* 5.4 (2018): 24-29.
[12] Mystakidis, Stylianos, Enrique Cachafeiro, and Ioannis Hatzilygeroudis. "Enter the Serious E-scape Room: A Cost-Effective Serious Game Model for Deep and Meaningful E-learning." *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, 2019.

sciendo

DOI: 10.2478/ijasitels-2020-0005

Part of Speech Tagging Using Hidden Markov Models

Adrian BĂRBULESCU[1], Daniel I. MORARIU[1]

[1]Computer Science and Electrical and Electronics Engineering Department, Faculty of Engineering, "Lucian Blaga" University of Sibiu, Romania

{adrian.barbulescu, daniel.morariu}@ulbsibiu.ro

Abstract

In this paper, we present a wide range of models based on less adaptive and adaptive approaches for a PoS tagging system. These parameters for the adaptive approach are based on the n-gram of the Hidden Markov Model, evaluated for bigram and trigram, and based on three different types of decoding method, in this case forward, backward, and bidirectional. We used the Brown Corpus for the training and the testing phase. The bidirectional trigram model almost reaches state of the art accuracy but is disadvantaged by the decoding speed time while the backward trigram reaches almost the same results with a way better decoding speed time. By these results, we can conclude that the decoding procedure it's way better when it evaluates the sentence from the last word to the first word and although the backward trigram model is very good, we still recommend the bidirectional trigram model when we want good precision on real data.

Keywords: Part of Speech, Hidden Markov Model, rule-based tagger, word structure analysis

1 Introduction.

Part of speech tagging is an important step in the domain of natural language processing, nowadays many companies use speech and language processing algorithms to develop different applications for their clients. Some examples of those applications used in the real world are: intelligent chatbots, various virtual assistance technologies such as Alexa or Siri, social network websites, most well-known search engines such as Google, Bing, DuckDuckGo, many smartphone mobiles, etc.

In this paper, we present an automated system that analyzes an English text and tries to correctly identify the parts of speech for each word using machine learning algorithms. The system uses a benchmark text corpus for the learning and for the evaluation phase, which is presented in section 2 and continues in section 3 with the necessary steps to prepare and preprocess the input data for the learning algorithm. In section 4 we present the theoretical aspects for the learning algorithms and how it was modified and adapted to work in the current

context. In section 5, we present methods used to evaluate the model and the obtained results are presented in section 6. The conclusions and possible further developments are presented in section 7.

2 The dataset

The used dataset is a collection of texts documents that were created and annotated specifically to have the possibility to evaluate the quality of the supervised learning algorithms used for the tagging process. For this paper, we have used the *Brown Corpus*, a collection of sentences and phrases written in English language, collected, and organized by Winthrop Nelson Francis & Henry Kucera from the language department at Brown University [1]. The corpus contains over 1 million words in total and exactly 500 documents. The documents from the corpus are divided in 2 main categories: informative prose containing 9 subcategories with a total of 374 documents and imaginative prose containing 6 subcategories with a total of 126 documents.

The dataset was divided using both the 70-30 approach and *k*-fold cross-validation approach. In 70-30 approach the documents from each subcategory were divided randomly: 70% for the training set and the rest (30%) is for the testing set. In the other approach *k*-fold cross-validation, divides each category into several distinct groups called folds and at each turn, a single fold will form the testing set and the other folds will form the training set.

3 Preprocessing

The dataset must be prepared in order to be used by the learning algorithm, so a series of actions are being applied and the text from the dataset is converted into a vector representation (bag-of-word type). This representation also contains annotated text with the appropriate part of speech (PoS) for each word. All the words in the dataset and the corresponding PoS are extracted based on separation rules. For Brown Corpus, space, tab and newline are very good separation elements.

This dataset contains many type of PoS tags (approx. 100 tags) and the analysis made by the learning algorithm would take a long time if the algorithm verifies each candidate tag at every step (the prediction would be slow and the results may suffer). For this reason, we have grouped all the existing part of speech tags into 10 basic categories, those being the basic speaking parts in English language. In table 1 in the first column there are listed the names for the basic categories and in the second column there are presented 2 or 3 examples of the original tags that were included in that category.

Before extracting the relevant words (also called tokens) from the dataset, first we remove the special characters (or stop characters), those are characters like

round brackets '()', square brackets '[]' and braces '{}' which are not important for the tagging process. Another process is eliminating tokens that contain only numbers, if the token contains numbers and letters, we will remove the numbers and keep the token as a word only if it will pass a certain remaining letters threshold.

In the preprocessing step in the training phase, we will also keep a list of words that start with a capital letter (capitalized words) and separately another list in which all words are converted to lowercase. The words in the test set will go through the same filter except for the last step, words that start with a capital letter will not be converted to lowercase.
In the test set, we will also remove repetitions for tokens that have a "End of sentence" tag, this is done because the algorithm doesn't evaluate these tags and repetitions like "?!?!" may issue errors later in the decoding phase.

Base part of speech	Brown Corpus tags	No. of words	Percentage
Noun (NN)	nn, nns, nps$, …	273608	23.56 %
Verb (VB)	vb, bem, hvd, …	176081	15.16 %
Article/Determiner (AT/DT)	at, ap, dt, …	142123	12.24 %
Preposition (PP)	in, to	137735	11.86 %
Others (OT)	cd, nil, *, …	108766	9.37 %
Adjective (JJ)	Jj, jjs, jjt, …	72125	6.21 %
Pronoun (PN)	pn, pp$, wps, …	71421	6.15 %
End of sentence (<s>)	"."	61254	5.28 %
Conjunction (CC)	cc, cs	60551	5.21 %
Adverb	rb, rp, qlp, …	57528	4.95 %

Table 1: Base part of speech and the frequency of occurrence for each one of them

4 Part of Speech - tagging

We used two approaches for the automatic part of speech tagging process. The first approach is the less adaptive one, those ones may either use the most frequent word tag as the model or using a default-tag to return all the tags or a combination between these two. The second approach is based on the Hidden Markov Models, these try to predict the PoS based on the transitioning probability from one hidden state to another state and based on the emission probability which returns the association probability of a tag with a certain word.

4.1 Less adaptive approaches

One simple approach to tag a text corpus is by using a default tagger, this one returns only a specific tag for every word in the text. Usually, the most used tag for this model is the "noun" tag, which is almost a quarter percent of Brown Corpus. Another less adaptive approach is using the most frequent class

baseline model. This uses a training set to form a dictionary with tag frequencies for each word, when choosing the tag for a certain word the model will search the dictionary for that word and it will pick the tag with the highest frequency, if the word doesn't have a dictionary (word not existing in the training set) then the model will return the "not found" tag. Instead of returning the "*not found*" tag, we can combine this model with the *default-tagger = noun*, this will return the "noun" tag when an unknown word is found in the test set. This process of combining models will be able to obtain a decent accuracy for unknown words and it will also increase the overall accuracy of the most frequent class baseline model. We called less adaptive approaches (not un-adaptive approaches) because the parameters for those methods depend on the training dataset, are not general valid parameters for all contexts (only "Default tag" can be considered general valid).

4.2 Adaptive approaches

4.2.1 Hidden Markov Model

4.2.1.1 n-Gram model

The Hidden Markov Model (HMM) can link the connection between observable states (in our case these are the words in the text) and hidden states (these are part of speech tags). A Hidden Markov Model has 2 important components, a matrix *A* that contains the transition probabilities and a matrix *B* that contains the emission probabilities [2], [3].
The *A* matrix contains the probability that a tag will appear after another tag appeared at the previous step. For example, knowing that the article "The" appeared on the previous step, it will be more likely that at the current step, the selected tag will be a noun "The car…". The probabilities are evaluated based on the test set and are calculated by counting every tag sequence that appears in the training set.
To be able to form a matrix with the transition probabilities for a "*n*-gram", the Markov model uses a training set from which these occurrence frequencies are extracted. To collect the occurrence frequencies, each sequence in the training set will be added or incremented in the transition matrix *A*.
After creating the Hidden Markov Model in the training phase, we only calculate the probabilities for unigram (1-gram), bigram (2-gram) and trigram (3-gram) [4] from the testing set using the following formulas.
Unigram:

$$P(t_i) = \frac{c(t_i)}{N} \quad (1)$$

where $c(t_i)$ represents the occurrence frequency for the tag t_i in the training set and N is the total numbers of tokens in the training set.
Bigram:

$$P(t_i|t_{i-1}) = \frac{c(t_{i-1},t_i)}{c(t_{i-1})} \quad (2)$$

where $c(t_{i-1}, t_i)$ represents the occurrence frequency for the tag t_{i-1} followed by the tag t_i in the training set and $c(t_{i-1})$ represents the occurrence frequency for the previous tag in the training set.
Trigram:

$$P(t_i|t_{i-1}, t_{i-2}) = \frac{c(t_{i-2}, t_{i-1}, t_i)}{c(t_{i-2}, t_{i-1})} \quad (3)$$

where $c(t_{i-2}, t_{i-1}, t_i)$ represents the occurrence frequency for the tag t_{i-2} followed by the tag t_{i-1} which is also followed by the tag t_i in the training set. $c(t_{i-2}, t_{i-1})$ represents the previous bigram frequency.
The *B* matrix, with the emission probabilities, represents the probability that a certain tag is associated with a certain word in the training set. The formula, used by the system, to describe the maximum probability estimate is as follows:

$$P(w_i|t_i) = \frac{c(t_i, w_i)}{c(t_i)} \quad (4)$$

where $c(t_i, w_i)$ represents the occurrence frequency for the tag t_i associated with the word w_i in the training set and $c(t_i)$ represents the occurrence frequency for the tag t_i in the training set.

4.2.1.2 Smoothing techniques

These techniques are used when certain sequences of probabilities are missing from the A matrix and the higher the rank of the selected *n*-gram is, the chance that some sequences of probabilities may be missing, is higher. Various data smoothing techniques are used in the literature to solve this problem, so that when a *n*-gram sequence is not found, its value is estimated by a smoothing function. In this article we used 2 estimated methods for the missing probabilities sequences.

One method of estimation is via linear interpolation [5]. This method involves calculating a new probability composed of the sum of the transition probabilities (unigram, bigram, trigram) multiplied by a predetermined weight:

$$P_{LI}(t_3|t_1, t_2) = \lambda_1 P(t_3) + \lambda_2 P(t_3|t_2) + \lambda_3 P(t_3|t_1, t_2) \quad (5)$$

The values of the coefficients $\lambda_1, \lambda_2, \lambda_3$ are estimated by the linear interpolation function [2], [5] and the probabilities are computed using formulas (2), (3) and (4).
Another estimation method used is the additive smoothing (α-estimate) [6]. This involves adding some constant values to the numerator and denominator in the probability function (usually correlated with the length of the dataset).

$$\theta_i = \frac{x_i + \alpha}{N + \alpha d} \quad (6)$$

Where $\frac{x_i}{N}$ represents the non-smoothed probability, α is the smoothing constant and d is the size of the data ($i = \overline{1, d}$). For $\alpha = 0$, then the above formula (6) does not use any smoothing and for $\alpha = 1$, then the new formula will be called *Laplace's Rule of Succession* or Laplace's smoothing technique.

4.2.1.3 Unknown words model

The Hidden Markov Model for unigram, bigram and trigram, that uses only this model on the tagging system, can get a decent accuracy but it can not reach a very good accuracy. These occur because there are many words that are not in the training set and the system presented so far does not predict these words at all. If the tagging system would be used as an application working with real data where the number of existing words is much higher than the words learned, the performance would decrease considerably. Because of this, we also expanded our system to be able to predict these unknown words (words that are not in the training set). There are several ways to tag unfamiliar words such as: using a rule-based system, unsupervised learning algorithms, word structure analysis algorithms, etc.

In this paper we have developed a system that uses two methods to tag unknown words. One is based on the analysis of the word structure and the other one uses manually added rules that were created based an analysis of the training dataset. The final function will combine these 2 methods and it will return a probability of associating the unknown word with a specific tag.
The method that is based on the analysis of the word structure tries to identify suffixes / prefixes that can appear in the specified word. In order to be able to choose the best suffixes & prefixes, those were not deduced and estimated from the training set (high computational time and mediocre results), but were chosen as the most representative, using a list of suffixes & prefixes provided by [7]. Following this analysis, the list of prefixes and suffixes chosen for this system are:

List of prefixes: "inter", "intra", "mis", "mid", "mini", "dis", "di", "re", "anti", "in", "en", "em", "auto", "il", "im", "ir", "ig", "non", "ob", "op", "octo", "oc", "pre", "pro", "under", "epi", "off", "on", "circum", "multi", "bio", "bi", "mono", "demo", "de", "super", "supra", "cyber", "fore", "for", "para", "extra", "extro", "ex", "hyper", "hypo", "hy", "sub","com", "counter", "con", "co", "semi", "vice", "poly", "trans", "out", "step", "ben", "with", "an", "el", "ep", "geo", "iso", "meta", "ab", "ad", "ac", "as", "ante", "pan", "ped", "peri", "socio", "sur", "syn", "sy", "tri", "uni", "un", "eu", "ecto", "mal", "macro", "micro", "sus", "ultra", "omni", "prim", "sept", "se", "nano", "tera", "giga", "kilo", "cent", "penta", "tech".

List of suffixes: "able", "ible", "ble", "ade", "cian", "ance", "ite", "genic", "phile", "ian", "ery", "ory", "ary", "ate", "man", "an", "ency", "eon", "ex", "ix","acy", "escent", "tial", "cial", "al", "ee", "en","ence", "ancy", "eer", "ier", "er", "or", "ar", "ium", "ous", "est", "ment", "ese", "ness", "ess", "ship", "ed", "ant", "ow", "land", "ure", "ity", "esis", "osis", "et", "ette", "ful", "ify", "ine", "sion", "fication", "tion", "ion", "ish", "ism", "ist", "ty", "ly", "em", "fic", "olve", "ope","ent", "ise", "ling", "ing", "ive", "ic", "ways", "in", "ology", "hood", "logy", "ice", "oid", "id", "ide", "age", "worthy", "ae", "es".

In order for the model to be able to use these affixes, it is first necessary to identify with which tags the words that have these affixes are associated in the training set and to calculate the probability of association with the tag encountered. If the affix is not found in the training set, then the additive smoothing will take place. To calculate this, we use the following formula:

$$P_{sp}(x_i|t_i) = \frac{c(t_i,x_i)+\alpha}{\sum_{k=1}^{T_n^{x_i}} k+\alpha d} \quad (7)$$

where $c(t_i, x_i)$ represents the occurrence frequency for the tag t_i associated with the prefix/suffix x_i, $\sum_{k=1}^{T_n^{x_i}} k$ represents the sum of the tags frequencies associated with the prefix/suffix x_i.

The second component of the tag identification function for unknown words is the manually based rule component. The following rules were used in this article: "words that start with a capital letter" are more likely to be nouns, "words that contain an apostrophe and end with the *s* character" are very likely to be nouns, "words that contain hyphen ('-') or slash ('/')" have a higher probability to be compound words of type OT (others) or JJ (adjective), "words that contain an apostrophe and end with *the t character*" are very likely to be verbs and "words that contain an apostrophe and end with the *'ve'* or *'ll'* characters sequence" are very likely to be pronouns.

In order to be able to combine these 2 methods presented above, the probability of the unknown word with the current tag is calculated according to the suffixes and prefixes associated with it (note P_{sp}) and the rule-based probability is calculated according to the passed conditions for the weights of rules (note P_r). These are combined in the following final probability:

$$P(w_k|t_i) = P_{sp}(x_i|t_i) + P_r(w_k|t_i) \quad (8)$$

Following this addition, the result may exceed the probability interval (0,1]. For this, a threshold function will be executed that will round the value to 1.0. Exceeding the limit only suggests that there is a probability of 100% (maximum confidence) that the current tag being tested is also the correct one, usually this value is obtained for the noun tag which in most cases is also the correct tag.

4.2.2 Decoder

In the previous section we presented how models are created for known words (emission and transition probabilities) and for unknown words (rule-based and word structure analysis). Next, we present the decoder part of the system, without it the tagger cannot determine the hidden variables sequence (the tag sequence) associated with the observations sequence (the words of a sentence) [2].

The algorithm used to decode an HMM, based on the dynamic programming, is the Viterbi algorithm [2]. The Viterbi algorithm can process the states of the trellis from left to right or opposite. The general formula to calculate any node at each step in the trellis is as follows (at time step t, where t ≠ 0):

$$v_t(j) = max_{i=1}^{N} v_{t-1}(i) P_{LI}(t_3|t_1,t_2) P(w_i|t_i) \quad (9)$$

where $v_t(j)$ represents the current Viterbi node processed for tag j and v_{t-1} represents the Viterbi node processed at previous time step.

To achieve better performances for the HMM, we came up with three decoding methods and their results are presented in this article. We will call these methods *forward* (goes from the first word of the sentence to the end of the sentence, then a backtrack is made to return the final tags), *backward* (goes

from the end of the sentence to the beginning of it, then the backtrack is made) and *bidirectional* (will execute both forward and backward method and then will backtrack either by the forward or backward methods, depending on the maximum value of the final node).

5 Model evaluation

The dataset used for this paper is a pre-labeled corpus, in this case the test set is also pre-labeled with the correct tags. These tags are not used in the prediction process, they are only used to evaluate the performance of the learning algorithms (evaluating models phase). To evaluate the model, we used 2 approaches, the first approach calculates a simple prediction accuracy, and the second approach calculates the metrics based on the confusion matrix. The first assessment involves calculating the percentage of correctly predicted tags (by comparing the predictive tag with the actual one from the test set) from the total number of predicted tags. We further note this evaluation metric as "Accuracy_1". For this approach, we could also calculate the accuracy for the known words and separately for the unknown words to get a better understanding of each individual model.
The second approach involves calculating the confusion matrix for the entire testing set and extracting the evaluation parameters from that matrix as: Accuracy_2, Precision, Recall, Specificity and F1-score [8].

6 Obtained results

Table 2 presents the results obtained by the bidirectional trigram model for each part of speech separately as well as an average for each evaluated metric (*TOTAL* line). We chose this model because it obtained the best results and integrated both forward and backward methods. Those results were obtained using the 70-30 approach to split the dataset and the confusion matrix was computed for each tag in the testing phase.

TAG	ACCURACY_2	PRECISION	RECALL	SPECIFICITY	F1-SCORE
NN	97.86%	95.70%	95.75%	98.57%	95.72%
OT	99.85%	99.19%	99.19%	99.92%	99.19%
CC	99.37%	91.88%	97.50%	99.48%	94.61%
JJ	98.83%	89.72%	93.16%	99.23%	91.41%
PP	99.41%	97.74%	97.59%	99.67%	97.67%
AT/DT	99.37%	98.19%	96.94%	99.73%	97.56%
VB	98.64%	96.67%	94.77%	99.38%	95.71%
PN	99.85%	98.96%	98.72%	99.93%	98.84%
RB	98.87%	90.19%	88.47%	99.46%	89.32%
TOTAL	99.12%	95.36%	95.79%	99.49%	95.56%

Table 2: Results obtained by the bidirectional trigram model

The most difficult tags to predict, which have the lowest score, were the adverb and the adjective. These 2 tags are the most context dependent tags and the context dependent tags do not have a high probability for only a tag and are, therefore, very difficult to predict in some contexts. The predicted tag with the best score is the Others tag, this tag contains interjections, numbers (cardinal numbers), compound words, etc. which most of the time have a unique morphological form and are not context dependent (for example, the word "one" will always have the cardinal number tag). These very good results can be explained by the fact that the training set and the test set are from the same dataset and have the same distribution. Even if there are other documents used for testing than those used for training, they come from the same sources (same authors) that tend to use the same words in the same contexts and automatically with the same part of speech.

For the forward bigram model, the average training time of the model is approximatively 1,41 minutes and the average decoding time for all sequences is 1,44 minutes. For the bidirectional trigram model, the average training time of the model is 1,38 minutes (similar to the average training time for the forward bigram) and the average decoding time for all sequences is 3,56 minutes (twice as long as the forward bigram model). The training time between these models does not differ too much because the training function uses parallel threads to use the most of the processor's capabilities. The decoding time is much longer for the bidirectional trigram model because it has to calculate the trigram transition probability (also the transition probabilities for bigram and unigram) and evaluate the model both forward and backward, after which it needs to decode the best sequence for each sentence. Processing times were estimated on a desktop system with Windows 10, a Quad-Core processor with a frequency of 3.60 GHz and 16 GB Ram.

Model parameters options	Unknown words percentage from test dataset (%)	Unknown words accuracy (%)	Known words accuracy (%)	Accuracy_1 (%)
Default-tag: Noun	-	-	-	24.96
Most frequent class baseline	13.76	0.00	95.85	82.66
Most frequent class baseline + Default-tag: Noun	13.76	52.35	95.85	89.87
Forward bigram	3.84	77.37	96.45	95.72
Backward bigram	3.83	82.20	96.51	95.96
Bidirectional bigram	3.83	82.23	96.50	95.95
Forward trigram	3.84	78.28	96.60	95.90
Backward trigram	3.80	81.46	96.63	96.05
Bidirectional trigram	3.83	81.54	96.63	96.05

Table 3: Comparatively results obtained by all approaches (adaptive and less-adaptive)

In table 3 we compare the results between the less-adaptive approach and the adaptive approach. The results are presented in terms of simple accuracy (Accuracy_1) for the known and unknown words but also for all the words all together. The known words are words in the test set that also appear in the training set and the unknown words are words that do not appear in the training set. The accuracy on unknown words verifies the tagging system performance that must adapt to unknown situations when it encounters new cases. Those results were obtained using the *k*-fold cross validation approach to split the dataset (in this case $k = 4$).

For less adaptive models (specifically most frequent class baseline model from table 3), we do not keep a separated list for the words that start with a capital letter. The default-tagger doesn't have any concept of known words or unknown words because it only predicts one tag, the most frequent class baseline and this model compound together with the default-tagger keep a list of unmodified words without converting the capitalized words to lowercase in the testing set. Without this conversion and without keeping a separated list with capitalized words only, the percent of unknown words for less adaptive models is way higher than the percent of unknown words for the adaptive models. As we can see, less adaptive approach can get good results (only on familiar words) but on a completely new dataset, the performance could decrease quite a bit.

From these experiments we noticed that a backward trigram model is almost as good as the bidirectional trigram. Analyzing in more detail, we noticed that in the case of bidirectional trigram, in 85.52% of cases the backward branch is chosen when using the bidirectional method. Figure 1 shows the percentage in which a bidirectional trigram model chooses the backward method as the optimal back trace path, being compared to the forward method. In 0.03% of cases, both models return the same result because the final nodes (final states) for these two methods have the same value.
Another interesting observation is that the bidirectional bigram model has the best accuracy for unknown words (82.23%) but does not have as good an accuracy as the bidirectional trigram for known words (96.63%). The forward bigram has a poor performance (comparing it with the others that use a Markov model) and the forward trigram has a better performance compared to the forward bigram. The problem of tagging of PoS is complicated because analyzing the dataset, 67% of tokens are ambiguous [2] (have more than one possible part of speech in different contexts), if we use a less adaptive approach like the most frequent class baseline + default tag: noun, we can almost achieve a 89.87% accuracy of correctly predicted tags. In the preprocessing section we showed that the noun tag represents approximately 23.56% of the dataset, eliminating the end of sentence tag and using only the first model with the default-tag = noun, an accuracy of ~ 25% need to be obtained, which implies that almost a quarter of the test set was "predicted" correctly. Because of the large dataset, the difference between these models' results are small, same applies for the results presented in table 2. The problem becomes

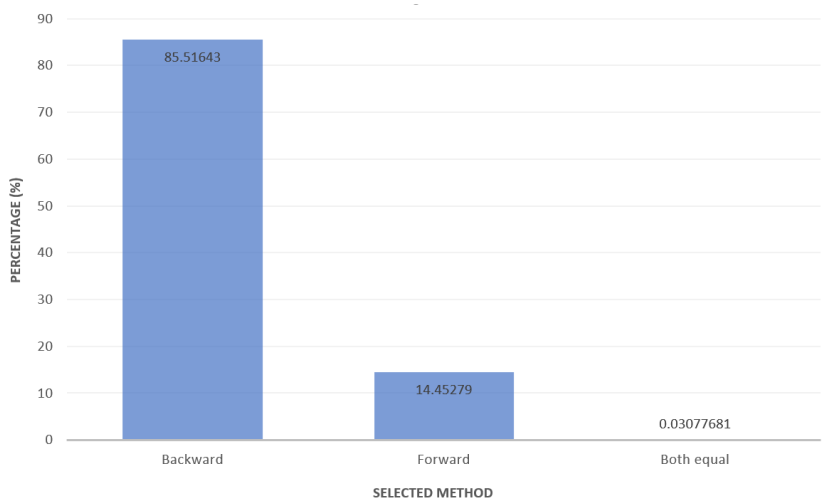


Figure 1:Percentage for each final node value chosen by the bidirectional method

complicated when the testing set have a different distribution than the training set (for example sentences from real world). In table 3, in the first column "Unknown words percentage from test dataset" the percent differs (is between 3.80% to 3.84%) because we use *k*-fold cross-validation for each test and the documents were randomly grouped in each fold.

7 Conclusions

The purpose of this paper is to analyze the performance of the Hidden Markov Model in the PoS tagging system. For this we implemented and tested 2 different models of HMM (bigram and trigram) and 3 different methods to compute the maximum probability values for a sentence (forward, backward and bidirectional). The trigram model obtains better results combined with the bidirectional or the backward method. In the bidirectional method, in almost all cases (85.51%) the best choice was the backward method. Also, we compare the obtained results by the HMM approaches with less-adaptive approaches as "predicting the same tag all the time" or predicting the "most frequent class baseline" to see the improvement from the accuracy standpoint, if we add the learning process in the PoS tagging process. The increase was from 89.87% (when we use a combination between all less-adaptive approaches) to 96.05% (when we use the learning approach).

In conclusion, the automatic speech part tagging system presented in this paper uses well-established algorithms in the field of machine learning and natural language processing to achieve state of the art performances. These performances were achieved on a test set extracted from the same dataset that the training set also uses.

The system presented in this paper was also tested with sentences entered manually by the user and the response was evaluated only by the user. Until

now the system works only for the English language and it can be extended easily to other languages.

References

[1] W. Nelson Francis and Henry Kučera at Department of Linguistics, Brown University Standard Corpus of Present-Day American English (Brown Corpus), Brown University Providence, Rhode Island, USA, korpus.uib.no/icame/manuals/BROWN/INDEX.HTM
[2] Dan Jurafsky, James H. Martin, Speech and Language Processing, third edition online version, 2019
[3] Lawrence R. Rabiner, A tutorial on HMM and selected applications in Speech Recognition, Proceedings of the IEEE, vol 77, no. 2, 1989
[4] Adam Meyers, Computational Linguistics, New York University, 2012
[5] Thorsten Brants, TnT - A statistical Part-of-speech Tagger (2000), Proceedings of the Sixth Applied Natural Language Processing Conference ANLP-2000, 2000
[6] C.D. Manning, P. Raghavan and M. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008
[7] Lois L. Earl, Part-of-Speech Implications of Affixes, Mechanical Translation and Computational Linguistics, vol. 9, no. 2, June, 1966
[8] Daniel Morariu, Radu Crețulescu, Text mining - document classification and clustering techniques, Published by Editura Albastra, 2012

sciendo

DOI: 10.2478/ijasitels-2020-0002

Imperialist competitive algorithm for determining the parameters of a Sugeno fuzzy controller

Stelian Ciurea[1]
"Lucian Blaga" University of Sibiu, Faculty of Engineering, Department of Computer and Electrical Engineering, E. Cioran Str, No. 4, Sibiu-550025, ROMANIA;
Research Center in Informatics and Information Technology.
E-mail: stelian.ciurea@ulbsibiu.ro

Abstract

We used an imperialist competitive algorithm to determine the parameters of a fuzzy controller of type Sugeno that would ensure a good unit step response of a second-order single-input and single-output automatic system.

Keywords: **S**ingle-input and single-output second-order linear system, Fuzzy Controller og type Sugeno, Imperialist Competitive Algorithm

1. Introduction

First described in 2007 by Gargari et al [1], the imperialist-competitive algorithm (hereinafter referred to as ICA) is an alternative to other heuristic methods of finding solutions as close as possible to the optimal ones in complete hard-NP problems. In previous papers we have studied how this type of algorithm can solve discrete problems such as the traveler salesman problem [15], or continuous problems such as determining the minimum for real functions with several arguments [16]. In this article we present how a competitive imperialist algorithm was used to determine the parameter values for a fuzzy controller of type Sugeno that was used to regulate a SISO second order linear system. The representation of the controller was done by 4 integer values and 199 real values. For this reason, ICA-specific operations have been implemented modeled on those used in the articles we referred to earlier ([15], [16]).

2. The mathematical model used for the second-order linear system

We have implemented an ICA (imperialist competitive algorithm) in order to determine the parameters of a controller so as to enable the best possible unit step response of a second-order automatic system. The

function at the output of such a system is given by the following differential equation [2]:

$$a_2 \frac{d^2 y}{dt^2} + a_1 \frac{d\ y}{dt} + a_0 y(t) = b_2 \frac{d^2 u}{dt^2} + b_1 \frac{d\ u}{dt} + b_0 u(t) \quad (1)$$

where u(t) is the input variable and y(t) the output variable. In case of downtime τ in the transmission of the data, the corresponding differential equation is the following:

$$a_2 \frac{d^2 y}{dt^2} + a_1 \frac{d\ y}{dt} + a_0 y(t) = b_2 \frac{d^2 y}{dt^2} + b_1 \frac{d\ y}{dt} + b_0 u(t-\tau) \quad (2)$$

Many papers such as [3], [4], [8], [10], [12] use this type of equation in describing the behavior of DC motors. We have considered the standard structure of an automated system consisting of a second-order system and a fuzzy controller, illustrated in Fig. 1 ([2],[8]).

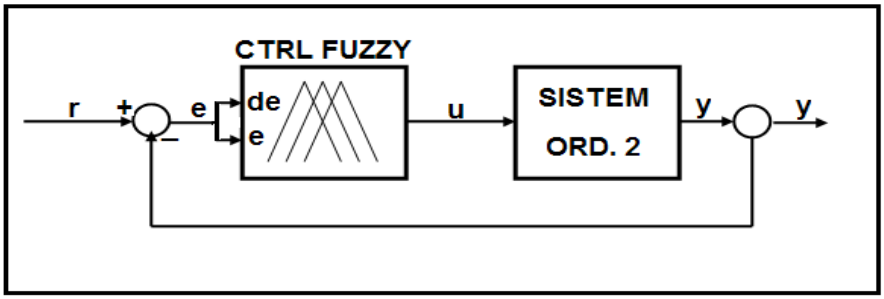


Figure 1. Automated regulating system

Our aim is for output value y to follow the evolution of reference value r as accurately as possible.

To simulate the operation of the system, we have considered its discrete model. The method, described in [2], consists in considering relation (1) at time intervals equal to dt and integrating the terms from that formula twice:

$$\int_{t-dt}^{t} \int_{t-dt}^{t} (a_2 \frac{d^2 y}{dt^2} + a_1 \frac{d\ y}{dt} + a_0 y(t)) = \int_{t-dt}^{t} \int_{t-dt}^{t} (b_2 \frac{d^2 u}{dt^2} + b_1 \frac{d\ u}{dt} + b_0 u(t)) \quad (3)$$

By using the trapezoidal rule for the calculation of integrals, we have obtained the following relation:

$$y_k = \frac{(b_2 + b_1\Delta + b_0\Delta^2)u_k + 2(b_0\Delta^2 - b_2)u_{k-1} + (b_2 - b_1\Delta + b_0\Delta^2)u_{k-2}}{a_2 + a_1\Delta + a_0\Delta^2} - \frac{2(a_0\Delta^2 - a_2)y_{k-1} - (a_2 - a_1\Delta + a_0\Delta^2)y_{k-2}}{a_2 + a_1\Delta + a_0\Delta^2} \quad (4)$$

where y_k, y_{k-1} and y_{k-2} are output values y at times t_k, t_{k-1} and t_{k-2}, $t_k - t_{k-1} = t_{k-1} - t_{k-2} = dt$; analogous u_k, u_{k-1} and u_{k-2} are values u, and $\Delta = (dt/2)^2$.

We have aimed at the operation of this system within the range of [0; tmax] seconds, considering that:

- at time $t_0=0$, a step signal of an amplitude equal to the unit is applied;
- prior to time t_0, the system was idle;
- dt=0.005 seconds.

3. Designed fuzzy controller

We have implemented a fuzzy controller of type Sugeno of type PD (for details see [13] and [14]) with two inputs and one output. The two entries have been the following:

- "error", marked *e*, is the percentage difference between the reference value and the system output. We have considered the universe of discourse for this value to be interval [-100; 100]. Its value at time t_k is given by the formula:

$$e_k = \frac{r_k - y_k}{r_k} \times 100 \quad (5)$$

- "error variation", marked *de*, having universe of discourse [-5; 5] and calculated at time tk by means of the following formula:

$$de_k = \frac{e_k - e_{k-1}}{\Delta t} \quad (6)$$

For the output of the controller noted *u*, the interval [-200; 200] was chosen.

We set seven fuzzy sets for the two entries. The associated linguistic terms were in the case of both inputs NB (big negative), NM (negative medium), NS (negative small), ZE (zero), PS (positive small), PM (positive medium), PB (positive big). The membership functions we employed are trapezoidal or built on two Gauss curves (Fig. 2 and Fig. 3).

We set seven fuzzy sets for the two entries. The associated linguistic terms were in the case of both inputs NB (big negative), NM (negative medium), NS (negative small), ZE (zero), PS (positive small), PM (positive medium), PB (positive big). The membership functions we employed are trapezoidal or built on two Gauss curves (Fig. 2 and Fig. 3).

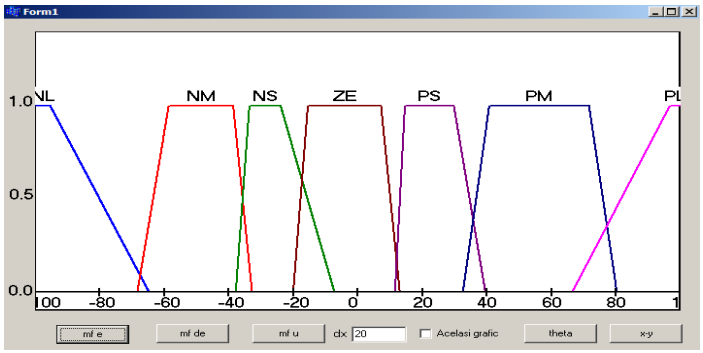


Figure 2. Example of trapezoidal members 1

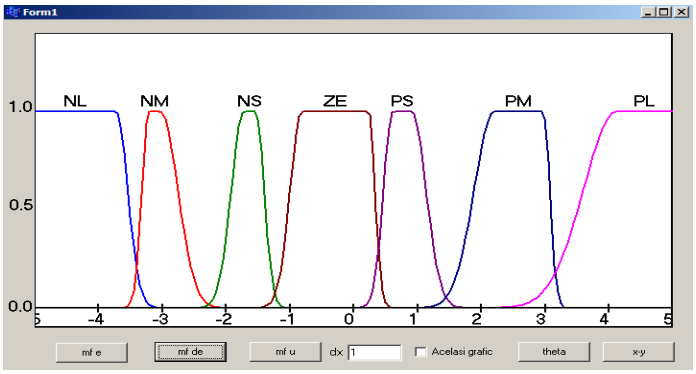


Figure 3. Example of membership functions shaped as a combination of two Gauss curves for variable *de*

For output variable u, the membership functions are linear functions in input variables. The rules used for this type of fuzzy controller are of the type below:

IF *e* is NB and *de* is NB, THEN $u_1 = a_1e + b_1de + c_1$

IF *e* is NB and *de* is NM, THEN $u_2 = a_2e + b_2de + c_2$ (7)

With 7 fuzzy sets for *e* and 7 fuzzy sets for *de*, we will have 49 controller rules (one rule for each pair of fuzzy sets *e* - *de*). Since the three coefficients in the above expressions (a_i, b_i si c_i, i=1,2,…,49) are different for each pair, we need 3 × 49 = 147 parameters for the controller rules.

The defuzzification method used for this type of control is the "weighted average"; thus, denoting by MFe_i one of the fuzzy functions that determine input *e*, by $MFde_i$ one of the fuzzy functions that determine input *de*, and by e_v and de_v the values of *e* and *de* at a given time, we calculate w_i = the weight of the appropriate inference rule:

w_i =AndMethod($MFe_i(e_v)$, $MFde_i(de_v)$)

We calculate the value provided by the controller using the following formula:

$$u = \frac{\sum_{i=1}^{49} w_i \theta u_i}{\sum_{i=1}^{49} w_i} \quad (8)$$

It follows that, for a complete definition of the fuzzy controller, we need parameters defining the type and position of the fuzzy functions on the discourse universe axis for the two input values (*e* and *de*), the

implementation of the fuzzy AND operation, and the values of the 147 parameters, a_i,b_i and c_i (i=1,2,…,49).

4. The Designed ICA

Imperialist competitive algorithms (ICA) are part of the category of heuristic search algorithms used to determine solutions as close as possible to optimal solutions for hard NP-complete problems. In these problems, the solutions are usually made up of a series of integer and/or real values. The basic idea is to generate a predetermined number of possible solutions (called in ICA „countries"), evaluate them, preserve the best (called in the ICA „metropoli") and explore their neighborhoods by attracting other solutions (called in the ICA „colonies") to the metropolis through the assimilation operation. To avoid stagnation of the solution set, the operation called revolution is used, which is applied to some of the colonies. By this some colony type solutions are thrown out of the vicinity of the metropolis type solutions.

The ICA algorithm we used in our application is similar to the one presented by Gargari et al in [1]:

1. Generate a random set of countries;
2. Evaluate these solutions and the best ones become metropolises;
3. Generate the initial empires by distributing the colonies;
4. Repeat
    5. Assimilation
    6. If a colony has better results than the imperialist country then
        a. Interchange the colony with the imperialist country
    7. Competition between empires:
        a. Assess empires
        b. Distribute the weakest colony of the weakest empire by another empire
        c. If the weakest empire has no colonies left then
            i. Remove this empire

Until one of the stopping conditions is reached.

The stopping conditions are:
- only one empire remained;
- a predetermined number of iterations of the repeat… until loop were performed.

We will present the particularities of the algorithm specific to the problem it has to solve.

4.1 Representation of controllers

The controllers (or "countries" from the point of view of ICA) have been represented in the form of structures having the following makeup presented in detail in [11]:

- 1 integer field for coding fuzzy sets for the three linguistic variables corresponding to the two inputs and to the output: trapezoidal or a combination of two exponential functions;

- 1 integer field representing the number of fuzzy sets in each linguistic variable (seven in our case);
- 1 boolean field stating that the membership functions of fuzzy sets have a symmetry with respect to the central value of the universe of discourse;
- 52 real fields for coding the form and locations on the axis of the universe of discourse of the fuzzy sets within each linguistic variable (e and de), with the following implications for any of these variables:
    - 6 values representing the ratio between the larger base of the trapezium and the average value of the universe of discourse – a value within the range [0.25; 2.0];
    - 6 values representing the percentage of the larger base of the trapezium overlapping the larger base of the trapezium placed on its left - a value within the range [0.05 ; 0.4];
    - 7 values representing the ratio smaller base/larger base – a value within the range [0.01; 0.65];
    - 7 values representing the position of the smaller base for CE – a value within the range [0.05 ; 0.95] out of the available interval calculated depending on the large basis.
- 1 integer fields that specify the mathematical formula for the logical operations AND (0=min, 1=product)
- 147 real fields for the coefficients that appear in the expressions [7] (a_i, b_i si c_i, i=1,2,…,49).

This mode of representation is useful because thespecific operations of ICA provide results with values in the same ranges, so correct. But for performing the operations necessary for fuzzy control or for graphical representations, some simple formulas (and related calculations) are needed to determine the classic shape of fuzzy sets.

4.2 Generating the initial set of countries

This has been done randomly: for the 147 specific coefficients the range [-100; 100]. These limits determine for the command u, the theoretical interval [-300; 300]

4.3 Evaluation of colonies

To evaluate the colonies in order to implement some of the specific ICA operations (empire determination, colony distribution, competition between empires), we have simulated the behaviour of the system when applying a step signal at time 0. The controller performance was calculated with the formula:

$$performance = (\sum_{t=0}^{2.5} |r_k - y_k|) + 10\, max(yfin_{max}) \quad (9)$$

where y_{max} is the maximum that the output value takes in the studied range, and y_{fin} is the final value of the output of the system.

This evaluation function was thought of as a penalty: the lower its value, the more efficient the controller. Thus, the value of the first term of this expression is lower the shorter the response time and the lower the steady-state error. The second term is lower the less the system exceeds the steady state value.

4.4 The assimilation operation

In the application, we have used different formulae depending on the two types of parameters that make up a colony:

• for integer parameters, we have taken into account that the value of any of these parameters does not condition the value of another parameter. This led to the following formula: the colony parameter receives the value of the corresponding parameter in the metropolis with a given probability, hereinafter referred to as assimilation probability and marked $prob_a$;

• for real parameters, we have used the following formula:

$$paramcol_i = paramcol_{i-1} + p(1 \pm d)(parammet_{i-1} - paramcol_{i-1}) \quad (10)$$

where $paramcol_i$ is the value of the colony parameter after iteration i, $parammet_i$ is the value of the parameter corresponding to the metropolis of the respective colony after iteration i, and d is a subunit random value, $d \in [-d_a/2 ; d_a/2]$; d_a is the assimilation deviation.

4.5 The revolution operation

We have used different formulae depending on the type of parameter whose value is to be changed.

• for each integer parameter, the following formula was applied: with a probability equal to value d_r (revolution deviation)

$$paramcol = nrval - 1 - paramcol \quad (11)$$

where paramcol is the integer parameter and nrval is the number of values available for that parameter.

• for real parameters, we have used the following formula:

$$paramcol_i = paramcol_{i-1} - p(1 \pm d)(parammet_{i-1} - paramcol_{i-1}) \quad (12)$$

$d \in [-d_r/2 ; d_r/2]$, in this case d_r being the revolution deviation.

5 Findings

The following values were chosen for the second-order system parameters:

$a_0=20$; $a_1=10$; $a_2=1$; $b_0=1$; $b_1= b_2=0$.

These values of the coefficients determine, for the system analyzed in open loop, a transfer function with two real poles having the values -7.2361 and -2.7639.

The observation interval was [0;2.5] seconds. Tests with ten initial sets of countries have been performed. The ICA parameters have been chosen as follows:

- number of countries in a set: 108;
- initial number of empires in a set: 8;
- maximum number of iterations in the algorithm: 300;
- share of a colony in the performance of the empire (w): 0.001, fixed;
- approaching step (p): 0.1 and 0.9;
- assimilation and revolution deviations (d_a , d_r): 0.01 and 0.9;
- the probability with which the revolution operation is applied to a country: 10% and 30%.

The entire application has been implemented in the C language.

The results obtained for each of the ten sets of countries (marked with #0, #1, ..., #9) are presented in tables 1. We have noted the following:

- the average performance of the best controllers was better for higher values of the parameters p, d_a and $prob_{rev}$;
- the yield of the algorithm representing the ratio between the average performance of the best controllers in each set after the first iteration and the average performance after the last iteration had the minimum value of 3.21 and a maximum of 4.19;
- In the case of the combination of ICA parameters for which the best average performance was obtained, in all ten tests the performance had values less than or approximately equal to 20 which implies a very good behavior of these controllers;
- the best value of the performance function obtained for a controller is 15.48. The simulation of the behaviour of the system with the help of this controller is presented in Fig 4. The system has an override equal to 0.7% and a response time of 0.19s. the system did not present oscillation around the reference value in steady state.

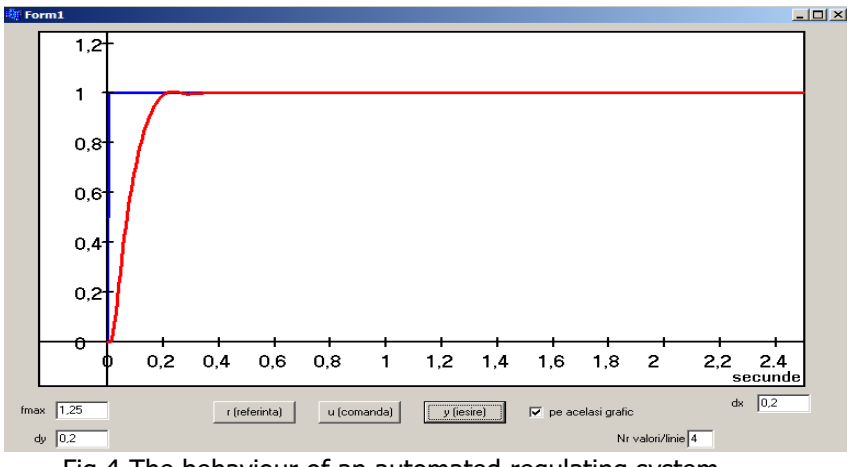


Fig 4 The behaviour of an automated regulating system with the best controller provided by ICA

Tabel 1 The performance of the best controllers obtained with the help of ICA

108 countries		Step 0.1 Assimilation deviation d_a 0.01 $prob_{rev}$		Step 0.1 Assimilation deviation d_a 0.01 $prob_{rev}$		Step 0.9 Assimilation deviation d_a 0.9 $prob_{rev}$		Step 0.9 Assimilation deviation d_a 0.9 $prob_{rev}$	
Set	Initial perform.	0.1	0.3	0.1	0.3	0.1	0.3	0.1	0.3
#0	31.28	18.56	16.58	17.24	17.62	19.87	16.29	18.55	17.03
#1	29.08	16.48	16.57	16.78	16.69	18.75	16.59	15.92	16.77
#2	77.24	16.8	17.71	16.34	18.86	20.87	16.85	16.97	15.78
#3	81.16	20.13	18.51	19.43	18.23	18.78	17.48	19.99	15.66
#4	46.2	17.26	17.87	16.44	18.82	15.75	16.30	15.78	15.62
#5	71.53	17.82	16.86	17.13	17.48	23.58	20.30	16.39	17.17
#6	147.2	17.74	17.00	19.14	17.85	25.92	17.98	19.08	16.06
#7	84.40	17.27	17.36	18.4	18.08	18.32	17.47	16.64	16.16
#8	78.04	23.27	18.76	22.85	17.27	31.95	15.81	20.29	16.37
#9	30.56	17.29	16.71	16.6	18.06	16.82	16.91	16.60	**15.48**
Average performance	67.669	18.26	17.39	18.03	17.89	21.06	17.19	17.62	16.12
Average yield		3,70	3,89	3.75	3.78	3.21	3.93	3.84	4.19

The ICA determined the minimum method for implementing the fuzzy AND operation. The fuzzy sets for the two inputs are illustrated in Fig. 5 and Fig. 6:

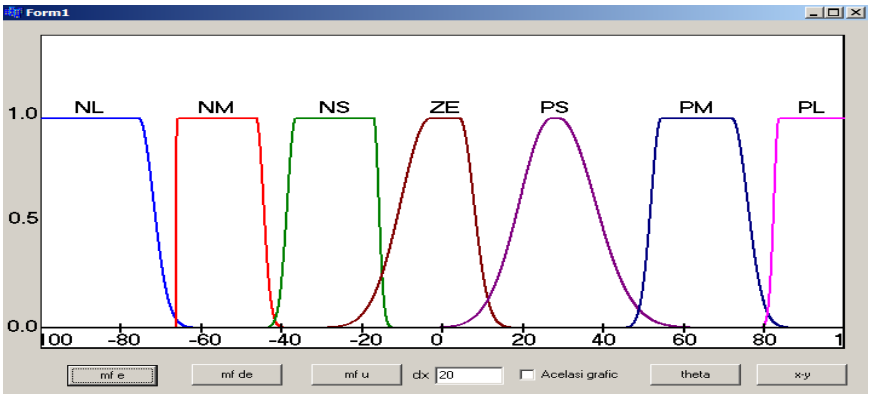


Fig. 5 Membership functions for input *e*

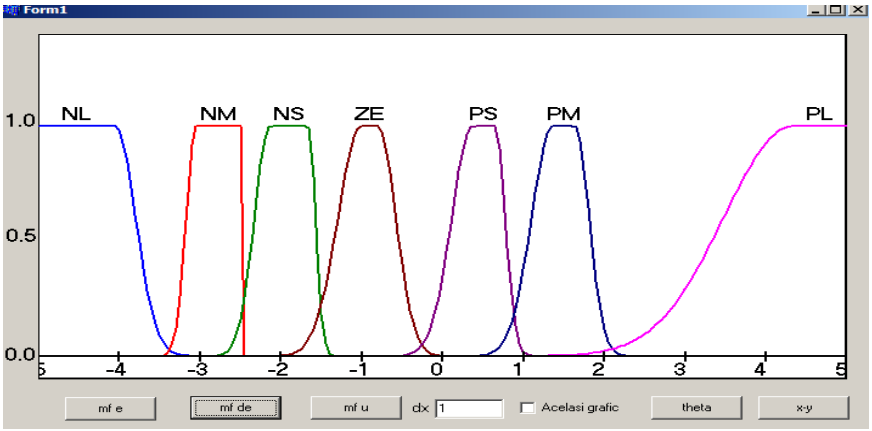


Fig. 6 Membership functions for input *de*

The values of the 147 coefficients are given in table 2.
Figure 7 shows the performance variation over the 300 iterations in the test that provided the best controller.

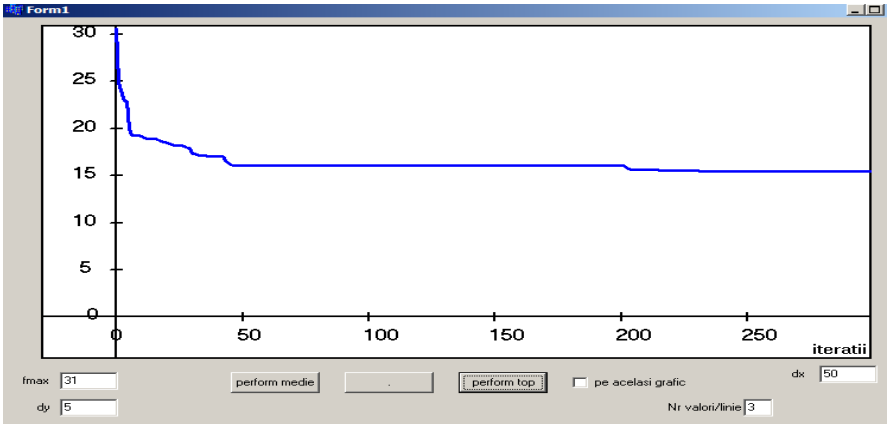


Fig.7 Performance of the best Sugeno controller vs iterations

Table 2 Inference coefficient values for Sugeno controller

i	a_{i1}	a_{i2}	b_i		i	a_{i1}	a_{i2}	b_i
1	20.85	70.12	-53.63		26	35.30	29.39	20.00
2	-5.61	-29.85	-9.81		27	-109.8	60.65	9.23
3	-26.18	-4.885	-14.49		28	4.048	-0.262	-42.66
4	-55.35	-0.978	-40.58		29	-0.742	118.5	-41.80
5	15.61	11.27	16.07		30	-26.10	74.35	-12.64
6	3.314	29.55	-18.94		31	4.725	12.03	-13.10
7	-16.81	-28.01	12.88		32	-69.34	-6.803	-0.238
8	-58.71	-0.971	30.26		33	-24.85	-10.99	-86.02
9	0.59	-18.71	-43.46		34	-26.61	10.85	-2.355
10	-15.47	-16.77	-15.59		35	33.53	3.745	17.70
11	3.144	-49.80	2.58		36	-34.51	-26.10	28.34
12	9.687	91.86	-44.88		37	120.4	54.30	-1.045
13	-20.31	-5.146	19.99		38	40.42	-16.14	3.061
14	8.541	-98.76	22.47		39	-7.942	27.90	0.809
15	2.390	-20.40	2.371		40	74.58	-4.596	-39.50
16	-16.19	10.45	-14.65		41	-16.11	-1.064	-5.001
17	-11.74	-3.282	15.52		42	-62.15	-2.779	-13.51
18	8.887	9.813	-68.76		43	11.75	97.83	36.00
19	-66.07	32.52	-6.918		44	-0.983	-70.08	68.86
20	50.68	106.0	7.714		45	6.143	32.92	-20.92
21	-51.87	-22.60	-7.216		46	5.611	6.246	-52.41
22	-7.590	-11.84	-79.67		47	5.206	7.311	12.74
23	-16.58	4.308	-3.190		48	-14.20	16.07	55.55
24	6.983	18.66	-91.15		49	-13.61	10.52	1.807

25	31.61	3.862	-65.30					

6. Conclusions

The imperialist competitive algorithm has been able to determine parameters for a fuzzy controller to ensure the efficient operation of the second-order automated regulating system, even in difficult conditions (non-negative controls of the controller). Better results have been obtained in the tests in which the values of the parameters p, d_a and $prob_{rev}$ was hight. Applying a step signal to an automated regulating system, the best controller ensured the operation of the system with an override equal to 0.7%, a response time of 0.19s. The system did not present oscillation around the reference value in steady state.

The average running times on a computer with an Intel Core i3 microprocessor at 2.93 GHz was equal with 37 seconds (for the sets of 108 countries and a maximum of 300 iterations). Sugeno-type fuzzy controllers are much faster than Mamdami-type fuzzy controllers (for the same values of ICA parameters, on the same type of computer Mamdami controllers had average running times of 117 minutes and 21 seconds).

References

[1] Esmaeil Atashpaz Gargari, Caro Lucas, *Imperialist Competitive Algorithm: An Algorithm for Optimization Inspired by Imperialistic Competition*, IEEE Congress on Evolutionary Computation, CEC 2007.

[2] Babutia I., Dragomir T.L., Mureşan I., Proştean O.: *Conducrea automată a proceselor*, Ed. Facla, Timişoara, 1985

[3] Bhartiya, S., Sehra, S., Bhuria V.: Dc Motor Speed Control using PID- Fuzzy Logic Based Controller, http://ijair.jctjournals.com/jan2013/t50.pdf IJAIR, ISSN: 2278-7844, 2013

[4] Bindu R., Mini Namboothiripad K.: *"Tuning of PID Controller for DC Servo Motor using Genetic Algorithm"*, International Journal of Emerging Technology & Advanced Engineering, Volume 2, Issue 3, March 2012, ISSN 2250-2459

[5] Ashkan MohammadZadeh Jasour, Esmaeil Atashpaz Gargari, Caro Lucas, *Vehicle Fuzzy Controller Design Using Imperialist Competitive Algorithm*, Second joint congress on fuzzy and intelligent systems, Tehran, Iran, 2008

[6] Danilo Pelusi, *Optimization of a Fuzzy Logic Controller using Genetic Algorithms*, 2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) 2011

[7] Z. Michalewicz, "*Heuristic Methods for Evolutionary Computation Techniques*", Journal of Heuristics, Vol.1, No.2, 1995, pp.177-206

[8] Dumitrache I., Călin S., Boţan C., Niţu C.: *Automatizări şi echipamente electronice*, Ed. Didactică şi Pedagogică, Bucureşti, 1982

[9] Stelian Ciurea, *"Determining the Parameters of a Sugeno Fuzzy Controller Using a Parallel Genetic Algorithm*", Proceedings of the 9th International Conference on Control Systems and Computer Science, University Politehnica of Bucharest, Romania, 2013", ISBN 798-0-7685-4980-4, pg 36-43.

[10] Kuma D., D., Dhakar, B., Yadav, R.: "*Tuning a PID controller using Evolutionary Algorithms for an Non-linear Inverted Pendulum on the Cart System*", International Conference on Advanced Developments in Engineering and Technology (ICADET-14), INDIA, International Journal of Emerging Technology and Advanced Engineering, website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 4, Special Issue 1, February 2014.

[11] Stelian Ciurea, "*A Better Genetic Representation of a Fuzzy Controller Enabling the Determination of Its Parameters with the Help of a Genetic Algorithm*", Proceedings of MIDS 2013, ISSN 2067-3965, pg. 49-58, Sibiu, Romania

[12] Yazdani A. M., Ahmadi, A., Buyamin, S., Fua'ad Rahmat, M., Davoudifar F., Abd Rahim, H.: "*Imperialist competitive algorithm-based fuzzy pid control ethodology for speed tracking enhancement of stepper motor*", International Journal on Smart Sensing and Intelligent Systems, vol. 5, no. 3, September 2012

[13] Terano, T., Asai, K., Sugeno, M.: *Fuzzy Systems Theory and Its Applications*, Academic Press, 1992.

[14] "*Fuzzy Logic Toolbox™ User's Guide R2011b*", The MathWorks, Inc., 2011

[15] Stelian Ciurea, *An Imperialist Competitive Algorithm Optimized to Solve the Travelling Salesman Problem ,* Proceedings of MIDS 2017

[16] Stelian Ciurea, *Imperialist Competitive Algorithm with Variable Parameters to Determine the Global Minimum of Functions with Several Arguments,* Proceedings of MIDS 2015

DOI: 10.2478/ijasitels-2020-0003

On Hagelbarger's and Shannon's matching pennies playing machines

Macarie BREAZU[1], Daniel VOLOVICI[1],
Daniel I. MORARIU[1], Radu G. CREȚULESCU[1]
[1]Computer Science and Electrical and Electronics Engineering Department, Faculty of Engineering, "Lucian Blaga" University of Sibiu, Romania
{macarie.breazu, daniel.volovici, daniel.morariu, radu.kretzulescu}@ulbsibiu.ro

Abstract
In the 1950s, Hagelbarger's Sequence Extrapolating Robot (SEER) and Shannon's Mind-Reading Machine (MRM) were the state-of-the-art research results in playing the well-known "matching pennies" game. In our research we perform a software implementation for both machines in order to test the common statement that MRM, even simpler, beats SEER. Also, we propose a simple contextual predictor (SCP) and use it to compete with SEER and MRM. As expected, experimental results proves the claimed MRM superiority over SEER and even the SCP's superiority over both SEER and MRM. At the end, we draw some conclusions and propose further research ideas, like the use of mixing models methods and the use of Hidden Markov Model for modelling player's behaviour.

Keywords: matching pennies, Hagelbarger, sequence extrapolating robot, Shannon, mind reading machine, contextual predictor

1. Introduction. Game description

One of the simplest games that can be played by two opponents is the "matching pennies" game [6]. Both players choose a face of their own coin ("head" or "tail") in secret and, after that, both expose their choices. One player wins if the choices are the same (hence the name "matching pennies"), the other when the choices are different. Even if theory says that, at random play, you neither win nor lose on long term, people's play falls in patterns and this can be used by the opponent. The common people's idea is that your smartness is what wins the game. Edgar Allan Poe's character from "The Purloined Letter" uses also body-language signs to predict opponents moves at this game.

Building a machine that plays this game requires the two fundamental steps in machine learning: building a model of the problem's world (in this case of the opponent) and using this model to implement the behaviour. Therefore, being

a redoubtable problem from the (today called) machine learning point of view, it got research focus from the beginning of the artificial intelligence era.

In the 1950', at Bell Laboratories, David Hagelbarger and Claude Shannon each build their own machines to play the matching pennies game with lab colleagues (Claude Shannon built also other interesting machines [7]). For today such machines can be considered trivial, but for that moment, when the memory was implemented with relays, these were state-of-the-art research results, and therefore the machines got their remarkable place in the history of computer science.

Hagelbarger calls his machine "SEquence Extrapolating Robot' (SEER) [1], focusing on the feature that extrapolates the series of the opponent's play. Shannon calls his machine "Mind Reading machine" (MRM) [2], focusing on the feature that it models the opponents' mind. The Shannon's MRM machine was a simplified version of the Hagelbarger's SEER machine, but even so it hoped for and got better results. Images with the SEER and MRM machines can be found in [8].

Different from the SEER and MRM papers, where "+" and "–" symbols were used to describe the play, we will describe the play using the "H" (Head) and "T" (Tail) symbols (according to the "matching pennies" game's name).

An example of a matching pennies game evolution can be:

#play	Player 1 (plays for matching)	Player 2 (plays for different)
1	T	H
2	H	H
3	T	T
4	T	H
5	H	H
6	H	T
7	H	T
8	…	…

Win/Loss (for player 2)	Change (for player 2)
1 (Win)	
0 (Loss)	0 (Same)
0 (Loss)	1 (Different)
1 (Win)	1 (Different)
0 (Loss)	0 (Same)
1 (Win)	1 (Different)
1 (Loss)	0 (Same)
…	…

In both Hagelbarger and Shannon machines, the machines were designed to play for "matching the pennies", i.e. to read the opponent's mind (hence the name "mind-reading machine" used by Shannon).

2. Hagelbarger's SEER

The Hagelbarger's SEER (SEquence Extrapolating Robot) is presented in [1]. Its following description is taken/adapted from that paper; all quotation marks refer to it. The block diagram of SEER is presented in Fig. 1.

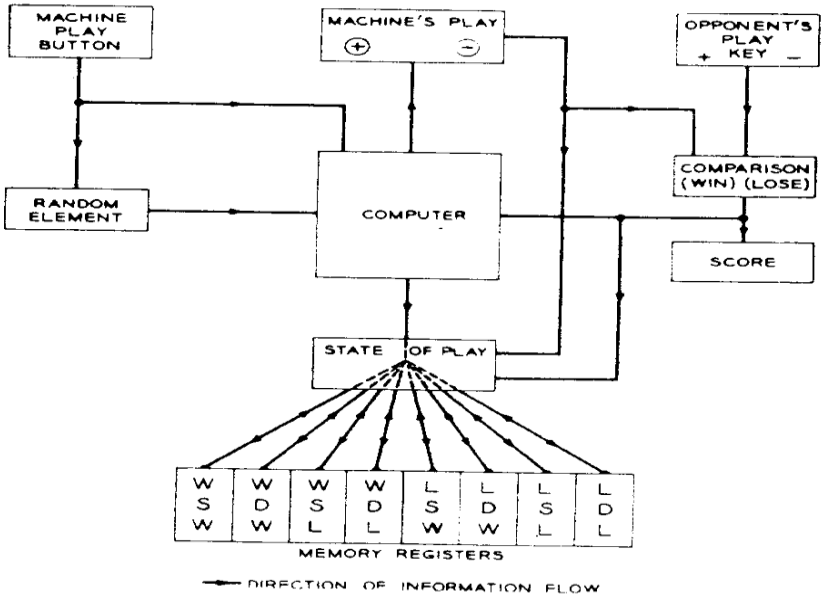


Figure 1. Hagelbarger's SEER block diagram [1]

"The "state of play" of the machine is determined by three things:" (presented in the order they appear in time)

- "whether it won or lost play before last play" (labelled W/L)
- "whether it played same or different last time" (labelled S/D)
- "whether it won or lost last play" (labelled W/L)

Based on those three bits of information (W or L, S or D, W or L) we have 8 different states of play, labelled in Fig. 1 from WSW to LDL.

For each state the machine's memory "stores two kinds of information:

a) Should the machine play same or different in this state in order to win?
b) Has the machine been winning in this state?"

"The a) part of the memory state is controlled by a reversible counter which starts at zero and can count up to +3 and down to -3. At the end of each play, if the machine should have played same, one is added to the counter. If it should have played different, one is subtracted. The counter will thus contain the number of times the machine should have played same in that state minus the number of times it should have played different. The stops at +3 and -3 in effect make the machine forget ancient history."

The b) part is implemented by "remembering whether the machine has won both, one, or neither of the last two plays in that state."

The machine's play is determined by the following rules:

"If the machine has lost the last two times in the present state, it plays randomly with equal odds on *same* and *different*.

If the machine has won one of the last two times in this state, it has three-to-one odds that it will follow the instruction in the a) part of the state memory.

If the machine has won both of the last two times in this state, the machine must follow the instruction in the a) part of the state memory."

After each play the machine updates the content of the state memory and the "state of play" and is ready for the next play.

For testing it, we have simulated the SEER machine in C++ language. An example of running our SEER implementation program (here against another SEER machine) is given in the following:

Game evolution (after 27 plays):

```
ME: HTTHHTTTTHHTHHHHHTTTTTTHTHT (SEER machine)
HE: TTHHTHHTTHTHHTHHHHHHHHHHTHH (opponent)
WH: 01010001110010111000000111 10 (win history - for machine)
CH:  10101000101100001000001111 1 (change history - for machine)
```

Machine's state at the end of the sequence:

```
Counts: -3  2 -2 -3  2  1 -1 -1 (the 0-7 state counters)
WinHR:   1  3  1  1  1  2  0  2 (the 0-7 state win history registers)
StateOfPlay: 6      (given by the gray 110 – WinOld_Change_WinNew)
```

In the previous example the win history register contains an integer value built from the 2 last win bits (previous win is MSB, last win is LSB). Because the win history register for `StateOfPlay=6` is 0, the machine will play randomly the next play.

3. Shannon's Mind-Reading Machine

The Shannon's Mind-Reading Machine (MRM) is described in [2]. As expected the machine looks a lot like SEER. The main difference is the fact that it tries to model the opponent's behavior; therefore the wins and changes are computed from the point of view of the opponent.

Like in SEER, MRM considers the 8 situations (states) based on the history from WSW to LDL (same labels as in SEER description). In each situation the player can do two things: "he may then play the same or differently". If, in one state, the same action of the opponent (play same or differently) appears in both the last 2 times, this behavior is considered to be repeated by the opponent's play and is used for the play of the machine. If this repetition of the opponent's play does not appear, the machine plays randomly.

From the description above we can notice that the machine does not use the StateOfPlay-based 7 state counters (hard to implement by relays) anymore, it uses only the StateOfPlay-based 2-bit change history register. This is in line with the statement made in [1]: "C.E. Shannon has built a machine using about half as many relays which follows a simplified version of the same strategy."

For testing it, we have simulated the MRM machine in C++ language. An example of running our MRM implementation program (here against another MRM machine) is given in the following:

Game evolution (after 27 plays):

```
ME: TTTTHTHHTTTHHHHHTTHHHTHTTTH (MRM machine)
HE: HHHHHHTTHTTTHTTHTTTTHTTTHTT (opponent)
WH: 11110111100101100011001010 1 (win history - for opponent)
CH:  0000010110011011000110011 0 (change history – for opponent)

ChgHR:  0  0  0  1  1  3  2  1 (the 0-7 state change history registers)
StateOfPlay: 1          (given by the gray 001 – WinOld_Change_WinNew)
```

In the previous example the win change register contains an integer value built from the last 2 change bits (previous change is MSB, last change is LSB). Because the change history register for StateOfPlay=1 is 0, the machine will assume the opponent will not change and therefore opponent will play same as previous – so that the machine will play same as the opponent, i.e. 'T'.

Obvious, in our C++ language implementations of SEER and MRM, we do not have the restrictions of the relay-based implementations of Hagelbarger and Shannon - i.e. we can maintain the whole histories for understanding/ debugging reasons – and also we can easily have a random number generator (the C library function `random`).

4. Experimental results (1) – SEER vs MRM

In [1] it is stated that "After much discussion an umpire machine was built which connected the two machines, and they were allowed to play several thousand games. The agility of the small machine triumphed, and it beat the larger one about 55-45."

In order to verify that statement, we made some tests in which the two machines play one against the other. Both machines are playing "same" (try to "match pennies"), it is the responsibility of the embedding environment to present to one of the machines the opposite selection of the other.

Because it is not clear from papers what was the goal of the machines: (1) to be the first winning a number of plays or (2) to achieve most wins from a fixed number of plays, we present the results for approach (2). This allows us to verify if the winner is the same after short or long games, therefore we present the results after 50, 100 and 200 plays for each game (winner in bold).

Table 1. SEER vs. MRM results

Game number	Results after 50 plays	Results after 100 plays	Results after 200 plays
Game 1	**26**-24	47-**53**	89-**111**
Game 2	18-**32**	38-**62**	77-**123**
Game 3	24-**26**	38-**62**	83-**117**
Game 4	23-**27**	49-**51**	88-**112**
Game 5	20-**30**	42-**58**	87-**113**
Game 6	**27**-23	48-**52**	95-**105**
Game 7	20-**30**	40-**60**	83-**117**
Game 8	22-**28**	45-**55**	88-**112**
Game 9	**27**-23	49-**51**	91-**109**
Game 10	24-**26**	46-**54**	91-**109**
Average	23.1 – **26.9**	44.2-**55.8**	87.2-**112.8**

We can notice from the results that, for the beginning of the games, SEER has some chances against MRM but, in long enough games, loses in all cases. The result can be explained by the fact that, at the beginning, state memory does not contain enough data for good predictions, so the results tend to be random.

In Fig. 2 we present the step-by-step evolution for the first 100 plays of Game 1. We can observe that at the beginning SEER has some advantage but MRM recovers and, in the end, it wins.

Even if it is not obvious what "the small machine … beat the larger one about 55-45" means, our average result (for 100 play games) of 55.8-44.2 in favour of MRM looks remarkable. The "55-45" statement suggests that the original games consisted of 100 plays. This is why we have chosen to present in figures the first 100 plays only.

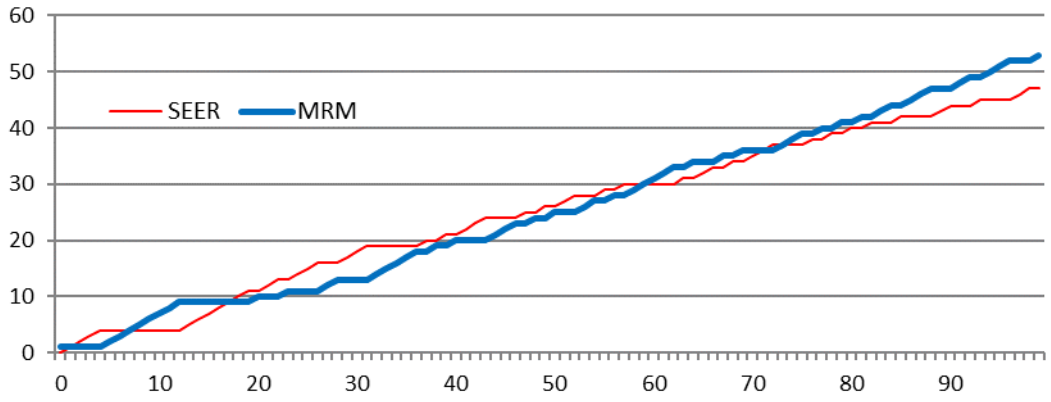


Figure 2. Score evolution for SEER vs. MRM Game 2 (first 100 plays)

5. A simple contextual predictor

It is obvious that both SEER and MRM suffer from the (hardware) limitation of history length that can be part of the stateOfPlay definition. Therefore we propose and test a simple contextual predictor (SCP), where such a limitation is overcome. The main idea is to find repeated patterns in opponents' play and suppose that he will repeat the same pattern over again.

In order to simplify the description of the algorithm and the implementation debugging we build a history string containing the description of each play (for the opponent's point of view) using the following symbols:

- `sw` = play same and win
- `sl` = play same and lose
- `dw` = play different and win
- `dl` = play different and lose

For a game evolution (after 33 plays):

```
ME: THHHHTHHTHHHHHTHHHTHHTTTHTHTHTTTT (SCP machine)
HE: HHTHHHHHTTHHHTTHTTHHHHHTHHHTTHHTT (opponent)
WH: 10100100010001001110011001001110O (win history)
CH:  011000010100101101000011001010O (change history)
```

(the win and change history are from the opponent's point of view)

we get the history string:

```
sldwdlslswslsldlswdlslsldwsldldwswdwslslswswdldlswsldlswdwswdlsl
```

We define the **context of length N** as **the last N** (**double) symbols** from the string. Because of the semantics of the build history string we search for matches only with a step of 2 individual symbols.

We start by searching in the history string the occurrences of the context string for the context length N=1. For each match we consider the following `sw,sl,dw` or `dl` symbol and increment its counter. Then we increase the context length and repeat the procedure until we get 0 or 1 total occurrences of the current context in the history string.

We then predict the next move of the opponent based on the longest context where the values for **`dl+sw`** (when we expect he will play *same*) and **`sl+dw`** (when we expect he will play *different*) are unbalanced (not equal). Certainly, if we have no match even at length 1 or, in all the analyzed contexts, the results are balanced we use a random play.

For the previous game example we have:

First step: length of context=1, context="`sl`"

```
sldwdlslswslsldlswdlslsldwsldldwswdwslslswswdldlswsldlswdwswdlsl
```

#next occurrences: `dl=3,sw=2,sl=3,dw=2, dl+sw=`**`5`**`, sl+dw=`**`5`**

Second step: length of context=2, context="dl

sldwdlslswslsldlswdlslsldwsldldwswdwslslswswdldlswsldlswdwswdlsl

#next occurrences: dl=0,sw=1,sl=1,dw=0, dl+kw=**1**, sl+dw=**1**

Third step: length of context=3, context="swdl

sldwdlslswslsldlswdlslsldwsldldwswdwslslswswdldlswsldlswdwswdlsl

#next occurrences: dl=0,sw=0,sl=1,dw=0, dl+sw=**0**, sl+dw=**1**

Because the total number of occurrences is 1, we do not further increase the context length and stop the analysis.

The longest unbalanced context is the one for length 3, so we predict based on it. The opponent will then choose to play **d** (different) relative to its last play '**T**', so the SCP predictor will predict also '**H**' (the SCP plays for "matching pennies").

The algorithm was described in "context" language, commonly used in data compression. In other research areas it is called "prediction by partial matching" (e.g. in [5]).

6. Experimental results (2) – SCP vs. SEER and vs. MRM

We also test our proposed SCP against SEER and against MRM, in the same approach like previous, i.e. in 10 games evaluated after 50, 100 and 200 plays. Table 2 and Fig. 3 present results of SCP against SEER while Table 3 and Fig. 4 present results of SCP against MRM.

Experimental results prove that SCP beats both SEER and MRM in long games. After 50 plays SEER and MRM win in few cases, but for longer games SCP recovers always.

Table 2. SCP vs. SEER results

Game number	Results after 50 plays	Results after 100 plays	Results after 200 plays
Game 1	**30**-20	**56**-44	**116**-84
Game 2	24-**26**	**53**-47	**111**-89
Game 3	25-25	48-**52**	**106**-94
Game 4	**30**-20	**63**-37	**124**-76
Game 5	**27**-24	50-50	**108**-92
Game 6	25-25	**53**-47	**105**-95
Game 7	**33**-17	**60**-40	**119**-81
Game 8	25-25	**55**-45	**110**-90
Game 9	24-**26**	**53**-47	**114**-86
Game 10	**29**-21	**55**-45	**113**-87
Average	**27.2**-22.9	**54.6**-45.4	**112.6**-87.4

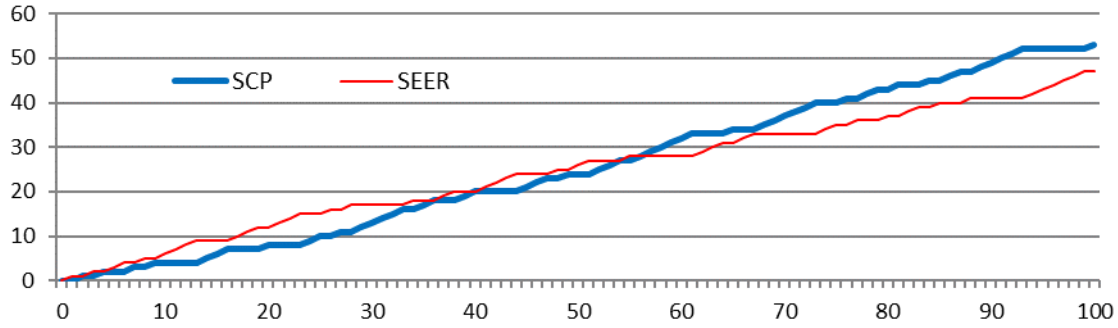


Figure 3. Score evolution for SCP vs. SEER Game 2 (first 100 plays)

Table 3. SCP vs. MRM results

Game number	Results after 50 plays	Results after 100 plays	Results after 200 plays
Game 1	**33**-17	**61**-39	**121**-79
Game 2	**31**-19	**61**-39	**115**-85
Game 3	23-**27**	**53**-47	**110**-90
Game 4	**33**-17	**65**-35	**120**-80
Game 5	**27**-23	**53**-47	**102**-98
Game 6	**35**-15	**61**-39	**117**-83
Game 7	**30**-20	**60**-40	**113**-87
Game 8	**28**-22	**53**-47	**112**-88
Game 9	24-**26**	**53**-47	**110**-90
Game 10	22-**28**	**53**-47	**117**-83
Average	**28.6**-21.4	**57.3**-42.7	**113.7**-86.3

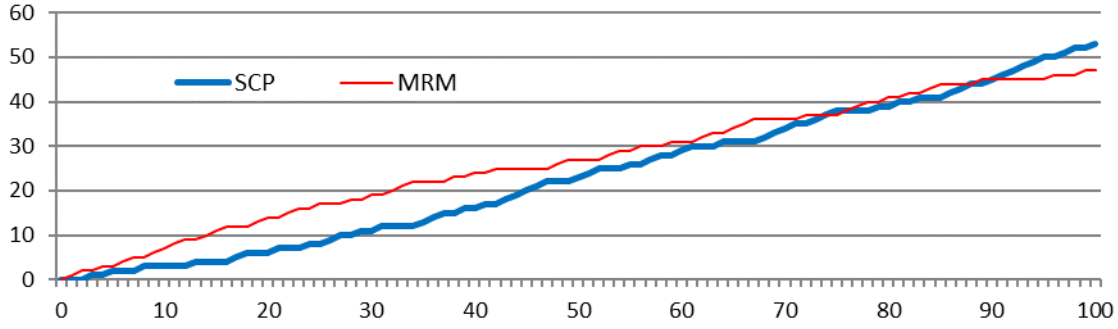


Figure 4. Score evolution for SCP vs. MRM Game 3 (first 100 plays)

The games selected to be presented in Fig. 3 and Fig. 4 are ones where, at the beginning, SCP loses, but the recovering process can be observed. The recovering at 100 plays is not by chance, it continues so that at 200 plays the win is consistent.

In order to evaluate what was the maximum length of context that was found during the play and how often it was used for prediction (i.e. it was not given up because it was balanced) we saved, for each prediction, the length of the longest context found (LLCF) and the length of the context used for prediction (LCUP). In table 4 we present the number of occurrences for each such pair, computed as sum for all the 10 SCP vs. MRM games analyzed.

Table 4. Number of contexts found and used

		Length of the context used for prediction (LCUP)						Total
		0	1	2	3	4	5	
Length of the longest context found (LLCF)	0	60						60
	1	48	86					134
	2	50	140	296				486
	3	32	82	183	638			935
	4	1	4	8	33	327		373
	5	0	0	0	0	0	12	12
Total		191	312	487	671	327	12	2000

From the previous table we notice that the maximum context length found and used (in 200 play games) is 5, and the longest context found is, in most cases, also used for prediction (corresponding to the numbers from the first diagonal).

In order to see how LCUP and LLCF evolve during play, we present in Fig. 5 their evolution during the first 100 plays of game 3 SCP vs. MRM (game also presented in Fig. 3). When the longest context found is also used for prediction the red diamond point is no more visible (it is overlapped by the blue square). As the game evolves longer contexts are found and used.

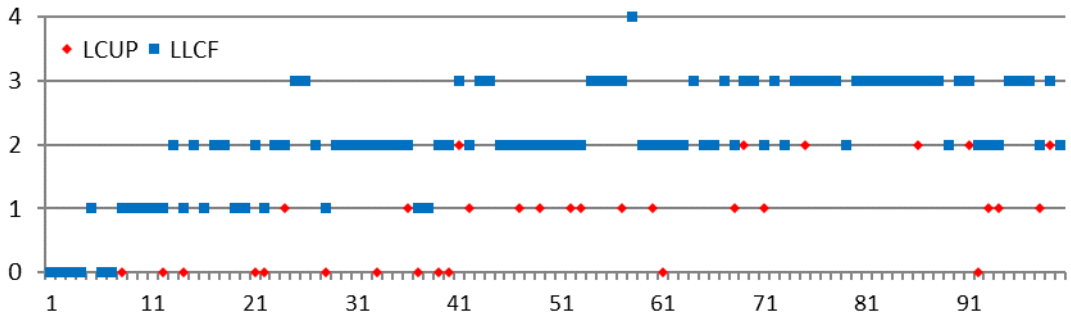


Figure 5. Evolution of LCUP and LLCF for SCP vs. MRM Game 3 (first 100 plays)

The SCP predictor was used also against some human players. We do not present such results because the results are very person dependent, and not statistically relevant. But, like with SEER and MRM tests done, the machine usually won. A rigorous set of tests for SCP (and maybe also for SEER and MRM) should be done in future, with enough games played so that the results became statistically relevant.

7. Conclusions and future work

The study of the SEER and MRM machines done in order to implement them in software reveals how remarkable was their design and implementation, considering the technological level at that time, i.e. implementing memory only by relays.

Our experimental results are consistent with the statement that MRM, although simpler, beats SEER. Also, the proposed SCP beats both SEER and MRM. This result is not surprising, previous hardware limitations being overcome

Another possibility that can be considered for SCP is to evaluate, in each context, the value **`sl+sw`** against **`dl+dw`** (that is, to consider the opponent less skilled, concerned only about changing or not in the current context). We assume that this strategy could be appropriate mainly for human opponents, but it needs to be tested.

In our implementation we have used for prediction always the longest unbalanced context. It could be also useful to use all the lower order unbalanced contexts (but weighted somehow differently). For context mixing we can use the approach from PAQ methods [3]. Unfortunately, because of the short length of the games, only simple contexts and mixing models can be considered.

The use of Hidden Markov Model HMM [4] looks promising for this application. After each play can we train the HMM model to learn the game history string (using a 4-symbol alphabet corresponding to the kw, kl, cw, cl cases) and therefore to model the behaviour of the opponent. After learning, the model is used to predict the next symbol and, from it, the opponent's move. After having the information of the next play the training and predicting process repeats for the entire game.

Our further research will focus on the idea of mixing different context length predictors and on the use of HMM for the matching pennies game.

References

[1] DW Hagelbarger. *Seer, A SEquence Extrapolating Robot*, IRE Transactions on Electronic Computers, page 1–7, March, 1956.
[2] Claude E Shannon. *A Mind-Reading(?) Machine*, Bell Laboratories Memorandum, March 18, 1953.
[3] Matthew V. Mahoney. *Adaptive Weighing of Context Models for Lossless Data Compression*, Florida Institute of Technology CS Dept. Technical Report CS-2005-16, 2005.
[4] Lawrence R. Rabiner, *A tutorial on HMM and selected applications in Speech Recognition*, Proceedings of the IEEE, vol 77, no. 2, 1989.
[5] Arpad Gellert, Adrian Florea. *Web prefetching through efficient prediction by partial matching*, World Wide Web, volume 19, pages 921–932, 2016.
[6] https://en.wikipedia.org/wiki/Matching_pennies
[7] https://cs.stanford.edu/people/eroberts/courses/soco/projects/1999-00/information-theory/ai.html
[8] http://william-poundstone.com/blog/2015/7/30/how-i-beat-the-mind-reading-machine

DOI: 10.2478/ijasitels-2020-0007

FitPi: Wearable IoT solution for a daily smart life

Sabin PRUNA[1], *Anca VASILESCU*[2]

[1]*Transilvania University of Brasov, Romania, sabinpruna@gmail.com*

[2]*Transilvania University of Brasov, Romania, vasilex@unitbv.ro*

Abstract

The extensive implementation of Internet of Things (IoT) solutions and its wide popularity to the public over the past decade enabled emergent applications that provide sophisticated, proactive health care solutions that can improve the quality of life of individuals. This work proposes an IoT architecture and implements a prototype solution that allows its users to improve their physical activity by collecting vital signs using wearables and other environmental and habitual information in order to monitor their activity and propose behaviors in a smart way that will allow them to achieve their preset goals. The focus is on the increased usability of the system employing refined solutions like voice recognition and smart visualization to enable its seamless use while offering an interoperable architecture that will enhance its flexibility. The prototype implementation offers a proof of concept evaluation of the proposed system, applying state of the art technologies and using existing hardware and popular gadgets.

.

Keywords: matching pennies, Hagelbarger, sequence extrapolating robot, Shannon, mind reading machine, contextual predictor

1 Introduction

The Internet of Things or IoT represents any system of interconnected devices which communicate with one another through internet-based solutions [4]. As the technology matured, numerous branches stemmed from it. These comprise from supply chain management, retail, surveillance to health care, data mining, research or numerous other directions which filled in the technological gaps of societal needs. A very useful and exhaustive survey on architecture, enabling technologies, applications and challenges of the Internet of Things is published in [6], and it mentions our targeted domain here, healthcare, as the first one ready to benefit from IoT potentiality. The diversity and variety of the IoT-based applications increases over time, and the interest of the software developers community intensifies accordingly to the point of writing platforms for ”rapid development of IoT applications” like the TinyLink system presented in [7].

A relevant area to this paper is the wearable market which consists of small electronic devices that can be worn on the body, like regular accessories, embedded in clothes or even implanted in the user's body. As more and more devices become connected to the internet via smartphones or directly through embedded chips, the wearable market has gained traction for allowing the user to be connected to the Internet spectrum with just something they wear [1]. Ranging from smart glasses to peacemakers, wearables are trying to enhance and improve how we carry on our daily lives. Due to progress in this area, this paper targets the use smartwatches, as mobile devices with a touchscreen display designed to be worn on the wrist, to provide a top view of one's daily activities wherever as long as the user has an Internet connection. For the implementation, a Fitbit [5] fitness tracker was used from the Versa lineup. This device already provides a plethora of fitness data, as this is its main purpose, however it lacks certain features related to lifestyle.

This paper aims to present an IoT-based system for supporting people who pay attention to their own healthcare and are moving from personal care towards a smart life. The purpose and, in addition, the novelty of our project are expressed in terms of providing a complete and unique project that covers the most important needs of a modern user by valuing the features of many hardware components and software technologies and tools. Having a portable hardware prototype and two interfaces, one website-based and one wearable, which are ready to communicate with the user about their current personal life status, the present FitPi is an original smartstyle solution. Moreover, it could be welcomed as a vector of IT&C for economics and life sciences since its implementation targets to improve the quality and standards of life from many perspectives, such as health, food, budget, working time log.

Our project, FitPi, has been developed mainly from a student perspective, as a wearable application that allows the user to quickly collect habitual and environmental information, and digest data using lists related to their logged records of personal health indices, money spent, food intake, as well as having measurements of the temperature in their choose enclosure and others. An interesting "framework for designing material representations of physical activity data" is presented in [9] as Shelfie, and it is connected with our project by the on-screen visualizations of physical activity data.

There are, evidently, numerous powered devices, at a higher price range, that integrate into different operating systems ecosystems and there are also options with a more sensible data-driven functionality such as the eButton proposed in [16]. All these solutions can measure many health data, as our FitPi project does as well, but, FitPi is distinguished by offering one application for processing the collected data and two synchronized solutions for interfacing these data. Therefore, the FitPi solution has many important characteristics, for example, it is cheaper, it is focused on daily and accurately tracking physical data, and it is smart-oriented in terms of involving its user in stepping to smart life.

Considering other similar Fitbit-based applications, we have remarked the Fitbit Garden app [3] designed to enable children to engage in more physical activity. Beyond this similarity of having the daily physical activity as a common goal, the FitPi app will provide the Fitbit collected data to a broader range of lifestyle features, from the wearable and weather station to the website and API. The IoT-based architecture adopted for developing the FitPi project with AWS cloud support [2, 13] has passed the test of analyzing other solutions, for example, the IBM- based weather station described in [10]. From the hardware point of view, our FitPi prototype accomplishes the necessary IoT infrastructure using boards, sensors, actuators, displays, as well as other interconnecting components described in the next sections. Combining the prefix from Fitbit Inc. name and the suffix from the Raspberry Pi board name, we have the name FitPi for our original project.

2 Platforms and tools as FitPi background

An important part of this paper context is around the notion of microservices. Such services are an approach to application development in which a large monolith product is built as a suite of modular services. ”Each module supports a specific business goal and uses a simple, well- defined interface to communicate with other sets of services” [14]. In SOA (Services Oriented Architecture), services use protocols describing how to transmit and parse messages using description meta-data. This meta-data describes both the functional features of the service and the characteristics of its service quality. Services-oriented architecture aims to allow users to combine large pieces of functionalities to form applications that are built exclusively from existing services and combine them in an ad hoc manner. A service presents a simple interface for the applicant that abstracts the essential complexity. Additional users can also access these independent services without knowing their internal implementation. A service consists of two parts: the Web Part where the API, models, and everything needed for communicating with a client, and the database section for persisting user data that contains only a Docker file so that the service can be containerized.

From the practical point of view, our results described here are based on the Fitbit submission. Fitbit Inc. is an American company headquartered in San Francisco, California. Its products ”are activity trackers, wireless-enabled wearable technology devices that measure data such as the number of steps walked, heart rate, quality of sleep, steps climbed, and other personal metrics involved in fitness.” [5] FitbitOS represents the patented operating system of the Fitbit company running on devices of different types, for which one can develop apps using the specific SDK.

Amazon Web Service has provided another significant support for our approach as it affords the cloud infrastructure needed to host applications like ours. The specific services can be accessed locally, but to be used by the client applications, they must be published in an online environment, facilitating access to them. For the FitPi project, we chose the Amazon Web Services ecosystem as it is one of the most popular platforms in the area, using the EC2 (Elastic Compute Cloud) and AWS (Amazon Web Services) Lambda to accomplish our goal [2, 13].

In order to have these software components physically interconnected by a hardware system that is capable of capturing ambient data from a room, a Raspberry Pi Arduino Uno board, a DHT-11 sensor, a microcontroller ESP8266, and an LCD were used. The cloud provider was chosen to be AWS. However, the Wi-fi chip has not been strong enough to pass the AWS authentication test with higher security, all these consequently required also to use the Raspberry Pi as an intermediary node between the station and the cloud.
As a branch of Amazon Web Services, AWS IoT allows the integration of IoT devices with the ecosystem and Cloud made available by Amazon. AWS IoT works using the concept of MQTT, each device having topics for specific operations like reading, updating, accepting or rejecting queue data. All these topics are incorporated in the Shadow concept that represents the device state at the current moment.

The architecture entails two UI applications, one for the wearable which will be the main focus of this paper, and a website which aggregates the data from both the tracker and the mobile application while also making it editable. As both applications require a common backend to fetch and post data, a decision was made to use microservices developed in Python. Therefore, the business logic has been treated as a service black-boxed from the users and exposed via a public API to the related applications. The services were hosted on AWS to provide reliability and scalability, making communication viable via HTTP protocols. Accordingly, the development time was halved and, most notably, the architecture had a common backend which resulted in uniformity. Firstly, the services were developed using a combination of the Python and C# programming languages in a domain-driven design workflow. The data was persisted using a document-oriented database (MongoDB) which was encompassed with the business logic layer and the API in a containerized image through Docker so that the service may run independently of the machine it resides on. Basically, AWS allows the services to be hosted with minimal costs and virtually no downtime, providing availability worldwide and scalability on the go, based on demands, in case of the user base would ever increase.
Moreover, a second IoT system was designed for the interest of this paper, namely a weather station, powered by Arduino and based on temperature sensors, a Raspberry Pi device and the Cloud, as we will describe in the next sections.

3 FitPi system components and features

This section is to present our FitPi system at the developing level: system architecture, components, internal applications, functionalities and features, workflows and specifically involved microservices.

FitPi is a system based on SOA design principles, requiring multiple frontends to run on. As a result, a backend capable of acting as a black box has been developed to assure the communication with many clients, regardless of their origin. It uses two client applications, as follows: (a) a website frontend that runs both in the browser, behind the scenes, and on the IoT device (Arduino and Raspberry) and (b) a wearable frontend available on Fitbit devices for the Versa and Ionic lineups, these versions being among the first running over the FitbitOS. These two components and their appropriate interconnecting modules are illustrated in Figure 1 and presented in detail in the next paragraphs.

The website application follows the ASP.NET MVC5 framework and the Model-View-Controller paradigm, the controllers being responsible for receiving incoming requests. After applying the required business logic, if any, the controller produces the desired model with the data received from the backend. This model is then applied to the desired view, which is finally returned to the user. To support reusability, extensibility, and also the separation of concerns, a 3-layer architecture was correspondingly developed for the website application, as follows:

1. Presentation layer as the MVC layer, i.e. controllers, views, models were created at this level;

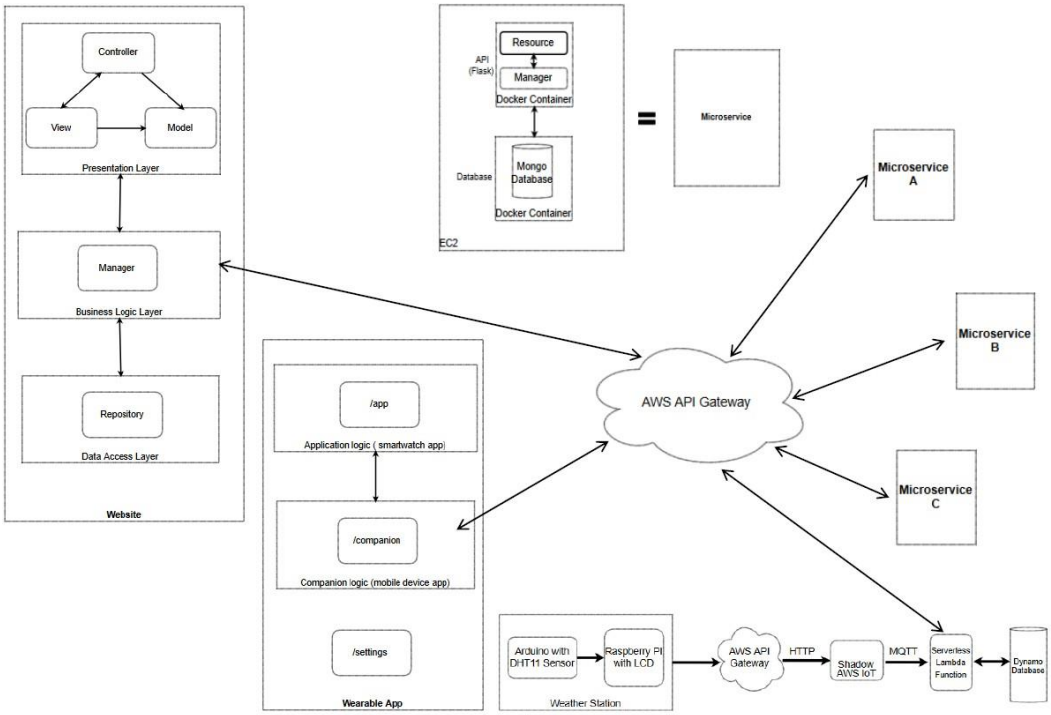


Figure 1: Application architecture

2. Business logic layer for residing the specific Services for Managers in order to keep the controllers light; for applying and housing the business logic, specific constructs are called by the controllers, acting as a black box to the Presentation MVC layer;

3. Data Access layer consisting of data manipulation processes and acting as a black box or an interface for working with the data source; this is where the repositories live and call on our data storage solution.

For the website, the Data Access layer was lightweight, in the sense that, owing to the shared backend, we had external APIs to call in the business layer. This solution exists mostly to persist information about the user's current state.
The second application, namely the wearable frontend, was created using the architecture required by FitbitOS. This application consists of a component running on the smartwatch and a mobile component running on the smartphone as a companion. Both communicate seamlessly with each other, acting like one. The wearable application structure is presented as follows:

• */resources* as a folder keeping all external resources like pictures, styles, or frameworks required by the app as well as the landing page (index.gui) which acts as a master container for the entire wearable app user interface;

• /app as a folder housing the application logic that runs on the smartwatch device; the files in here are responsible for interacting with the master view to display user-requested views as well as communicating with the companion app to fetch the data;

• /companion containing the companion logic side of the app which runs on the mobile device; a specific socket communication solution has been assumed here to ensure the interaction with the watch device and make the JavaScript fetch API to communicate with the backend, i.e. sending and fetching data through HTTP endpoints;

• /settings as a separate page of the companion app that establishes application-level configuration, for example, the application colour map.

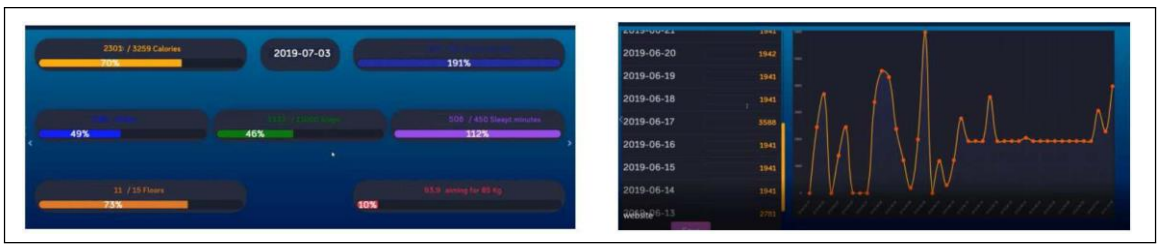


Figure 2: Daily activity module - dashboard and details on steps

Besides these two client applications, in Figure 1 we have an extra third client in the sense of a weather station that accomplishes our FitPi ecosystem. This

station is a hardware device based on two components: (1) an Arduino board that tracks the temperature data and (2) a Raspberry Pi that is connected to a small LCD where the default browser is shown in full screen with the website version of the app. The website has integrated the JavaScript speech- to-text engine functionality to allow the users to communicate with the system via their voice, similar, in this regard, to Amazon Alexa or Google Home.

The step forward to having a real software architecture based on these frontend layers was in the direction of using SOA to add a suitable shared backend. This backend application is user-oriented and responsible for doing the heavy lifting, applying complex business logic and calculations and acting as an interface for consuming the client applications. Therefore, our FitPi product performs as an original and complete solution for managing user's daily activities and allowing easy access to information related to the personal progress achieved. Current backend implemented functionalities are described as follows.

Daily activity is for manipulating data tracked by the wearable device, considering: the number of steps performed in a day, calories consumed, floors climbed, duration of intense activity, body weight, hours of sleep, distance travelled and heart rate. These are available in a daily format, with an appropriate overview, but also a detailed management tool for each element throughout the use of the application is available. We add in Figure 2 two captures from this module, one from the activity dashboard and the other with the steps target presented in details.

Nutrition is for keeping a micro- and macro-nutrient-oriented view upon the daily food in- take. Data on consumed foods such as calories and macro-nutrients like carbohydrates, fats, proteins, vitamins, sugars are managed. Again, data are presented in a daily format, with additional information on the user's monthly progress.

Budgeting is an original part for tracking the spent money, being focused on categorizing transactions made in budgets, compared to the transaction orientation of popular applications on the market, and setting thresholds for each of them. It is possible to add new budgets, in the monthly context and related transactions. At the same time, FitPi informs the user about exceeding the thresholds and about the percentage remaining in each budget.

Timetable is an educational topic for keeping track of the users' timetable, for example, classes at university. Although anyone can use the application, there is this custom section dedicated to students working parttime. Thus, FitPi presents, in an easily understood format, the courses, laboratories and seminars to which the student must attend, mentioning the current time at which they must be active.

Work-log is a work-time management tool for monitoring the hours done as usual work. As we already mentioned before, should one be working under a flexible schedule, that user may want a way to record the number of hours they spend at their day job, whilst also keeping track of having the hours required monthly.

Weather is an extra functionality of our project for persisting data about temperature for both the room and the city where the weather station device

is positioned. FitPi provides information about the temperature and humidity of the room where the physical system is located, as well as providing weather forecast data over several days.
Applying an SOA-based solution comes with powerful extensibility as each functionality or domain is isolated in its own microservice. Such architectures may enable multiagent solutions leading to emergent applications [12]. This way, in case of a business need for adding another requirement, a separate microservice can be created without affecting the existing functionality. The main microservices benefit is that each one can be created with a different tech stack, the best way to solve the services use case. Most services were created in Python programming language so that the Flask framework was used for fast bootstrapping of the exposed API. A service follows the three-layer architecture as specified in the frontend description with the attached details:

1. Presentation layer, where the domain models are present and the API side of the service is created; for this particular use case, we have resources (a resource construct consists basically of the endpoints of particular functionality, similar to a controller); for simple services, this is mostly the only layer used, owing to the simple CRUD business logic;
2. Business logic layer, where the managers are present with the role of encompassing business logic if any;
3. Data access layer, where we house the data storage solution, in this case, a NoSQL database: MongoDB; from a developer point of view, this could be considered as the other visible structural part of the microservice since there is a section for the API and another one for the mongo database.

Having these separate sections, namely API part and Database part, which now need to communicate with the frontends, we have accordingly used the built-in solutions provided by Amazon Web Services. Each microservice section is placed inside of a docker container, allowing the microservice to run on any machine that has that docker installed. We have used EC2 machines to host our microservices in the cloud, and these machines were then secured by hiding them behind Amazons API Gateway, which is responsible for load balancing and validating the requests. Only the frontend applications can access the backend endpoints using token validation.
We conclude this system description section by adding both a principal workflow and one
for the weather station part. The principal workflow example could be:

1. the user requests a specific page
2. a frontend controller or a function from the wearable app section receives this request
3. the controller calls the manager from the business logic to get the data
4. the controller either calls the data access layer or makes a call to a microservice for the data

5. the API gateway intercepts this request and validates it; in case of approval, it is then forwarded to one EC2 machine which is specific to the microservice called; load balancing part decides to which machine the request is forwarded (as there are several machines for each microservice), also considering the performance reasons
6. the EC2 machine forwards the request to the docker container
7. the docker than forwards the request to the resource of the microservice (API section)
8. the resource calls its manager if needed or uses the mongoDriver to communicate with its database or fetches the data from an external data source
9. once the service has finished manipulating the date, a roundtrip is done all the way back to the frontend application
10. the controller receives the data, puts it in a model and renders the required view for the user.

The weather station part has a similar but slightly different workflow as it uses AWS IoT, a SaaS functionality created by Amazon. For example, such a workflow could consist of next steps:

1. the Raspberry Pi sends the data collected by the Arduino board to AWS via the AWS Gateway
2. this is then sent to AWS IoT and stored in the device's shadow (this is a concept in AWS
1. which basically means the state of the device)
2. AWS IoT uses an MQTT protocol as the shadow acts as a message bus; we have a serverless function (in AWS this is called a Lambda function) that subscribes to this message bus; on each shadow change given by a new temperature data, the function stores this in a NoSQL database, DynamoDB
3. the function also exposes an endpoint for retrieving this data when a frontend application demands it.

4 User perspective on FitPi as an IoT system

Apart from being a modern and up-to-date IoT system, our FitPi project is significant and valuable support for any user interested in monitoring their everyday life and publishing the daily activities results on a social network, here Twitter. This feature was achieved by implementing a TwitterBot [15], and Python programming language has been used because it provides modules ready to use the Twitter API quickly, acting as a black box over low-level implementation. Since it has been observed that text messages sent to the user are not effective in keeping them motivated beyond a week, perhaps having the information published on a social media platform would motivate the user to keep trying to improve their lives and continue using the FitPi

product to pursue their goals. For example, in Figure 3 we have a page of statistics prepared for keeping the user's motivation at a high level.

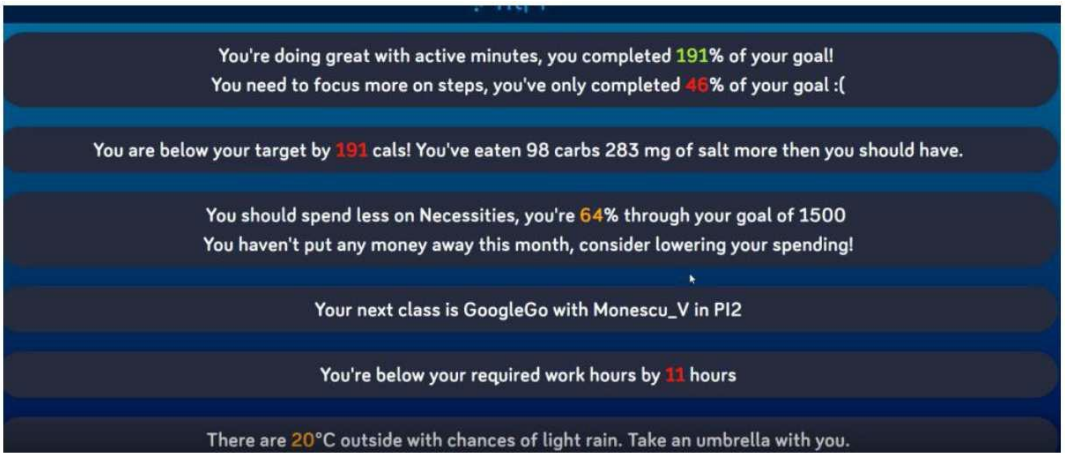


Figure 3: Statistics - user view

Figure 4: Nutrition module - website and wearable views

As a Fitbit-based project, our system has access to all the physical activity data which are stored in the Fitbit data service. The user simply must carry on with their daily life, and the tracker will monitor all that is related to steps taken, calories burnt, floors climbed, and some other. Data that are related to weather is either retrieved from OpenWeatherMap when the user wants to know more about temperature from other areas or by using the provided weather station which records information about temperature and humidity every one minute. For monitoring food consumption, we used the database provided by MyFitnessPal (MFP). Our users can either log their meals into this MFP as it is synchronized with FitPi or directly into the FitPi website. The timetable has data directly fetched and calculated from the Transilvania University of Brasov timetable. Other sections have their own in-house database so that when a user logs something into the website, data is stored and can be displayed both in the website and the wearable application. This mirroring solution for the specific case of the Nutrition data is represented in Figure 4.
Each of our system modules is implemented with respect to the IoT approach, and it succeeded by featuring the specific aspects of security, privacy, scalability and other given self- properties in different manners, as we highlight in the following.

Our system faces the security problem from many perspectives, for example. The services were hosted on AWS to provide a higher level of reliability and scalability (making communication viable via HTTP protocols), and the configuration files contain both private keys and other data requiring higher security. Moreover, we accessed the AWS console for EC2 services and chose a Linux AMI (Amazon Machine Image), being optimally a hosting environment for services. In the management console, we created a thing to which we assigned a certificate, a public key and a private key (used in the Raspberry python script), so data transmission can be done securely and encrypted via HTTPS.

As we have already mentioned before, the service that deals with physical activity has numerous managed data elements such as the number of steps taken in a day, calories consumed, floors climbed, duration of intense activity, body weight, hours of sleep, distance travelled and heart rate. The uniqueness of these data is given by their retrieval from a compatible physical activity monitoring device, for example, Fitbit Versa. The fitness tracker could be synchronized with any mobile application that sends data to the Fitbit Cloud. Using the access provided by the Web API, as a system developed by Fitbit to access data from tracking devices, we were able to fit it out in our architecture and synchronously obtain the data needed for the FitPi domain of interest. Since these data need to be retrieved at regular intervals to synchronizing the information between FitPi and fitness tracker, we used Flask and APScheduler (Advanced Python Scheduler), a Python tool that can configure jobs to run at regular time intervals on a separate thread, thus maintaining the functionality of the services in Flask. The APScheduler tool is also used to synchronize data from FitPi with data from MFP for the current day. Data is stored in an aggregate document, with information about all the tables in four options defined by the MFP (breakfast, lunch, dinner and sweets).

The nutrition service manages data on consumed foods such as calories and macro-nutrients, like carbohydrates, fats, proteins, vitamins, sugars. This data comes from the MyFitnessPal database, the food products being registered in the application provided by MFP arriving in the cloud. Because the MFP access via API is private and it is not possible to access it by an individual developer, we used a web scrapper that directly parses the desired data (if it is in the public domain) from the MyFitnessPal website based on the user id and successfully passed the privacy problem.

5 Discussion and future work

The more these new branches of technology continue to emerge, the more humans become involved in personal wellbeing, and the abundance of data their smart devices provide them allow a deeper insight into their lifestyle. Even

if are there countless health applications available on a multitude of wearable platforms, this FitPi project aims to complement these apps in the direction of daily activity monitoring based on smart solutions.

We may conclude now that the lifestyle management system, FitPi, targets to cover most of the daily activities adopting an intelligent approach. The FitPi website frontend assembles all the information related to the physical activity using real data and presents it in an user-friendly and easy-to-understand format. By means of the voice assistant, it becomes a reactive system that notifies the user about the progress made in reaching the personal set goals, as well as about a possible decline in the effort made, thus determining the user to better their approach. This improvement is also supported by the other functionalities of the application, such as monitoring the food intake and finance management. At the same time, FitPi facilitates certain elements of the user's life and, by suggesting possible actions to be taken accordingly to the recorded vital signs values, it performs like a self-adapting IoT system. Also, by developing an application available on a wearable device, the user could access the FitPi data transparently, which increases their motivation to maintain a healthy lifestyle.

For now, the current FitPi version is a single-user version, the developer. The biggest step that FitPi has to do in the near future is to create and integrate a service that allows multiple users to register, for example, using ASP.NET Identity support. This also involves versioning existing services, implementing new specific security features and personal data protection by saving unique data to each user account.

FitPi allows users to access internet services through a browser or Fitbit devices. Still, to grow their demographic and fully satisfy existing users, there is a need for native applications on mobile devices. Thanks to the adopted service orientation architecture, this can be achieved much faster - already having the backend functionality, it requires only the creation of native client applications to use these services.
Last but not least, integrating the voice system in all areas of the application is a valuable step forward in the direction of having an intelligent notification system. Now, FitPi offers limited support for data aggregation, and this allows the user to be aware of his or her progress or regress. But, by implementing an independent active notification system based on o set of recommended actions that the user should take, we add artificial intelligence to the system and greatly improve the user's success rate.
Once these extensions are implemented, we can say that FitPi is not only an alternative to current solutions, bringing various improvements, but a complete IoT system, able to positively influence its users every day, inspiring them in a smart way to reach their lifestyle goals.

References

[1] Aloi, G., Caliciuri, G., Fortino, G., Gravina, R., Pace, P., Russo, W., Savaglio, C. Enabling IoT interoperability through opportunistic smartphone-based mobile gateway, Journal of Network and Computer Applications, March, 74-84, ISSN: 1084-8045, 2017
[2] Amazon Web Services homepage, https://aws.amazon.com/. Last accessed 9 November 2020
[3] Amresh, A., Lyles, A., Small, L., Aloysius Gary K., FitBit Garden: A Mobile Game Designed to Increase Physical Activity in Children, Proceedings of the 2017 International Conference on Digital Health (DH'17), 200-201, Elsevier, London, United Kingdom, ISBN: 978-1-4503-5249-9, 2017
[4] George Eleftherakis, Dimitrios Pappas, Thomas Lagkas, Konstantinos Rousis, Ognen Paunovski, Architecting the IoT Paradigm: A Middleware for Autonomous Distributed Sensor Networks, International Journal of Distributed Sensor Networks (IJDSN), 139735:1-139735:17, ISSN: 1550-1477, 2015
[5] Fitbit Inc. homepage, https://dev.fitbit.com/build/reference/. Last accessed 20 September 2020
[6] Arindam Giri, Subrata Dutta, Sarmistha Neogy, Keshav Dahal, Zeeshan Pervez, Internet of things (IoT): a survey on architecture, enabling technologies, applications and challenges, In Proceedings of the 1st International Conference on Internet of Things and Machine Learning (IML '17), Association for Computing Machinery, New York, NY, USA, Article 7, 112, ISBN: 978-1-4503-5243-7, 2017
[7] Gaoyang Guan, Borui Li, Yi Gao, Yuxuan Zhang, Jiajun Bu, Wei Dong, TinyLink 2.0: integrating device, cloud, and client development for IoT applications, In Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20), Association for Computing Machinery, New York, NY, USA, Article 13, 113, ISBN: 978-1-4503-7085-1, 2020
[8] Kasun Indrasiri, Prabath Siriwardena, Microservices for the Enterprise: Designing, Developing, and Deploying, APress Media LLC, Springer Science + Business Media New York, ISBN: 978-1-4842-3857-8, 2018
[9] Rohit Ashok Khot, Larissa Hjorth, Florian Mueller, Shelfie: A Framework for Designing Material Representations of Physical Activity Data, ACM Transactions on Computer-Human Interaction, Vol. 27, No. 3, Article 14, Publication date: May 2020, ISSN: 1073-0516, 2020
[10] Kishore Kodali, R., Mandal, S., IoT Based Weather Station, International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 680-683, IEEE Kumaracoil, ISBN: 978-1-5090-5241-7, 2016
[11] Michael Papazoglou, Web Services and SOA: Principles and Technology, Pearson Education Canada, ISBN: 978-0-2737-3216-7, 2012
[12] . Paunovski, O, G. Eleftherakis, and A.J. Cowling, Disciplined Exploration of Emergence using Multi-Agent Simulation Framework, Computing and Informatics, 28(3):369-391, ISSN:1335-9150, 2009
[13] Sarkar, A., Shah, A., Learning AWS - Second Edition, PACKT Publishing House, ISBN:978-1-7872-8106-6, 2018
[14] SHIELDX homepage, https://www.shieldx.com/. Last accessed 20 September 2020
[15] Twitter development support homepage, https://developer.twitter.com/en. last accessed 11th November 2020
[16] Yicheng Bai, Chengliu Li, Yaofeng Yue, Wenyan Jia, Jie Li, Zhi-Hong Mao, Mingui Sun, Designing a wearable computer for lifestyle evaluation, 38t h Annual Northeast Bioengineering Conference (NEBEC), 93-94, IEEE, Philadelphia, PA, ISBN: 978-1-4673-1140-3, 2012

sciendo

DOI: 10.2478/ijasitels-2020-0009

Modeling Electricity Consumption and Production in Smart Homes using LSTM Networks

Miroslav-Andrei BACHICI[1], Arpad GELLERT[1]

[1]Computer Science and Electrical and Electronics Engineering Department, Faculty of Engineering, "Lucian Blaga" University of Sibiu, Romania
{miroslav.bachici, arpad.gellert}@ulbsibiu.ro

Abstract

This paper presents a forecasting method of the electricity consumption and production in a household equipped with photovoltaic panels and a smart energy management system. The prediction is performed with a Long Short-Term Memory recurrent neural network. The datasets collected during five months in a household are used for the evaluations. The recurrent neural network is configured optimally to reduce the forecasting errors. The results show that the proposed method outperforms an earlier developed Multi-Layer Perceptron, as well as the Autoregressive Integrated Moving Average statistical forecasting algorithm.

Keywords: electricity prediction, Long Short-Term Memory, smart home, energy management system, photovoltaics

1. Introduction

The worldwide energy consumption has seen a substantial growth over the past century and ensuring that everyone has access to energy is still a major and ongoing challenge. In this work we will focus on the electricity consumption and production at household-level, however the modelling of a larger area is also of interest, as we can see in [11] and [12]. In order to cope with the environmental degradation caused by the usage of fossil fuels, the demand of alternative, "renewable energy" sources are undoubtedly increasing. Solar energy is the most important source of this kind, mainly because of the Sun's lifespan which is approximately 5 billion years. Therefore, taking into consideration human timescale, it is safe to say that this is an inexhaustible source of energy. Photovoltaic solar energy, one of the Solar energy's subcategories is captured by photovoltaic panels in the process of producing electric energy. The panels are composed of multiple arrays of photovoltaic cells, usually made of silicon, which in contact with the light from the Sun causes agitation in the electrons of the cells, and therefore, produce

electricity. To analyze how efficient this method of producing electricity in a household is and whether it is feasible or not, different prediction methods have been applied in order to evaluate both the energy production and consumption on such a smaller scale.

In the current stage of the research, we analyze the efficiency of a Long Short-Term Memory (LSTM) recurrent neural network (introduced by Hochreiter and Schmidhuber in [10]) in forecasting the electricity consumption and production in a smart house. In our use case, the electricity production is assured by two photovoltaic panels. We already evaluated some stochastic methods (Markov chains, prediction by partial matching) and some statistical algorithms (ARIMA, TBATS). Other methods will be further implemented and evaluated. We are interested in determining the method with the lowest prediction error, which then will be integrated into a smart energy management system whose role is to keep a balance between the consumption and production of a certain household.

The rest of this paper is organized as follows. Section 2 presents the relevant related work. Section 3 describes the proposed LSTM-based modelling of the electricity consumption and production in a smart house. Section 4 discusses the evaluation results in a comparative view with other existing methods. Finally, Section 5 concludes the paper and presents some further work possibilities.

2. Related Work

In [3], Stefan Feilmeier presented the software architecture of the FENECON energy management system. The author recorded the electricity production of two photovoltaic panels and the loads on three phases in a household during five months, with a step of 5 minutes. The obtained dataset was used later for evaluations in different researches (including our current work). In [3], the dataset was used to evaluate a Multi-Layer Perceptron (MLP) in predicting the electricity consumption and production. The author reported a mean absolute error of 211.07 Watts. Other works which are applying neural networks to predict the electricity consumption of buildings are [13] and [9]. Sensor-based forecasting of the electricity consumption in a large entertainment venue with neural networks and support vector regression is investigated in [8].

In [4], we adapted a Markov model to be able to work with the electricity consumption and production levels and to provide a short-term prediction of the upcoming levels. To reduce the state-complexity of the model, we downscaled the input data by dividing all the values to the same scaling factor. The value returned by the model is upscaled to obtain the predicted electricity level, by multiplying it with the same scaling factor. The scaling factor, the length of the input vector, as well as the order of the Markov Model, were the main parameters varied in the experiment. The mean absolute error, measured

on the same dataset which is used in our current work, was 34.43 Watts. A stride predictor and a hybrid predictor composed of the Markov model and the stride predictor have been also evaluated. However, the stride predictor proved to be inefficient and, thus, the more complex hybrid predictor could not provide better results than its Markov prediction component.

To predict the next values of time series, several machine learning techniques can be considered as good candidates, a design pattern for an efficient implementation being presented in [6]. The efficiency of the prediction model is highly dependent on the type of the input data. Therefore, we evaluated different prediction methods with the goal of finding the most appropriate one for the electricity consumption and production datasets. In [5], we used statistical methods to predict the electricity consumption and production. The Autoregressive Integrated Moving Average (ARIMA) algorithm can describe time series based on the past values or lags and the forecast errors. The other statistical method is based on Trigonometric Seasonal, Box-Cox Transformation, ARMA Residuals, Trend and Seasonality components (for short TBATS), is decomposing seasonal time series into trend, seasonal and irregular components. The evaluations have shown a mean absolute error of 198.27 Watts for the ARIMA algorithm and 73.62 Watts in the case of the TBATS model.

In [14], Monteiro et al. evaluate short-term statistical prediction of photovoltaic electricity production. Two models are proposed: one of them is analytical and the other one is using a MLP. The prediction relies on weather forecasting tools focused on the location of the photovoltaic plant, as well as on hourly recorded photovoltaic electricity production. The analytical model computes the sky irradiation based on hourly radiation forecasts and adjusts it with irradiation attenuation index and photovoltaic production attenuation index. The neural network was selected and configured using genetic algorithms and is using weather forecasts as input information. The proposed models were evaluated and compared on the same data collected from a grid-connected photovoltaic plant. The authors concluded that the two models have similar results, and both are usable in the sight of selling electricity to the markets.

In [2], Fan et al. evaluated the hybrid prediction through data mining techniques of the next-day energy consumption in buildings. The proposed method has three steps. In the first step, an outlier identification and removal is performed. In the second step, a recursive feature elimination is applied in order to use the optimal inputs for the eight different predictors. In the final step, an ensemble model is optimized for the predictors through a genetic algorithm. The authors concluded that the proposed ensemble model can be efficient in fault detection.

3. LSTM-Based Forecasting of Electricity Consumption and Production

This section focuses on describing the proposed model of this paper and the functionalities that it has. For our paper, we implemented and used an LSTM, which is a model of Recurrent Neural Network (RNN). RNNs represent one of the most optimal choices when working with data organized in time series models. Their work principle is based on combining nonlinear activation functions in a recurrent structure, which makes prediction possible and provides improved prediction accuracy, as stated in [1]. In contrast to the standard Neural Networks, which are usually represented using feedforward architectures, RNNs allow the information to be transferred both forward and backward, with the help of their feedback connections. Therefore, these neural networks benefit from the ability to work with dynamic data. An analysis regarding the applicability of RNNs for prediction purposes is presented in [15].

An RNN can have multiple layers, steps or stages. Their work principle is described in Fig. 1. Each stage from the above schema corresponds to a given time T. The RNN at the time T+1 will use the RNN from the time T as one of its inputs. Each stage will send its output to the next stage. The key mechanism which makes the RNN to work well is represented by the hidden state information propagated from a certain stage to the next. The hidden state works as a memory capable of retaining information of the current stage. A layer from the RNN is processing the input data and is returning its internal state which is going to be used as an input in the next stage. More specifically, each stage is trained to transform the target sequence from a moment T into the input sequence but with a T+1 timestep offset. To achieve this, a backpropagation algorithm is used. The value of the loss obtained for each parameter is used to change the parameter values in the reverse direction with the purpose of minimizing the loss. As this movement is time based, each timestep contains its own loss value. In the process of modeling the dependencies between value sequences, the gradient of the timestep T depends on the gradient of the timestep T-1 and so on, and because of this, the further we progress with the timesteps, the gradient of the latter timestep matters less and less. This is known as the "vanishing gradient problem" whose effect is that the network cannot learn from long term dependencies, because the gradients of the early stages become smaller. LSTM networks are a solution to this problem.

LSTMs are RNNs that work with data that varies through time or sequentially, like language, stock market prices, weather recording sensors, etc. The way they work is similar to other RNNs, by using the outputs of a layer at a timestep T as inputs for the same layer at a timestep T+1. They have a component that acts as a memory which helps to transfer information learned at the timestep T to the next timesteps, and they can also forget irrelevant information from the preceding state and update the current state, allowing

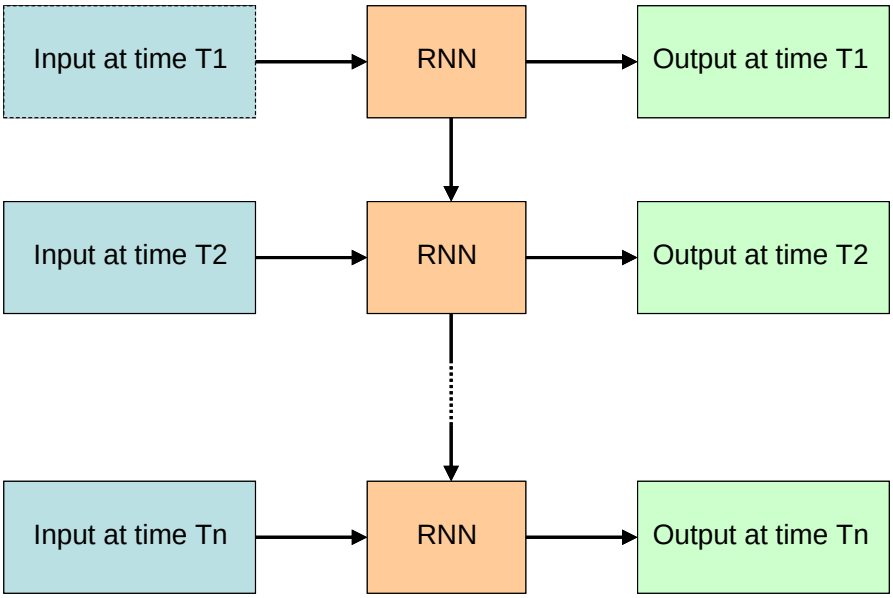


Figure 1. RNN

only important parts of the state to reach the output. The networks use activation functions to induce nonlinearity to the data.

Among the most used activation functions are the sigmoid and the hyperbolic tangent. As in our datasets we had no negative values, we decided to use the sigmoid function in our LSTM network, as the interval of this function is [0,1]:

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

As mentioned before, LSTMs are capable of remembering and choosing which data is relevant as future inputs. They do this by using three gates that release data between hidden state and cell state. These gates are called "forget gate", "input gate" and "output gate". An LSTM neuron incorporates a cell, an input gate, an output gate, as well as a forget gate. The transformation process of information passing through a cell is described in [7] as follows. All the gates of the cell are collecting activations from the block and from the outside. A recurrent connection with the weight 1 keeps the current internal state of a cell. The input and output gates scale the input and output of the cell by using activation functions. The forget gate decides which information must be eliminated from the cell state. That is a sigmoid layer, which provides output values between 0 and 1, and scales the internal state, so as the values exiting the gate are ranged in the interval mentioned above.

For our experiment, we implemented the LSTM network using Python 3 and the TensorFlow framework with the Keras API. As input data, we used the datasets provided by the FENECON Energy Management System (FEMS) described in [3]. Fig. 2 shows the schema of this system, with "PV1" and "PV2" being the producers and "Ph1", "Ph2" and "Ph3" being the consumers.

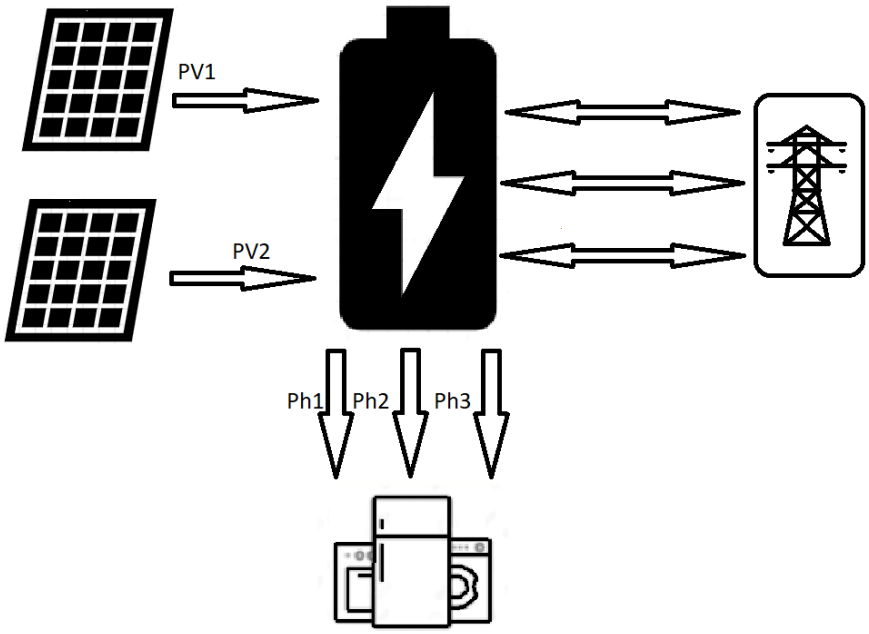


Figure 2. FEMS

The datasets from the system consist in 5 files, two of them containing values from the energy produced by "PV1" and "PV2" photovoltaic panels, and the other three files ,each having recordings for "Ph1" , "Ph2" and "Ph3" consumers. The recording interval for each value is 5 minutes and the period is 5 months. Each file consists of a 1D array of values.

To determine the accuracy of our LSTM network, we used the Mean Absolute Error (MAE) metric:

$$MAE = \frac{\sum_{i=1}^{N} |R_i - P_i|}{N} \tag{2}$$

where R_i is the real value at time i, P_i is the predicted value (the output provided by the LSTM) at time i, whereas N is the total number of evaluated electricity levels. All the values from our datasets are measured in Watts and therefore, the MAE values which we are going to report further are also represented in Watts.

4. Experimental Results

In this section we focus on the results we obtained from our experiment. We tuned the LSTMs parameters in an effort to try and find the best configuration that would produce the smallest value for the MAE. We started with a standard configuration of 5 inputs, two hidden layers each containing 50 neurons, a learning rate of 0.01 and 30 epochs. The first parameter that we varied was

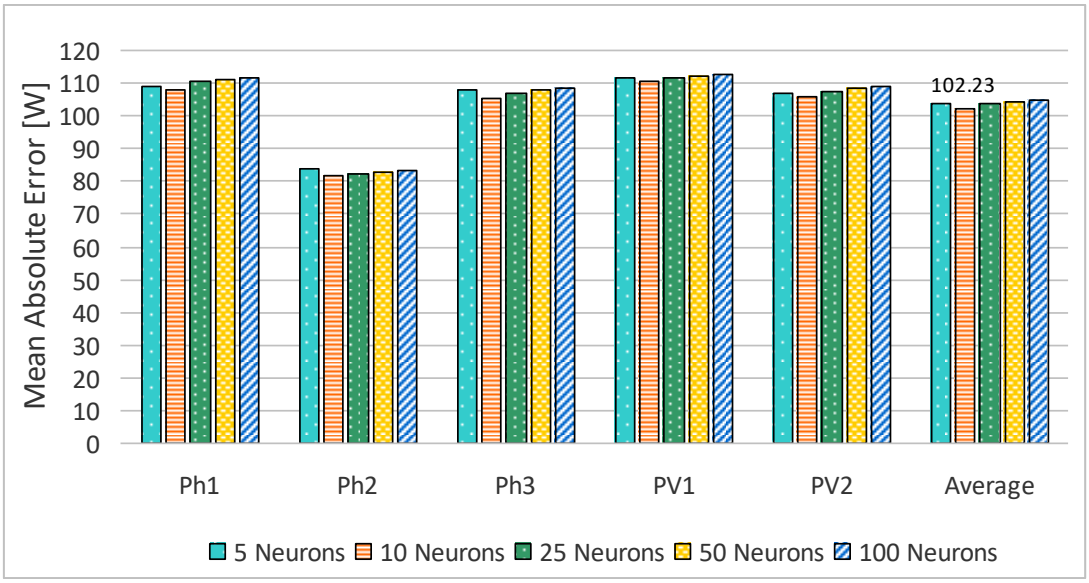


Figure 3. The influence of the number of neurons from the first hidden layer

the number of neurons in the first hidden layer, going from 5 values to 10, 25, 50 and 100, leaving the rest of the configuration unchanged. Due to the fact that the LSTM provides slightly different results in different runs because of its random initialization, ee ran each dataset through the network 5 times for each changed parameter, and calculated the average of the MAE values we obtained. By increasing the input number, we noticed the MAE value was increasing. After a series of experiments towards this direction, we concluded with 10 being the optimal value, and thus we obtained the MAE equal to 102.23 for this configuration. Fig. 3 shows a graph with the values obtained following the tests.

Next we varied the number of neurons from the second hidden layer, following the same pattern that we used for the previous varied parameter. Starting with the base configuration and adding the optimal value 10 for the first hidden layer, we experimented with the second layer starting with 5 neurons, then 10, 25, 50 and 100, and the smallest MAE value we obtained was 101.47, for 5 neurons on the second hidden layer. We concluded with this value being the optimal tune for this parameter. Fig. 4 describes the results obtained by experimenting with the above mentioned values through all the datasets.

The next parameter we tuned was the learning rate, starting from a value 0.01 and slightly increasing it to 0.02 and 0.03. We noticed that by increasing the learning rate, the MAE value also increased, to the point where we reached the value 107.1 with a 0.03 learning rate value, so we decided to stop increasing it. The optimal configuration here is with a value of 0.01, having a MAE of 101.47, which means that our configuration up until this point was accurate with this parameter already having an optimal value. The graph with the results can be seen in Fig. 5.

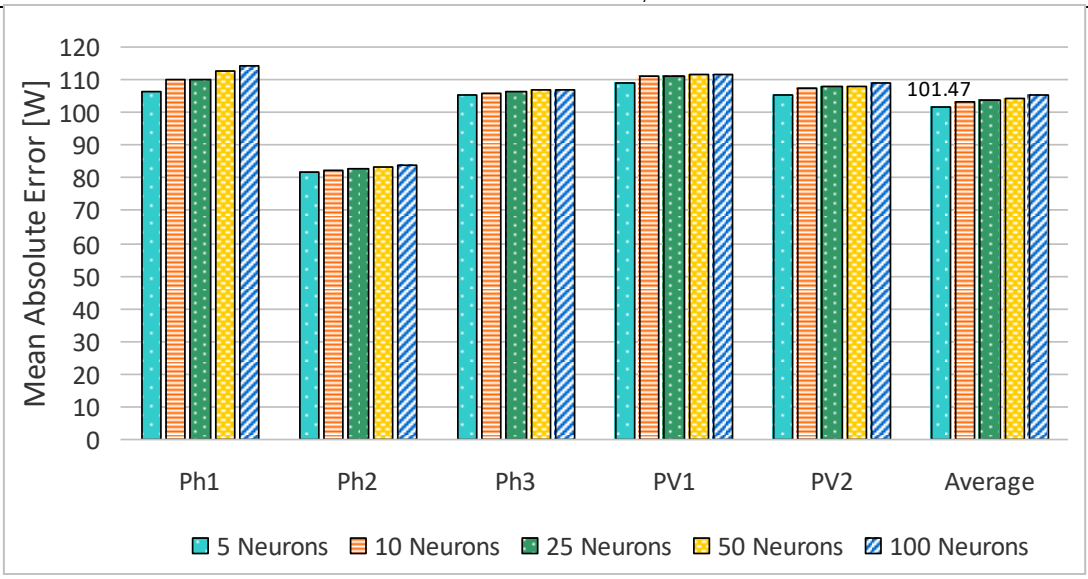


Figure 4. The influence of the number of neurons from the second hidden layer

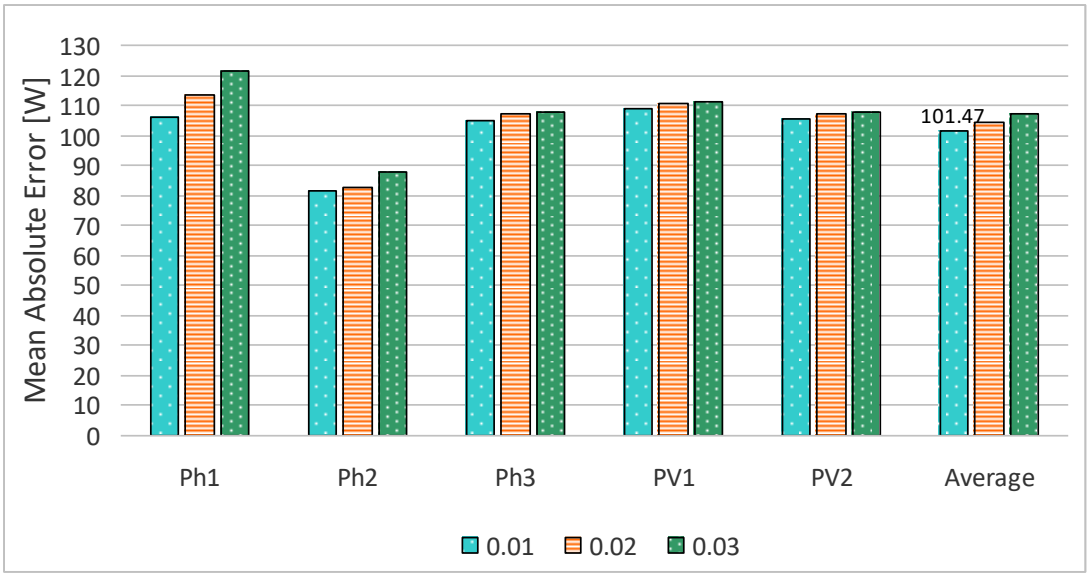


Figure 5. The influence of the learning rate

The next parameter we varied for our configuration was the number of inputs. Having reached the MAE equal to 101.47 with our current configuration using 5 inputs, we increased the number to 10, 15, 20 and 25. We noticed that the higher the number of the inputs, the higher the value of the MAE became. So we also decided to try a smaller number than the starting one and, thus, we went with 4 inputs. This proved to be the right decision, as we reached a MAE equal to 100.99. The results are visible in the graph from Fig. 6.

The last parameter that we decided to vary was the number of epochs. Our base configuration had 30 epochs which achieved the above mentioned MAE, so we decided to increase this number. We varied through 50, 100 and 500 epochs. The results we obtained drove us to the conclusion that 50 epochs was the best configuration, having obtained a MAE equal to 100.77.

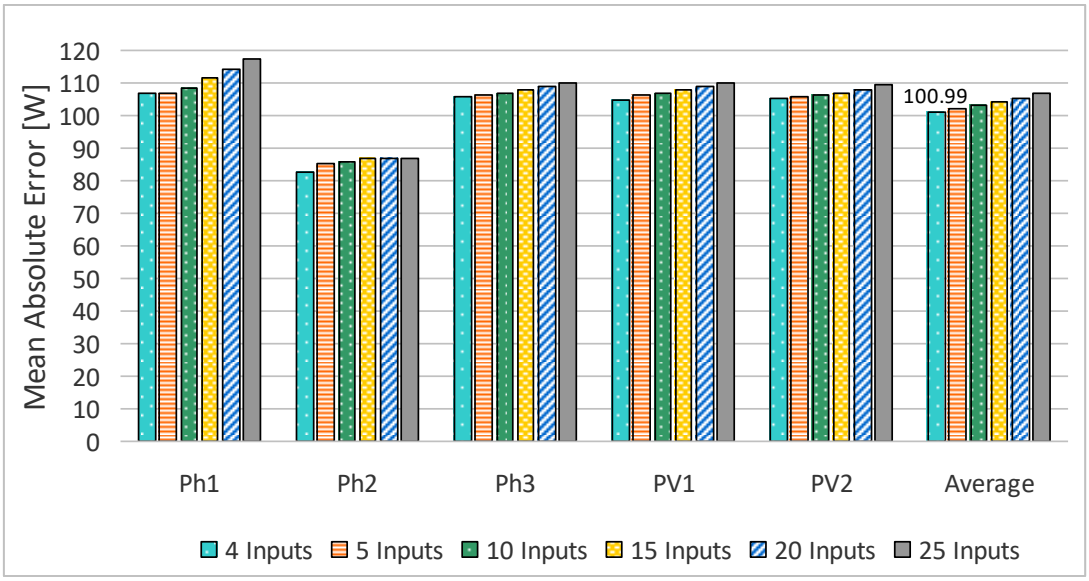


Figure 6. The influence of the input vector size

We also tried to go below our starting value and we decided to run a series of tests with 25 epochs, but as we can see in the graph from Fig. 7, the MAE was higher than the one we obtained with our optimal configuration.

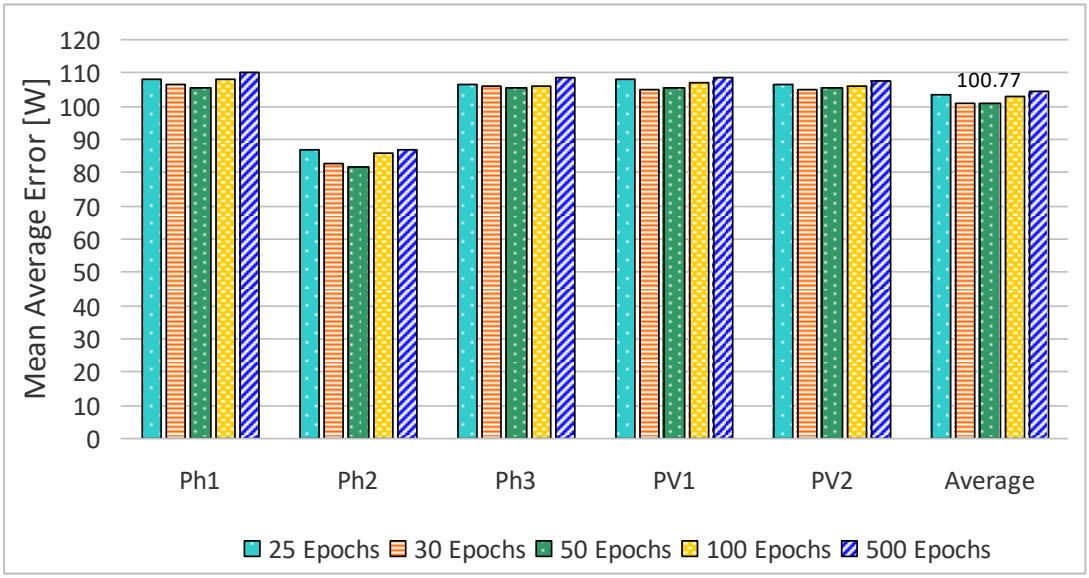


Figure 7. The influence of the number of epochs

After the experiment we concluded that the optimal LSTM configuration has 4 inputs, a first hidden layer with 10 neurons, a second hidden layer with 5 neurons, a learning rate of 0.01 and 50 epochs. Next we made a comparison of our results with other methods used to calculate the MAE on the same datasets.

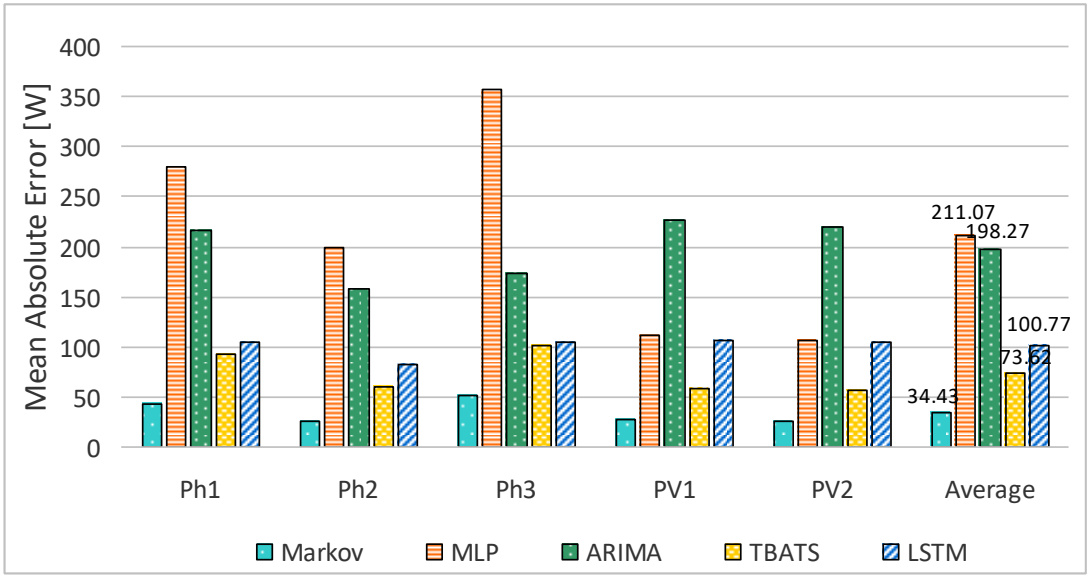


Figure 8. Comparison with other forecasting methods

As the graph in Fig. 8 portraits, with a MAE of 100.77 Watts, our LSTM network outperformed a MLP [3], which had a MAE equal to 211.07, and an ARIMA model [5], with a MAE equal to 198.27. Our LSTM had poorer performance than a TBATS algorithm [5], which had a MAE equal to 73.62, and also than a Markov predictor [4] with MAE 34.43.

5. Conclusions and Further Work

In this paper, we analyzed the LSTM used as a predictor of the electricity consumption and production in a smart house. The goal is to integrate such a predictor into a smart energy management system of a household, that might keep a balance between the electricity consumption and production avoiding demands from the grid. The evaluations performed on the datasets collected from a real household have shown that the LSTM's mean average error is 100.77 Watts, which is half of the mean average error encountered by the MLP and by the ARIMA [5] statistical algorithm, respectively. The LSTM proved to be less accurate than the Markov predictor [4] and the TBATS algorithm [5], but we can classify it among the best methods.

Taking into account the constructive and functional differences of the best performing methods (the Markov model-based stochastic predictor, the TBATS statistical predictor and the LSTM neural predictor), we are interested to develop a hybrid prediction mechanism able to exploit all these three predictors as components. Another further work direction is to develop and evaluate a prediction method relying on fuzzy logic.

References

[1] S. Abdulkarim, *Time series prediction with simple recurrent neural networks*, Bayero Journal of Pure and Applied Sciences, Vol. 9, No. 1, 2016.

[2] C. Fan, F. Xiao, S. Wang, *Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques*, Applied Energy, Vol. 127, 2014.

[3] S. Feilmeier, *Loads management based on photovoltaic and energy storage system*, M.Sc. Thesis, Lucian Blaga University of Sibiu, 2015.

[4] A. Gellert, A. Florea, U. Fiore, F. Palmieri, P. Zanetti, *A study on forecasting electricity production and consumption in smart cities and factories*, International Journal of Information Management, Elsevier, ISSN 0268-4012, Vol. 49, pp. 546-556, December 2019.

[5] A. Gellert, U. Fiore, A. Florea, R. Chis, *Forecasting Electricity Consumption and Production in Smart Homes*, Submitted to Pervasive and Mobile Computing, November 2020.

[6] A. Gellert, A. Florea, *Investigating a New Design Pattern for Efficient Implementation of Prediction Algorithms*, Journal of Digital Information Management, Vol. 11, Issue 5, ISSN 0972-7272, pp. 366-377, October 2013.

[7] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Studies in Computational Intelligence, Springer, ISBN: 978-3-642-24796-5, January 2012

[8] K. Grolinger, A. L'Heureux, M. Capretz, L. Seewald, *Energy Forecasting for Event Venues: Big Data and Prediction Accuracy*, Energy and Buildings, Vol. 112, pp. 222-233, 2016.

[9] L. Hernández, C. Baladrón, J. Aguiar, B. Carro, A. Sánchez-Esguevillas, J. Lloret, *Artificial neural networks for short-term load forecasting in microgrids environment*, Energy, Vol. 75, pp. 252-264, 2014.

[10] S. Hochreiter, J. Schmidhuber, *Long Short-Term Memory*, Neural Computation, Vol. 9, No. 8, pp. 1735-1780, 1997.

[11] K. Kavaklioglu, *Modeling and prediction of Turkey's electricity consumption using Support Vector Regression*, Applied Energy, Vol. 88, Issue 1, pp. 368-375, 2011.

[12] T. Khatib, A. Mohamed, K. Sopian, M. Mahmoud, *Solar Energy Prediction for Malaysia Using Artificial Neural Networks*, International Journal of Photoenergy, Vol. 2012, 2012.

[13] H. Khosravani, M.D.M. Castilla, M. Berenguel, A. Ruano, P. Ferreira, *A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building*, Energies, Vol. 9, Issue 1, 2016.

[14] C. Monteiro, L.A. Fernandez-Jimenez, I.J. Ramirez-Rosado, A. Muñoz-Jimenez, Pedro M. Lara-Santillan, *Short-Term Forecasting Models for Photovoltaic Plants: Analytical versus Soft-Computing Techniques*, Mathematical Problems in Engineering, Vol. 2013, 2013.

[15] J. Park, D. Yi, S. Ji, *Analysis of Recurrent Neural Network and Predictions*, Symmetry, Vol. 12, No. 4, 2020.

DOI: 10.2478/ijasitels-2020-0008

Study of the influence of the exhaust line ultrasounds over the performance of the Blind Spot Warning System

Cătălin MEIROŞU[1]

Universitatea Politehnica din București, Facultatea de Electronică, Telecomunicații și Tehnologii Informaționale

6meirosucatalin86@gmail.com

Abstract

During the previous years, the vehicle manufacturers have tried to equip their vehicles with as much technology as possible, making the driving experience for people easier than ever. Most of the modern vehicles come today with ADAS (Advanced Driver Assistance Systems) either for driving (E.g. Cruise Control, Blind Spot Warning) or Parking (E.g. Rear Ultrasonic Sensors, Rear View Camera). Since the vehicle come equipped with more technology, a major task in developing vehicle remains the integration of these ADAS system in the vehicle context with the other components. Since most of the components cope with each other on the vehicle level, some technologies are more affected by other components – such as the case of an ultrasound vehicle scanning system (Blind Spot Warning) and the Exhaust line that emits ultrasounds from the exhaust muffler. The aim of this paper is to study the influence of the exhaust line ultrasounds (ultrasounds that are emitted by the engine cycle and filtered in the exhaust line of the vehicle) over the detection performance of the Blind Spot Warning Ultrasound system. Since vehicles are sold with a wide variety of powertrains, the solution presented took into account also these differences between powertrains equipped. In order to test the solution, mock-ups of the vehicle were made in order to proof the robustness of the method.

Keywords: ADAS, BSW, Blindspot, Ultrasounds

1 Introduction

This paper presents the performance influence of the Exhaust line of the vehicle on the performance of the Blind Spot Warning System. Also, the phenomenon will be explained along with the various configuration of the powertrains and the exhaust layouts that can directly affect the performance of the Blind Spot Warning detection. Tests have been carried out on a prototype vehicle to observe the sensor positioning impact for the Blind Spot Warning performance.

The influence of the ultrasonic sensor positioning is observed in this paper, along with the various powertrain configurations that has a strong influence over the Blind Spot Warning detection performance. To determine the impact that the vehicle exhaust system has over the Blind Spot Warning detection on the prototype vehicle, the results of the preliminary tests based on a prototype vehicle will be presented. The prototype vehicle) will be tested in different sensor implementation scenarios and different filtering scenarios, so that the final position for the upcoming marketed vehicle – can be established. The influences of the exhaust ultrasounds over the Blind Spot Warning detection was tested for both prototype and final vehicle both on the test track but also on rolling road conditions for a better result interpretation. Moreover, various filtering scenarios were tested to establish the best performance for each setup, but also to discover the best suitable configuration for all powertrains that will equip the upcoming vehicle.

2 Phenomenon explanation

The false Blind Spot Warning detection scenario is described by a positive alert (The Blind Spot Warning System senses a vehicle in the blind spot) when no vehicle is present in the blind spot, as illustrated in Figure 1 below.
A false detection signals the driver by lighting up the light emitting diodes present in the side view mirror glass. An example of the false warning when there is no vehicle detected on the rear side is presented in the Fig. (1) below:

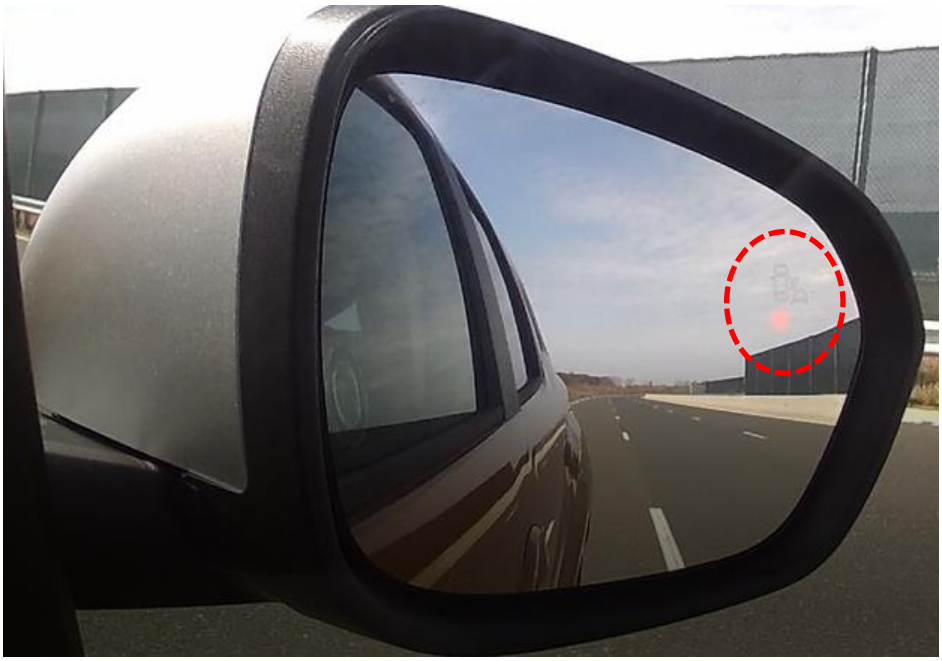


Figure 1. False warning alert example on the tested vehicle [1]

The false detection scenario for the prototype vehicle is repeated for a medium and high engine speed (rpm) at a certain torque level demanded by the driver

(by pressing the accelerator pedal). In order to consider a clean detection (good vehicle detection), the Blind Spot Warning detection shall be centered in a 3m x 3m square in the rear end of the leading vehicle, both right and left-side, as presented in the Fig. (2) below.

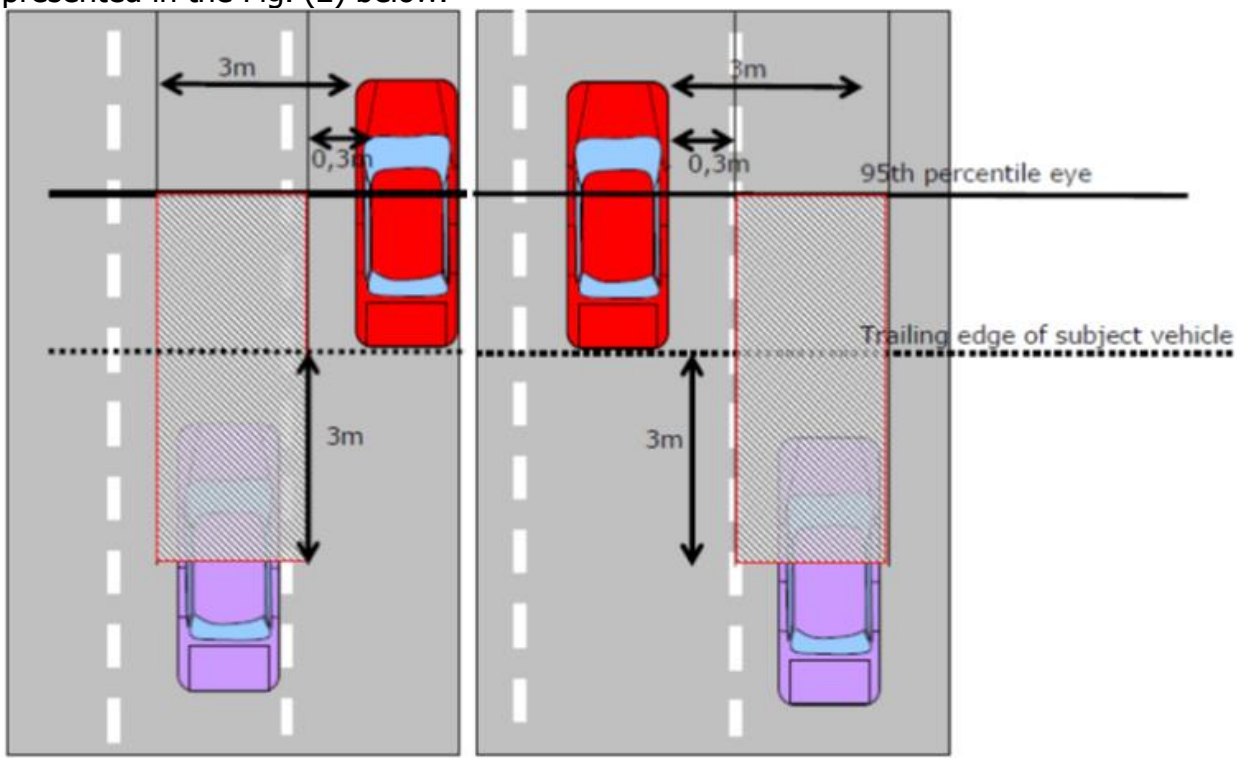


Figure 2. Clean detection of the BSW system is within the hashed square [1]

3 Testing procedures

3.1 Testing track procedure

In order to determine if the Blind Spot Warning detection is affected or not by the Exhaust noise and/or if a false detection may appear, the testing procedure will contain the following tests that simulate a possible driving scenario. This below-mentioned procedure comes to enforce the severity on the BSW system as the exhaust is being heated up during the consecutive tests and the exhaust gases increase in velocity as the exhaust heats up, this generating even more false detections. The testing procedure for testing the false detections contain the following tests:

1. Acceleration starting from standstill (0 km/h) until 80 km/h with 25% accelerator pedal pressed with a gear shift at 3000 rpm for Diesel engines and 3500 rpm for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
2. Acceleration starting from standstill (0 km/h) until 80 km/h with 50% accelerator pedal pressed with a gear shift at 4000 rpm for Diesel

engines and 4500 rpm for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.

3. Acceleration starting from standstill (0 km/h) until 80 km/h with 100% accelerator pedal pressed with a gear shift at 4500 rpm (regulator engine speed) for Diesel engines and 6500 rpm (regulator engine speed) for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
4. Acceleration 50 - 110 km/h starting from 3rd gear with 25% accelerator pedal pressed with a gear shift at 3000 rpm for Diesel engines and 3500 rpm for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
5. Acceleration 50 - 110 km/h starting from 3rd gear with 50% accelerator pedal pressed with a gear shift at 4000 rpm for Diesel engines and 4500 rpm for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
6. Acceleration 50 - 110 km/h starting from 3rd gear with 50% accelerator pedal pressed with a gear shift at 4500 rpm (regulator engine speed) for Diesel engines and 6500 rpm (regulator engine speed) for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
7. Acceleration 80 - 130 km/h starting from 4th gear with 25% accelerator pedal pressed with a gear shift at 3000 rpm for Diesel engines and 3500 rpm for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
8. Acceleration 80 - 130 km/h starting from 4th gear with 50% accelerator pedal pressed with a gear shift at 4000 rpm for Diesel engines and 4500 rpm for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
9. Acceleration 80 - 130 km/h starting from 4th gear with 100% accelerator pedal pressed with a gear shift at 4500 rpm (regulator engine speed) for Diesel engines and 6500 rpm (regulator engine speed) for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording.
10. Acceleration from standstill 0 - 130 km/h starting from 1st gear with 50% accelerator pedal pressed with a gear shift at 3000 rpm for Diesel engines and 3500 rpm for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording. This test scenario is also known as the Aggressive Level 1.
11. Acceleration from standstill 0 - 130 km/h starting from 1st gear with 100% accelerator pedal pressed with a gear shift at 4500 rpm (regulator engine speed) for Diesel engines and 6500 rpm (regulator engine speed)

for Gasoline engines. This test will repeat 3 times and the false detections will be noted down and compiled from the testing CAN recording. This test scenario is also known as the Aggressive Level 2.

3.2 Testing track results on the prototype vehicle

The test results show false detections for the harshest testing scenarios for the powertrains mentioned, presented as below:
Aggressive level 1: acceleration from standstill 0-130 km/h at 50% accelerator pedal press with gear changes at at 3000 rpm for Diesel engines and 3500 rpm for Gasoline engines
Aggressive level 2: acceleration from standstill 0-130 km/h at 100% accelerator pedal press with gear changes at at 4500 rpm for Diesel engines and 6500 rpm for Gasoline engines
For each powertrain available for the prototype vehicle, all the testing scenarios mentioned above have been carried out, with false detections for the aggressive level 1 and aggressive level 2. Some powertrain configurations do not have any false detections and therefore it will not be presented.=
The powertrains that have been reported with false Blind Spot Warning detections are presented below:

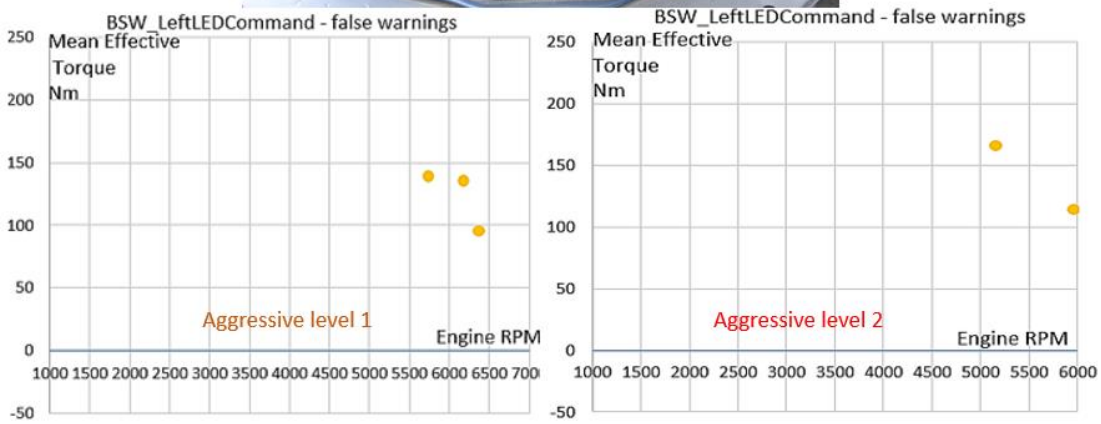


Figure 3. False detections for H5FT, 1.2-liter 16V Turbocharged Gasoline Euro 6 engine – Aggressive level 1&2 [1]

3.2.1 Gasoline 1.2 liter Turbocharged Engine

For this powertrain, the noise level is lower than in the case of a Diesel engine, as it can be seen in the graph below. This powertrain, however, does show some false detections for a very high engine speed (above 6000 rpm). Nevertheless, this engine speed is not so often used during day-by-day driving and can be considered insignificant.

In this case, due to the low incidence of the false detections both on aggressive level 1&2, the false detection can be considered as the same between the two aggressive levels, even though during aggressive level 2, the occurrence is lesser.

3.2.2 Gasoline 2.0 liter Normally Aspirated Engine

For this particular engine, used only in extreme cold areas (Russia, Ukraine and the Ex-Communist Russia block) and extreme hot areas (Golf countries, Persian Gulf, UAE, Arab countries), even though it is Natural aspirated (no turbocharger), the noise level is higher than the 1.2-liter engine, the principal differences between the two are:
Increased Engine Displacement (almost 0,8l)
Different exhaust line, including catalyst that is different from Euro4 to Euro6, the Euro 6 corresponding to 1.2-liter engine is more restrictive, therefore can be quieter.
Older technology, corresponding to lower requirements for the exhaust pass-by-noise regulations.
This engine has a very high noise level registered between 4500-5000 rpm, as it can be observed in the graph below. This engine speed range can be used by the drivers while overpassing or highway entries.

In this particular case, where a filtering is needed, the detection distance will drop with respect to the 4m threshold established in the project requirements for the Blind Spot Warning system. For a lower aggressive level, a higher number of false detections can be observed in the 4500-5500 engine speed threshold.
For aggressive level 2, the use case in which more of the false detections shall happen due to a high engine speed and a high torque demand by the driver (remember that aggressive level 2 stands for 100% accelerator pedal pressed), the false detections as actually less frequent. This behavior is due to the fact that during Aggressive level 1, the echo sent by the exhaust is received by the BSW sensors as there is time to process the incoming ultrasounds. In the case of aggressive level 2, the engine speed increases so fast that the echo sent by the exhaust is ignored by the BSW sensors (the sample rate is faster than the processing sample of the BSW system).

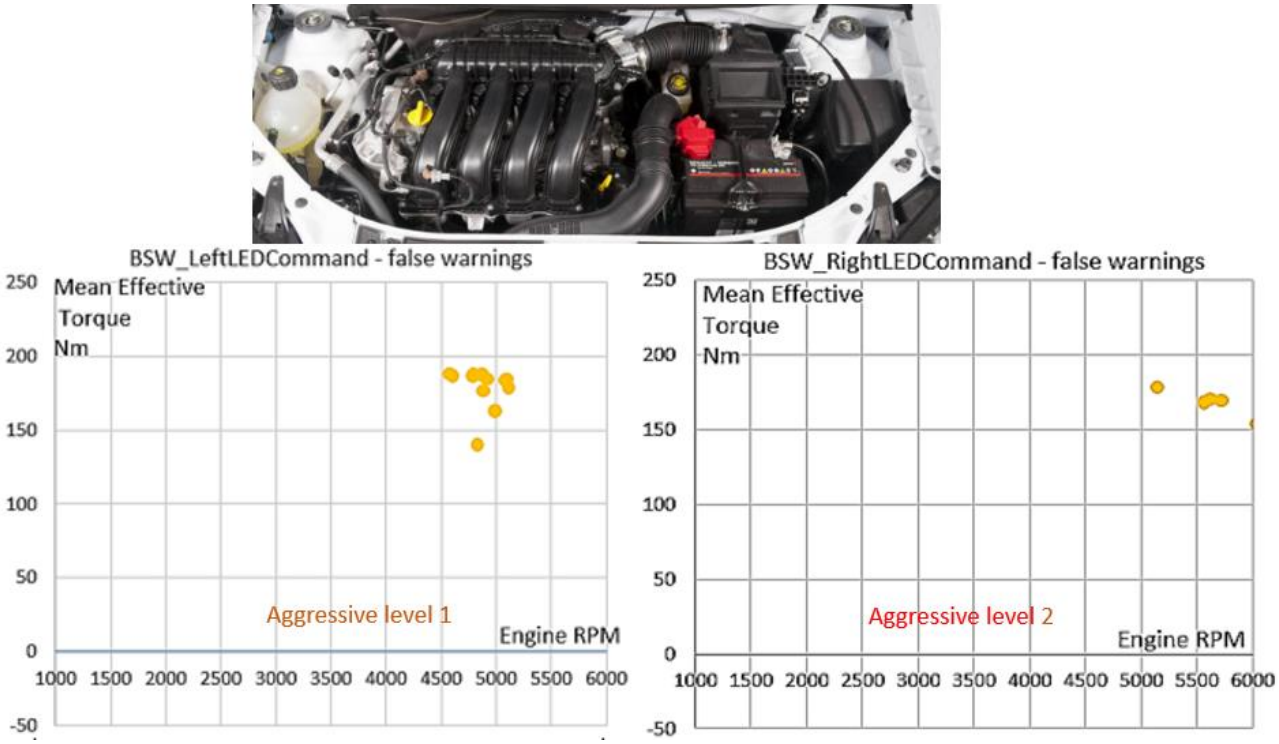


Figure 4. False detections for F4R, 2-liter 16V Gasoline Euro 4 engine – Aggressive level 1&2 [1]

3.2.3 Diesel 1.5 liter Turbocharged Euro 6 Engine

For this 1.5-liter Euro 6 engine sold in the Northern-Africa region, it can be observed in the below figure that the false detection is very pronounced both on aggressive level 1 and aggressive level 2. This is mainly due to the Exhaust line designed specifically for high temperatures and fast gas flow (in order not to maintain heat and help cooling).

As in the case of the F4R engine presented above, the occurrence of the false detections is lesser for aggressive level 2, as the BSW sensors have less time to process the echo coming from the Exhaust line, the sample rate of the BSW system being over the frequency of the exhaust noise.
However, this powertrain has a lot of occurrences for aggressive level 1, in the engine speed threshold 3000-3800 rpm, this threshold being used by the drivers for overpassing or for highway acceleration lanes. This specific interval needs to be filtered out as the tests on the prototype vehicle show a very high probability that a false detection occurs.

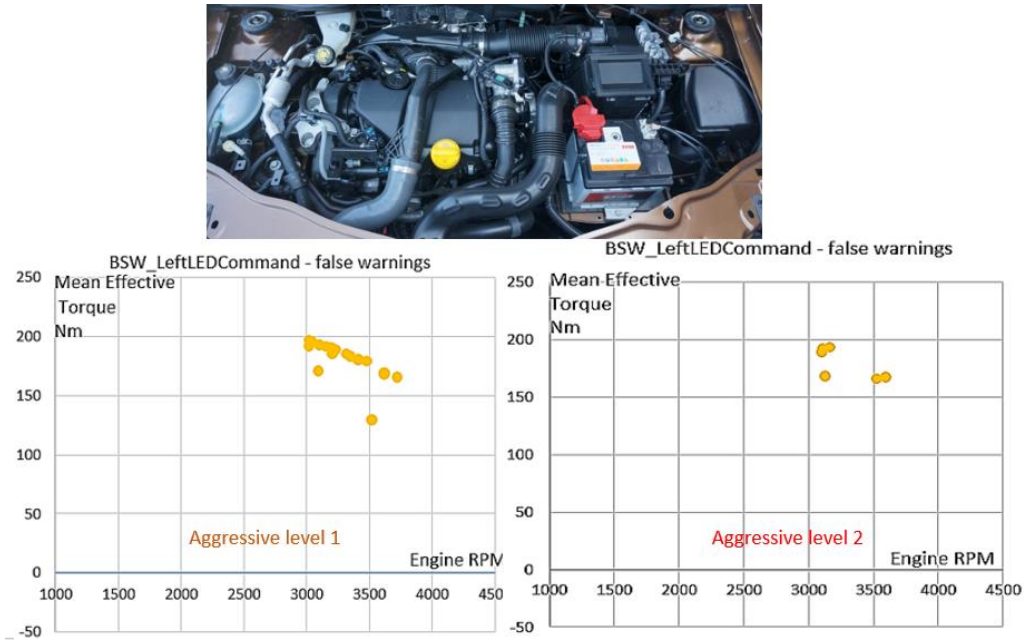


Figure 5. False detections for K9K 1.5-liter Gen 5 - Euro 6 Diesel Engine – Aggressive level 1&2 [1]

Diesel 1.5 liter Turbocharged Euro 6 B Engine

This 1.5-liter Diesel engine is currently used on the vehicle, the vehicle that is sold with the studied BSW system. This calibration for the 1.5-liter DCI engine is made specifically for the European market, being a version that respect the harshest environmental CO_2 regulations. In the figure below, it can be observed that for the same type of engine as above mentioned but only a new Software calibration and a new exhaust gas treatment system can influence so much the BSW detection.

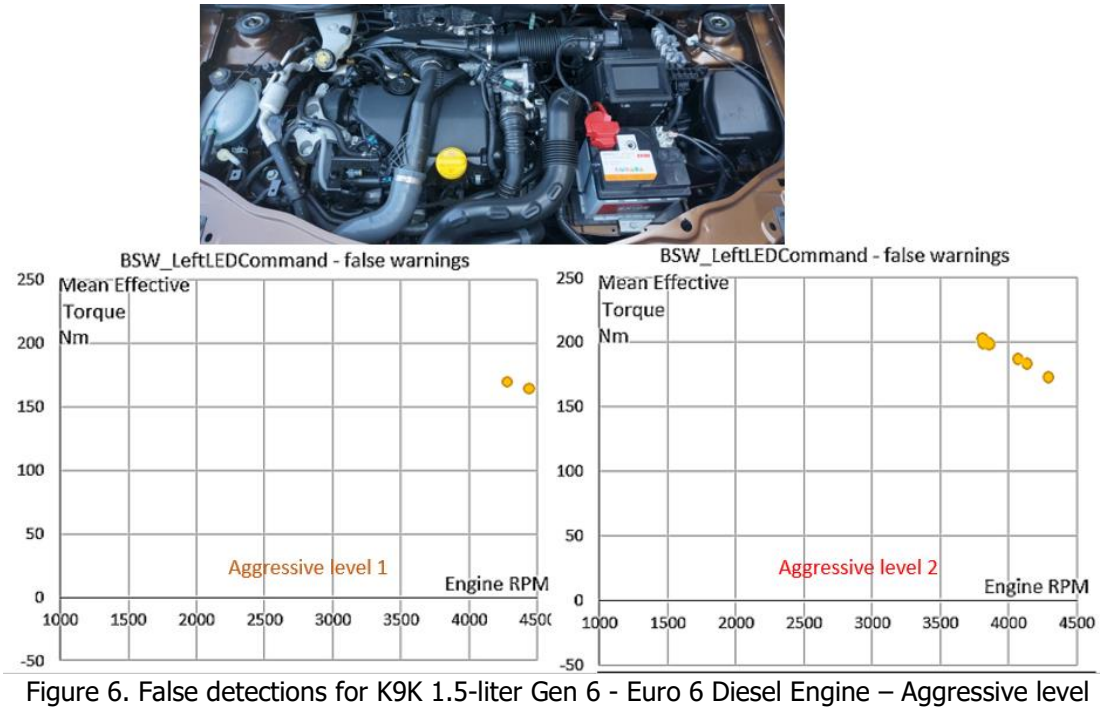


Figure 6. False detections for K9K 1.5-liter Gen 6 - Euro 6 Diesel Engine – Aggressive level 1&2 [1]

4 Preliminary test results

The preliminary tests on this prototype vehicle show a different behavior of the false detections between a Diesel and Gasoline engines, the Diesel engine having false detections between 3500-4500 rpm, this being the usual engine speed during an overtaking. In the case of a Gasoline engine, the false detections appear in the threshold 6000-6500 rpm, a very high engine speed (close to the regulator cut-off). This Gasoline-engine false detection engine speed threshold is not used during day-to-day driving, even for extreme overpassing scenarios, as at this engine speed, the torque value is low and not efficient for overpassing.
These false detections are caused by the exhaust gases that generate some ultrasonic waves on the same frequencies as the operating frequencies of the Bosch® Blind Spot warning system. These exhaust gases ultrasounds are received as echo by the Ultrasonic sensors used by the BSW system and the system triggers the alert even though there is no vehicle present in the Blind Spot. A short schematic explaining the steps is presented in the Fig. (7):

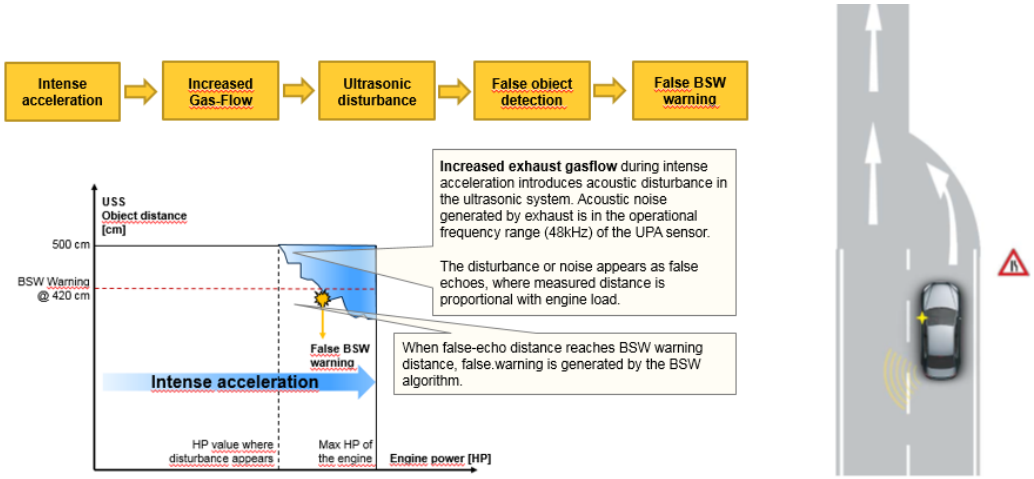


Figure 7. False BSW detection occurrence short explanation [1]

Solution

As seen above, multiple powertrains of the vehicles are affected by the false detection of the BSW system due to the ultrasounds generated by the various exhaust systems. Therefore, to sort out this flaw in detection, a transversal solution had to be developed. The best solution for this variety of powertrains is a software filter applied in the BSW ECU that ignores the false warnings coming as an echo from the exhaust. In order to obtain the data for building the software filter that can be applied successfully for all the powertrains, a data acquisition with an analog sensor mounted in the place of the rear sensor closest to the exhaust line was made. The results conducted on the prototype vehicles in TITU facility show the presence of false detections for a noise level above 98 dB(A) – corresponding to 28 mV in some scenarios, and a sure false detection for noise level above 104 dB(A) – corresponding to 54 mV on the analog sensor that measured the noise level. For a noise level in the threshold 98 -104 dB(A), the false detection is not confirmed all the time, since the ultrasound frequency duration as received as echo represents a short sample

that is not processed by the system's sensors. Therefore, the filter must be applied for a noise level above 104 dB(A), so that the occurrence of a false detection shall be as low as possible. In the figure below, it is represented a data acquisition result for a H5FT engine (1.2-liter Turbocharged Gasoline engine) showing the thresholds for possible and certain false detections occurrences.

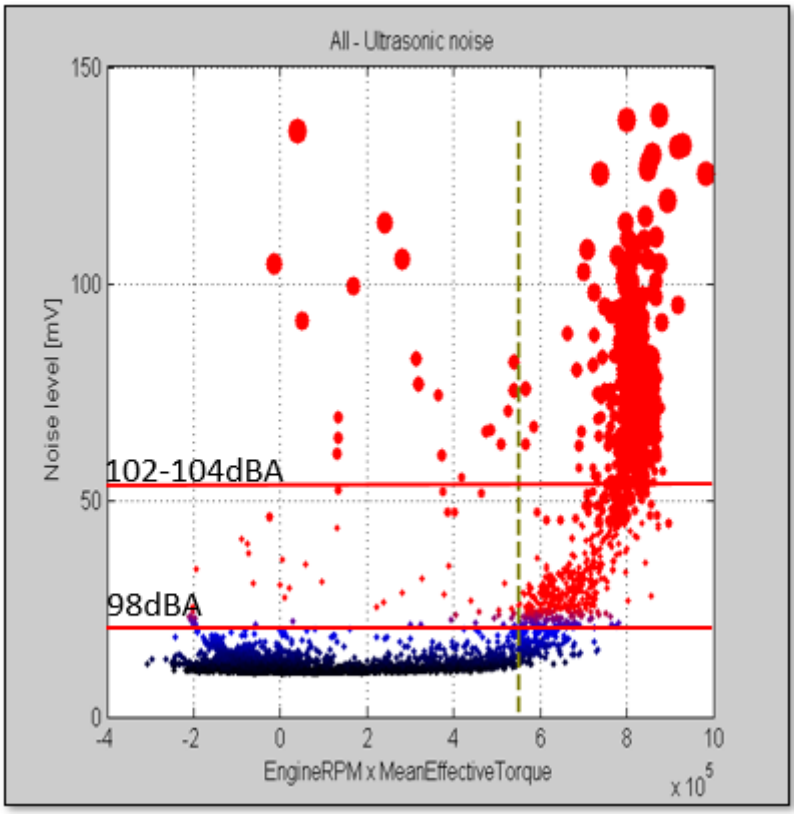


Figure 8. Exhaust noise level data acquisition for the H5FT engine (1.2-liter Turbocharged Gasoline engine) [1]

As it can be seen in the Fig. (8), all the points below 28 mV or 98 dB(A) is safe from influencing the performance of the Blind Spot Warning system, marked on the figure with small blue dots. For the threshold 28mV-54 mV, translated to a noise level between 98-104 dB(A) marked with small red dots, there is a possibility that the false warnings can be triggered but the exhaust ultrasound emitting period is lower than the sample rate of the Blind Spot Warning system, this meaning that the Blind Spot Waning ultrasonic system may not be affected by the echo received for a noise level between 98-104 dB(A). For a noise level above 54 mV (measured with the analog sensor), the false alerts are very likely to be triggered, therefore this noise level needs to be filtered.
As observed on the figure, some false warnings may happen also for low engine speeds (3000 rpm), this being caused by a high torque demand on the driver's behalf for a high noise lever registered (over 100 mV). In most of the scenarios, the false detections happen after 5000 rpm for this particular engine (H5FT), as it was shown in the real prototype tests presented in the chapters above.

Since for these high engine speeds, the probability is higher for a false warning to occur, the filter shall be proposed to treat precisely these scenarios. In order to fully define the filter entirely, the data acquisition needs to be made for each powertrain and exhaust architecture used for the prototype and final series vehicle.
The software filter represents a compromise between not having false detection that may perturb the driver and the detection performance of the Blind Spot Warning system. Therefore, the tradeoff is false warnings for detection distance. Since the detection distance cannot be negotiated as there is also a minimum detection distance imposed by the NCAP advisory manual, the filter must comply with the 3m x 3m detection distance. Below is a table that lists the correlation between the noise value obtained from the data acquisition and the maximum detection distance that the sensor can achieve for this particular noise without having any false warnings.

Table 1. Correlation between Sensor noise level and the maximum detection distance [1]

Noise value [mV]	Detection range [cm]
100	230
90	237
80	244
70	250
60	285
50	315
40	350
30	400
20	450
10	500

As presented above, to respect the requirements given by the project and imposed by NCAP® for being allegeable in scoring points in their rating system, the noise level must be maximum 54 mV at sensor level. The table also shows that the lower the noise, the longer the detection distance can be achieved without any false warnings.
The filter also shall take into consideration the engine speed (direct influence on the exhaust line noise) but also on the engine torque as the demand from the engine can also influence the exhaust noise by changing the timing and advance laws on the engine injection process (more Diesel/Gasoline injected, adding more air means the burn inside the engine amplifies which turns to a higher-pressure difference, meaning exhaust noise).
For each powertrain and exhaust architecture, a separate filter characteristic is built, a typical filter can be seen in the figure below. This filter takes into

consideration the two variables – engine speed and torque demanded but also the noise level acquired at sensor level.

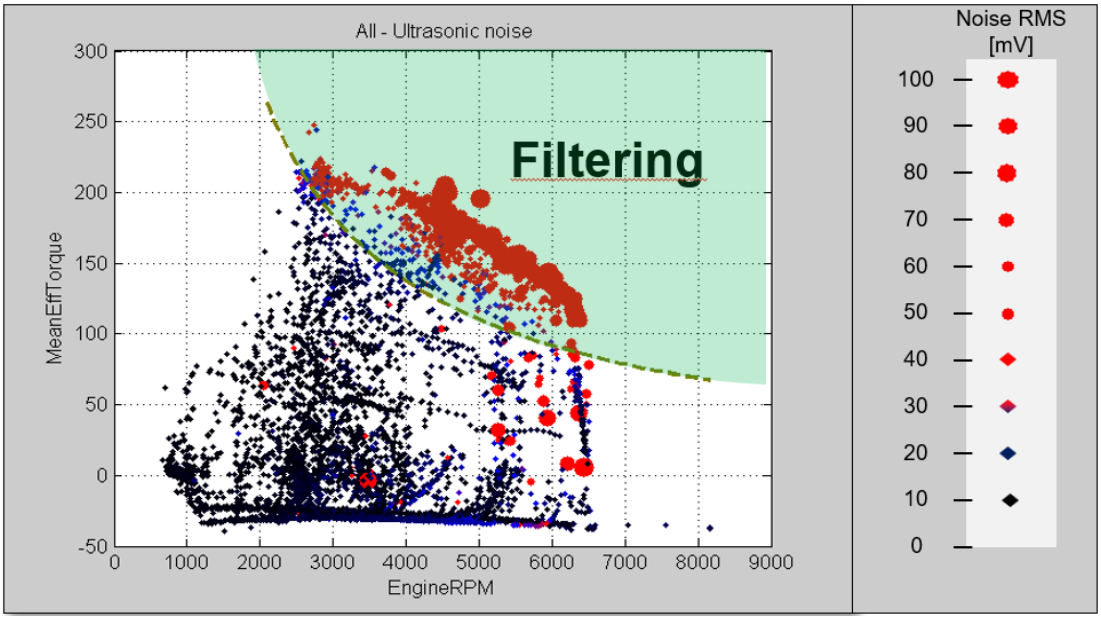


Figure 9. Exhaust noise filter applied on the H5FT (1.2-liter Turbocharged Gasoline engine) [1]

In this case, the large red dots are delimited by a hyperbole between 4000 – 6500 rpm and 100-200 Nm of torque demanded. This is the particular area on which the filter will cut out the frequencies received as echo from the exhaust line, so that it doesn't report false warnings. Outside this hyperbole, towards the lower left corner of the chart, the Blind Spot Warning system is working properly by the required specifications.

5 Conclusions

For the theme studied – the influence of the exhaust ultrasounds over the Blind Spot Warning ultrasound-based detection system, the following tests and research has been carried on:
Tests on the former body-type prototype vehicle with the current installed position and sensors on the TITU tracks and on rolling-road conditions.
Tests on all the possible powertrains and exhaust architecture existing that covers all the range of the sold vehicle worldwide.
6000 km of road testing in normal traffic conditions, based on a test track that covered most of the current driving conditions in Europe.
Testing in rolling road conditions with another 2 different vehicles to confirm and compare the results.
Testing on all 4 seasons and for all types of weather in order to cover all possible conditions that are met in Europe.

Aerodynamic simulations for establishing the best positions for the Blind Spot Warning ultrasound sensors.
Numerical simulations to determine the precise pattern of detection for the rear sensors so that the implementation is a success.
Developing a filter that takes into account two main parameters of the powertrain – engine speed and torque demanded.
Testing with and without a Software filter in order to spot the differences between the two conditions.
These all together conclude that the system used on the sold vehicle is a reliable system that can be used in most of the weather conditions (except rain and wet surface), on most of the paved roads existing in Europe, and on all possible sold powertrains. However, the system has some limitations due to the filtering but only under heavy acceleration.
Since not all the research leads have been studied, this including the rear mud guards, larger exhaust pipes and new prototype catalysts have been studied, the thesis can be a departure point for studying the influence of these elements in the attenuation of false detections for the Blind Spot Warning ultrasound system on the sold vehicle. Also, a very important influence is the temperature rise in the exhaust line caused by extreme engine speed, but also very high external temperatures (>37°C) that causes sudden false warnings for normal driving conditions even though there is no vehicle in the blind spot at that time.

6 Further work

While studying the phenomenon and testing different powertrain layouts and engines, a very interesting aspect was the different behavior for the same engine but different power output calibrations. Besides this, the study for other solutions was not approved for the following leads:
- Rear mud guards for all the vehicles equipped with the Blind Spot Warning system. These additional plastic covers were mounted on the rear side of the wheel arch and prevents the water mist to be gathered around the Blind Spot Warning sensor. This solution could provide a decent protection against false detections. Since the mud guards were not tested and represented an additional cost not taken into account at the beginning of the design phase of the vehicle, this solution was abandoned before the test were started.
- New catalysts, middle silencers and final silencers. Since the whole noise level is to be attenuated by the silencers and the catalyst, the development of new exhaust lines in 2 years-time since the problem was discovered until the vehicle was first sold was not possible nor approved by the project. This solution could have solved the problem only if the whole exhaust system was built to treat this particular problem. But since there are also other requirements form the exhaust system, this problem could not convince the board to accept the budget for the modification. The modification of the whole exhaust lines for all the powertrains was more difficult and expensive than the whole vehicle testing, therefore it was abandoned.

- Larger dimension for the exhaust pipes. This solution was suggested by the NVH testing team that had an expertise with the exhaust noise level for the audible threshold. The larger the diameter of the exhaust pipes, the lower the noise level since the flow section is bigger and the exhaust gas speed can decrease, this causing less ultrasound frequencies emitted on the muffler. However, this solution also has its drawbacks for the NVH exterior noise homologation, as a larger diameter means a higher resonance, therefore a higher noise level on the rolling road homologation. Since the homologation is crucial for vehicle selling, this solution was also abandoned.

7 Acknowledgement

The work has been funded by the Operational Program Human Capital of the Ministry of European Funds through the Financial Agreement 51675/09.07.2019, SMIS code125125

References

[1] Renault Technologie Roumanie, Titu – Blind Spot Warning implementation on the HJD Project 2016-2017
[2] Bosch Engineering Center Cluj, Cluj-Napoca – Development of the Blind Spot Warning System for the HJD project 2016-2017
[3] Renault Technocentre Guyancourt – Development of Engines and Powertrains
[4] EuroNCAP Official European New Car Assessment Programme – Ratings

International Journal of Advanced Statistics and IT&C
for
Economics and Life Sciences

VOLUME 10 Number 1 CONTENTS 2020

ISSN 2067 – 354X
eISSN 2067 - 354X