WIIS2022

Proceedings of Workshop on Intelligent Information Systems

Vladimir Andrunachievici Institute of Mathematics and Computer Science
October 06-08, 2022, Chisinau

CZU 004+519.7(082)
P 93

VLADIMIR ANDRUNACHIEVICI INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE
5, Academiei str., Chisinau, Republic of Moldova, MD 2028
Tel: (373 22) 72-59-82, Fax: (373 22) 73-80-27
E-mail: imam@math.md
WEB address: http://www.math.md

Editors: Dr. I. Titchiev, Prof. S. Cojocaru, Prof. C. Gaindric

Authors are fully responsible for the content of their papers.

Descrierea CIP a Camerei Naţionale a Cărţii

Proceedings of Workshop on Intelligent Information Systems WIIS2022, October 06-08, 2022 Chisinau / editors: I. Titchiev [et al.]. – Chişinău : Institute of Mathematics and Computer Science, 2022 (Valinex). – 197 p. : fig., tab.

Antetit.: Vladimir Andrunachievici Inst. of Mathematics and Computer Science. – Referinţe bibliogr. la sfârşitul art. – Ind. de aut. : p. 194-195. – 150 ex.

ISBN 978-9975-68-461-3.

004+519.7(082)

This issue is supported by the research project 20.80009.5007.22, "Intelligent information systems for solving ill-structured problems, processing knowledge and big data".

Firma poligrafică „VALINEX" SRL, Chişinău, str. Florilor, 30/1A, 26B, tel./fax 43-03-91, e-mail: info@valinex.md, http:\\www.valinex.md

Coli editoriale 10,83. Coli de tipar conv. 11,45. Format 60x84 1/16. Garnitură „Times". Hirtie ofset. Tirajul 150.

COMMITTEES

Program Committee

Inga Țițchiev (chair),	Panayotis Yannakopoulos (Greece),
Svetlana Cojocaru (co-chair),	Serghei Ivanov (France),
Constantin Gaindric (co-chair),	Serghei Verlan (France),
Nataliya Pankratova (Ukraine),	Mircea Petic,
Alexei Averkin (Rusia),	Constantin Ciubotaru,
Horia Teodorescu (Romania),	Artiom Alhazov,
Stefan Kramer (Germany),	Alexandru Colesnicov,
Florin Filip (Romania),	Liudmila Burțeva,
Stefan Pickl (Germany),	Olesea Caftanatov,
Volodymyr G. Skobelev (Ukraine),	Iulian Secrieru,
Vasile Apopei (Romania),	Elena Guțuleac,
Adrian Iftene (Romania),	Ina Ciobanu,
Lenuța Alboaie (Romania),	Corina Negara
Diana Inkpen (Canada),	

Organizing Committee

Inga Țițchiev,	Ludmila Malahov,
Svetlana Cojocaru,	Tudor Bumbu,
Constantin Gaindric,	Tatiana Verlan,
Mircea Petic,	Elena Gutuleac,
Olesea Caftanatov,	Irina Gheorghita,
Iulian Secrieru,	Alexandru Parahonico
Olga Popcova,	

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Explainable Artificial Intelligence

Invited paper

Alexei Averkin

1 Abstract (extended)

Every decade technology makes revolutionary shifts that become the new platforms on which application technology is built. Artificial intelligence (AI) is no different. AI has moved from 1st Generation shallow learning and handcrafted features to 2nd Generation deep learning, which has been effective at learning patterns. We have now entered the 3rd Generation of AI which is machine reasoning-driven – where the machine can interpret decision-making algorithms, even if they have the black-box nature. Explainable artificial intelligence and augmented intelligence are the main part of the 3rd Generation of AI.

But in the 2030s we will see their role in the 4th Generation AI with machines that are learning to learn and that will dynamically accumulate new knowledge & skills. By the 2040s, 5th Generation AI will see imagination machines that are no longer reliant on humans to learn.

Explainable artificial intelligence now represents a key area of research in artificial intelligence and an unusually promising one in which many fuzzy logics could become crucial. Research in the area of explainable artificial intelligence can be divided into three stages, which correlate with the 3 generations of AI: in the first stage (starting from 1970), expert systems were developed; in the second stage (the mid-1980s), the transition was made from expert systems to knowledge-based systems; and in the third phase (since 2010), deep architectures

of artificial neural networks, which required new global research on the construction of explainable systems, have been studied.

According to the DARPA approach, the first wave is the era of handcrafted, declarative knowledge; the second wave is statistical learning; and the third wave is the future. DARPA termed the third wave Contextual Adaptation, but we instead refer to it as DARPA's initiative, AI Next. In addition to the three waves, we have identified six major phases of DARPA's AI investment: AI Beginnings, Strategic Computing, Knowledge/Planning, Cognitive Systems, Data Analytics, and AI Next.

DARPA's initiative, AI Next is related to third-generation explanatory systems and the DARPA program, which began in 2018. The DARPA explainable AI (XAI) program seeks to create AI systems whose learning models and solutions can be understood and properly validated by end users. Achieving this goal requires methods for constructing more explicable models, developing effective explicable interfaces, and understanding the psychological requirements for an effective explanation. Explainable AI is needed for users to understand, properly trust, and effectively manage their smart partners. DARPA sees XAIs as AI systems that can explain their decision to human users, characterize their strengths and weaknesses, and how they will behave in the future. DARPA's goal is to create more human-readable AI systems through effective explanations. XAI development teams solve the first two problems by creating and developing Explainable Machine Learning (ML) technologies, developing principles, strategies, and methods of human-computer interaction to generate effective explanations. Another XAI development team tackles the third challenge by combining, extending, and applying psychological explanatory theories that the development teams will use to test their systems. The development teams evaluate how a clear explanation of XAI systems improves the user experience, confidence, and productivity.

Explainable artificial intelligence now represents a key area of research in artificial intelligence and an unusually promising one in which many fuzzy logics could become crucial. Fuzzy systems can help to ap-

proach various aspects of Explainable Artificial Intelligence (XAI). The pioneering works of L.A. Zadeh offer precious tools for the current XAI challenges. They are not limited to approaches of natural language, but they also help to assist the user in understanding the meaning of the decisions made by artificial intelligence-based systems and to provide explanations about the way these decisions are made. Nowadays, XAI is a prominent and fruitful research field, where many of Zadeh's contributions can become crucial if they are carefully considered and thoroughly developed. It is worth noting that about 30% of publications in Scopus related to XAI, dated back to 2017 or earlier, came from authors well recognized in the Fuzzy Logic field. This is mainly due to the commitment of the fuzzy community to produce interpretable fuzzy systems since interpretability is deeply rooted in the fundamentals of Fuzzy Logic.

Alexei Averkin

Department of Intelligent Systems at Federal Research Center.
Informatics and Computer Sciences of Russian Academy of Sciences
E–mail: averkin2003@inbox.ru

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Machine Learning Under Real-World Constraints

Invited paper

Stefan Kramer

Abstract

For a long time, research in machine learning has focused almost exclusively on the development of algorithms for learning and inference. As machine learning components are now deployed in countless technical systems, various constraints arising from real-world scenarios are gaining importance and need to be considered. In software engineering, these constraints would run under the heading of non-functional properties and requirements. Constraints may be required response times, privacy and confidentiality, fairness, transparency and explainability, and safety. In the talk, I will give examples of how to address these constraints and how fulfilling multiple constraints at the same time may be solved and be a good research topic for the future.

Stefan Kramer

Prof. Dr. Stefan Kramer
Chair of Data Mining
Managing Director of the Institute of Computer Science
Honorary Professor of
The University of Waikato

Johannes Gutenberg University Mainz
Staudingerweg 9, 55128 Mainz, Germany
Phone: +49-6131-39-26990
Fax: +49-6131-39-23534

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Why fuzzy logic systems (FLS) are difficult to tame: General FLS are not expressed by algebraic operations

Invited paper

Horia-Nicolai Teodorescu

1 Abstract (extended)

We point to several persistent misconceptions and wrong beliefs in the field of fuzzy logic and enounce propositions that clarify and, hopefully, will contribute to eliminating those misapprehensions. That may improve many engineering designs and reduce confusion for those learning fuzzy logic systems.

In their paper of 2015, Reshma and John state that "Fuzzy is the logic of approximations. It can be used ... to describe vague and imprecise information" [1]. Such statements are frequently heard at conferences and read in publications. As a first remark, there is no single fuzzy logic. How to choose one version of the infinite number of infinite-valued logics (sometimes named fuzzy logics) for a given application is not an issue solved in the frame of these (fuzzy) logics. This problem of what kind of fuzzy logic to use for building a fuzzy logic system suitable for a specific application rather pertains to metalogics and has not yet been addressed in any significant manner; we are not discussing it. Some glimpses of this issue are found in [2].

In their popular volume on Fuzzy Logic and Control, Jamshidi, Vadiee, and Ross [3] describe the state of the art in 1993 as *"Among many new technologies based on AI, fuzzy logic is now perhaps the most*

popular area... fuzzy logic is enjoying an unprecedented popularity in the technological and engineering fields including manufacturing." Interestingly, after 30 years from this book and almost 60 years from Zadeh's seminal papers, the domain of fuzzy logic is still very much alive and popular. One may ask what contributes to its continuous popularity. One explanation is summarized by Jamshidi et al. [3]: *"Fuzzy logic offers design rules that are relatively easy to use in a wide range of applications, including nonlinear robotic equations"... "Fuzzy logic also allows for design in cases where models are incomplete, unlike most design techniques."* (Section 14.1 in [3]). It is not clear from that volume what are those design rules that fuzzy logic offers, but the popular interpretation of the above is that fuzzy logic is based on If-Then rules that are easy to enunciate in an intuitive manner and that always work more or less well. Unfortunately, the misconceptions about fuzzy logic include:

Misconception #1. "Fuzzy logic offers design rules that are relatively easy to use."

Fuzzy logic is not providing any indication of how to choose the rules for a fuzzy logic system, even less "design rules". While it is easy to enunciate If-Then rules, there is no indication inside fuzzy logic theory to estimate the quality and suitability of those rules.

The easiness of enunciating fuzzy If-Then rules does not mean the easiness of finding a "good enough" set of rules for a specific application. The latter was implicitly recognized by the need to develop myriads of intricated, combined techniques (many based on genetic algorithms or other biology-inspired methods) for the optimization of the rule systems. In addition, fuzzy logic and fuzzy logic systems (FLS) theories are related, yet distinct. Among others, fuzzy logic says nothing about how to choose a defuzzifier, which is an essential part of a FLS.

Another vastly popular misconception is implied by (and well summarized in) Ngo and Tran [4], who say *"The membership functions of the fuzzy sets are triangles and trapezoids due to simple and effective programming."* The implied delusion is

Misconception #2. Using simple, triangular, or trapezoidal membership functions (m.f.s) is suitable in virtually all applications of FLSs.

The fallacy here is more subtle, involving (i) the disconnection of the issue of choosing the set of rules from the issue of choosing the membership functions; (ii) avoiding discussing the choice of the membership functions as an approximation problem (which would need also the consideration of the set of rules). In addition, saying that the triangular or trapezoidal membership functions are simpler to include in programs than, say, Gaussian m.f.s is possibly unjustified. Further on, saying that the program is more effective using triangles and trapezoids is unjustified: the duration or running a program achieving a specified error is often larger for this choice of m.f.s.

Jamshidi et al. in the cited volume [3] are not falling into the same type of error; they are saying *"... determining the optimum shape of a membership function may not always be easy and can sometimes be obtained empirically at best. To avoid these 'black art' approaches to membership functions ..."*. However, several other issues are illustrated by the above quotation, including:

Misconception #3. There is an *"optimum shape of a membership"* function for any application.

In the first place, this is an ill-posed problem, because the optimality criterion is not stated. In addition, again, the issue of m.f.s is disconnected from the issue of the rule system, which is not possible. Assuming that the optimality means that the system approximated with a specified error a given function on a specified interval of the input variable(s), the *"optimum shape of a membership function"* does not exist. It can be easily proved that there is an infinity of m.f.s that, for a specified set of rules, a specified defuzzifier, a specified interval of the input variable(s), and a specified error satisfy the approximation error (at least if the function to approximate has no discontinuity in the interval).

Misconception #4. The design of membership functions and in general of FLSs is much of a "black art" or should necessarily be done

using some adaptive or learning system, such as a neural network (see Jamshidi et al. [3] and many other texts).

In fact, *when* there is a well-defined problem, with a given input-output function and a criterion of approximation, there are several mathematically sound solutions to the design, either by interpolation or approximation, as presented first in a series of volume sections in [5-8], and in a series of papers as in the reference list of [9], with additional details in [10-15].

The main part of the presentation will focus on this issue: the optimization of the design of membership functions, for a specified approximation or interpolation problem. During the development of methods for solving this type of problems it becomes apparent that some of the simplest m.f.s, the polynomial ones, face difficulties with analytic representations of the solutions because fuzzy logic systems' input-output functions cannot be represented by algebraic operations (a result from [12]).

As said, there are several other issues related to FLS design, such as choosing the fuzzy logic [2], deciding on the problem approximation or interpolation and choosing the type of FLS for it, and choosing the right optimization criterion [10-15]. These are not problems pertaining to fuzzy logic systems or to proper FLSs theory and these problems should be addressed apart for completing the well-posed problem of design. However, these problems will also be briefly discussed during the presentation.

Concluding, the theory of FLSs is a mature domain, which reality is not always reflected in the published articles and books, even those published by the most respected publishers. Improving the understanding of engineers of the underlaying FLS theory could improve the design results, decrease the design time, and largely reduce the waste of resources.

References

[1] Reshma V and A. John. *Aspect based summarization of reviews using naïve Bayesian classifier and fuzzy logic.* 2015 International Conference on Control Communication & Computing India (ICCC), 2015, pp. 617–621, doi: 10.1109/ICCC.2015.7432970.

[2] M.H.M. Teodorescu and H-N.L. Teodorescu. *Noncommutative logic systems with applications in management and engineering.* International Journal of Computers, Communications and Control, vol. 16 Issue 1, 2021, Article No: 4082, https://doi.org/10.15837/ijccc.2021.1.4082.

[3] M. Jamshidi, N. Vadiee, and T. Ross. *Fuzzy Logic and Control: Software and Hardware Applications.* Vol. 2, Pearson Education, 1993.

[4] H. Q. T. Ngo and A. S. Tran. *Using Fuzzy Logic Scheme for Automated Guided Vehicle To Track Following Path Under Various Load.* 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), 2018, pp. 312–316, doi: 10.1109/GTSD.2018.8595504.

[5] H.N. Teodorescu and H.J. Zimmermann. *Fuzzy Systems and Signals.* AMSE Press, France, AMSE Monographs Series A, 1992 (158 p.), 1992.

[6] H.N. Teodorescu, T. Yamakawa, and A. Rascanu (Eds). *Fuzzy Systems and Artificial Intelligence.* Litografia Universitatii, Iasi, Romania, 240 p. 1992.

[7] Tofan I., J. Gil Aluja, O. Costinescu, and H.N. Teodorescu (Eds.). *Advances in Fuzzy Sets and Applications.* Editura Universitatii, Iasi, Romania, 1992 (215 p.).

[8] H.N. Teodorescu and E.P. Klement (Eds). *Fuzzy Systems and Applications.* Iasi, Romania, 1992.

[9] H. Ying, Y. Ding, S. Li, and S. Shao, *Typical Takagi-Sugeno and Mamdani fuzzy systems as universal approximators: necessary conditions and comparison.* 1998 IEEE Int.

Conf. Fuzzy Systems Proceedings. pp. 824–828 vol. 1, doi: 10.1109/FUZZY.1998.687597.

[10] H.-N. Teodorescu. *Taylor and bi-local piecewise approximations with neuro-fuzzy systems*. Studies in Informatics and Control, vol. 21 (4), pp. 367–376, 2012.

[11] H.-N.L. Teodorescu. *On the characteristic functions of fuzzy systems*. International Journal of Computers Communications & Control, vol. 8 (3), pp. 469–476, 2013.

[12] H.-N. Teodorescu. *Improve the design and testing of fuzzy systems with a set of (almost) simple rules*. International Journal of Computers, Communications and Control, vol. 17 (1), 2022.

[13] H.-N. Teodorescu. *Coordinate Fuzzy Transforms and Fuzzy Tent Maps–Properties and Applications*. Studies in Informatics and Control. Vol. 24, Issue 3, pp. 243–250, 2015/9/1.

[14] H.-N. Teodorescu. *Non-linear systems, fuzzy systems, and neural networks*. Proc. of the 3rd Conference on Fuzzy Logic, Neural Nets and Soft Computing, Iizuka, Japan, 1994.

[15] H.N. Teodorescu. *On fuzzy sequences, fixed points and periodicity in iterated fuzzy maps*. International Journal of Computers Communications & Control 6 (4), pp. 749–760, 21, 2011.

Horia-Nicolai Teodorescu

Horia-Nicolai Teodorescu
Institute of Computer Science of the Romanian Academy,
Iasi, Romania

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

The Universe as a Universal Quantum Computer: A Road Map for Its Possible Architecture

Veaceslav Albu

Abstract

The paper presents a coherent framework for the architecture of a Universal quantum computer (UQC) under the hypothesis that the Universe represents a complex cybernetic system (CCS). For this purpose, the author poses some questions that such an architecture should answer and construct a persuasive account of UQC's ontology by defining and postulating main concepts, categories, and their relationships. The proposed account permits the UQC's functionality to agree with fundamental concepts of computer science, complex systems theory, and all known laws of nature. Moreover, it pretends to answer some science puzzles such as the quantumness of 3D reality, black matter, and black energy.

Key words: Plirophoria, Principle of dual nature of Information in the Universe, Pliroholon, Cyberholon, Baryoholon, Plirostrand, Pliroknot.

1 The Exordium

The last decade shows the tendency in scientific publications in theoretical physics, astrophysics, computer science, math, systems theory, and philosophy of cosmology to show the importance of the concept of *Information* not only as a research tool. They refer to *Information,* as a fundamental concept of the Universe's ontology with the same degree

of the ontological status of Matter/Energy. They speculate on different aspects regarding the Universe as a quantum computer, a computer simulation, a computer program, a conscious computational being, and active universal information storage [1-4]. This tendency underlines the fact that the modern scientific paradigm based on the reductionism of theoretical physics gets stuck when trying to answer some scientific enigmas, such as expanding the Universe, dark energy, dark matter, and the nature of gravity and quantumness of the real world. This list is not complete. That is exactly what Norbert Winner tried to point out in his famous citation: "Information is information, not matter or energy. No materialism, which does not admit this, can survive at the present day" [5]. This paper represents an attempt to map the main features and principles of a possible architecture for a Universal Quantum Computer (UQC) that resembles the way the Universe functions. Such proposals have to be consistent also with the main scientific facts and theories of modern cosmology. The present framework follows from the author's scientific stand for the equivalence of the E/M/I ontological status in the Universe's ontology and from its *Infogravity* theory that gives birth to the fundamental view on the dual nature of the concept of *Information*. The latter implies the need for the *Postulate of Dual Nature of Information* and the concept of *unchanged evolutionary information* or *Plirophoria* [6]. For future references and the audience's convenience, we will also use the expressions "Wiener's information" to point to the Information of dual nature – as a superposition of embedded in matter information with its ever-produced evolutionary information or Plirophoria, and "Shannon's information" when referring to the classical concept of embedded in matter *Information* of modern scientific paradigm. In addition, in this paper, we will introduce the definition of "*Wiener's Oracle*" that refers to the functionality of the evolutionary information as a mechanism of the universal loop of feedback control in the proposed architecture of the Universal Quantum Computer (UQM). The functionality of any cybernetic system mirrors and resembles its inner architecture and its organization. As we know the main features of the Universe's functionality, the proposed UQM's

architecture and ontology have to obey not only the laws of physics and quantum mechanics but also those of cybernetics, systems theory, and theoretical computer science. For this reason, during the argumentation provided in this work the author will focus mainly on the laws of the latter domains. Nevertheless, references to the relevant research paper from the first domain will be provided when needed.

To construct an ontology of quantum computation by natural kinds that occurs in the Universe, one has to answer the following questions. "*What does the Universe compute?" "Where do the data come from for this computation?" "What is the algorithm and the rules of calculating the input data?" "Where are data stored and input data processed?" "How are Universe's cybernetic systems of the first, second, third, and higher orders organized?" "How are their control centers organized and how does the feedback control loop work?" "How can life and consciousness emerge from the proposed account of computation?" "If the Universe as a computer really exists, how do its ontology's components help to overpass the seminal Observer problem of QM?*" and finally, "*Is the Universe conscious?*" This list is not complete.

2 Main Initial Conditions

The following initial conditions will govern the reasoning of this work. First, in the conditions when the proposed account and framework for the UQC should coincide with the things and objects happening in the real world, the author and the readers of the present work have to be placed inside the Universe's computer, by analogy with any hero placement within the virtual world of a computer game. Having us as referential points or as observers during the virtual tour of present discourse it will increase the level of subjectivity of this approach. As the computational outputs have to obey the laws of nature, any subjective influence on drawing conclusions is undesired. Instead, the referential points are placed in a computer processor or processors, whichever is the case. One can say that the author presents a thought experiment on "how the Universe calculates or produces the real world".

Second, as the UQC output will be the emergent reality governed by laws of nature, only the calculations of natural kinds in the Universe will be discussed. We will shortly remind you that natural kinds refer to the calculations that occur by laws of nature, without any inference of humans or their artifacts.

Third, the proposed UQC's architecture framework has to obey the known laws of physics and resemble the entire Universe's evolutionary history from the Bing Bang in any time and place in the Universe, as demonstrated by science.

Fourth, the work describes the functions, meanings, and performances of the four known interactions namely, gravity, electromagnetic, and strong and weak forces during the performing universal calculations.

Fifth, for the purpose of this work, we take "qudits" instead of "qubits" for quantum calculations, as qudits supposedly possess 32 and more quantum states in the superposition [7-8], whereas qubits do only two, "0" and "1".

3 Postulates and Definitions for the Universe's Computer Ontology

The proposed framework is constructed under the hypothesis of the dual nature of information, by the analogy of quantum duality of matter and energy. In addition, the term *Plirophoria* as evolutionary information and its four properties were presented in [6]. In this section, the main definitions and postulates are formulated.

Postulate_01. ***On dual nature of Information in the Universe.*** The Information represents a fundamental concept in the Universe with the same ontological status, as *Matter/Energy* possesses. The dual nature of *Information* by natural kinds manifests in reality from one side, as the classical (corpuscular, collapsed) state of information embedded in the Universe's matter during the evolution of all classical objects and phenomena of the real world. From the other side, it manifests as unchangeable by its content and physical volume, gravitationally and

topologically encoded *quantum evolutionary information* (*or Plirophoria*) about the history of all collapsed quantum states during the evolution of each classical object of any scale in the Universe as totality of all emerged classical objects during the Universe's lifespan.

From postulated requirement that *Plirophoria* does not change over the Universe's lifespan, neither by content nor by the physical volume occupied in the 3D space of the Universe, it follows that *plirophoria* has to be protected from the influence of other classical and quantum objects and phenomena in the Universe. Gravity remains the only one interaction in the Universe that does not strongly manifest and react on a quantum level. That is why, by analogy with *Energy/Matter* duality carried on by electromagnetic interaction and remaining neutral to gravity at the quantum scale, the Information duality has to be gravitational by nature and neutral to electromagnetic, weak, and strong interactions. In addition, any topological and gravitational encoding remains neutral at a quantum level to all three remaining interactions. That is a reasoning bridge for the introduction of the next definition of a ***Plirostrand***.

Definition_01. ***A Plirostrand*** as a bearer of quantum evolutionary information of an atom (or any quantum particle of Standard Model or any molecule), created by natural kinds, represents a quantum two-dimensional self-assembling structure of specific topology, which gravitationally encodes all its quantum incoherence history from the Big Bang, for all Universe's hydrogen and helium atoms, or for star formation for heavier elements. A *plirostrand* possesses only gravitational and topological properties remaining neutral to the other three interactions that preserve the atom's *plirophoria* content and volume, unchanged during the Universe's lifespan.

Definition_02. ***A Pliroknot*** or ***a plironode*** is encoded topologically and gravitationally on a given *plirostrand* hosted by an atom (or any quantum particle of Standard Model or any molecule) each time when in its evolutionary information a new event of quantum incoherence occurs.

Definition_02A. ***A basic pliroholder*** represents any atom or any

quantum particle of the Standard Model or any Universe's molecule of natural kinds.

Definition_02B. ***A plirohead*** represents any basic *pliroholder*.

Definition_02C. ***An active plirohead*** represents a *plirohead* with an active quantum state, free from any quantum effects with other *pliroheads*. ***A dormant plirohead*** represents a *plirohead* with all its quantum states bounded by other *pliroheads* and acts as a constituent part of an *active plirohead*.

The proposed notion of a two-dimensional *plirostrand* can be compared with a two-dimensional tape of a Turing machine, on which an active *plirohead* writes (embeds) the output information with one exception. If in a Turing machine the head freely moves along its tape in both directions, the *plirostrand* has a *robust* connection with its hosting basic *pliroholder* and *self-assembles* in opposite direction from it, when a collapse of a wave function of this *pliroholder* occurs. Being of a quantum nature, the *plirostrands* of two atoms, for example, of hydrogen, are able to form their superposition or entanglement and form their commonly shared *plirobraid*, as their hosting atoms are quantum connected by chemical covalent bonding.

Definition_03. ***A plirobraid*** is formed when two or more atoms form organic or other types of chemical molecules by quantum entangling together with their individual *plirostrands* and *pliroknots*.

For example, a hydrogen molecule's *plirobraid* will consist of two *plirostrands* of coupling two hydrogen atoms to form the molecule. The water molecule's *plirobraid* is formed from the *plirostrands* of two hydrogen atoms and one oxygen atom. As a science of chemistry and biology describes the emergence of objects of higher complexity levels, like living cells, for example, the increased levels of complexity of a given emerging classical object will have its *plirobraid* as layered networks of relevant *plironodes* and *plirostrands* of their inner *basic pliroholders*, accordingly with its level of complexity.

For the sake of the completeness of the Universe's Ontology, we will admit that a *plirostrand* also consists of a superposition of *plirostrands* that correspond to all constituents of protons and neutrons that form

the given atom. In their turn, the *proton's plirostrand* contains *plirostrands* of respective quarks and gluons forming this respective proton. From this consideration, a *plirobraid* of a hydrogen molecule contains *two layered plirobraids* of both atoms together with their constituents – protons, quarks, gluons, and one electron. Moreover, from the above definitions and *Postulate on the dual nature of Information* in the Universe follows that any real object in the Universe at any scale – from atoms to galaxies – consists not only of known baryonic matter but is always surrounded by self-assembling quantum evolutionary information attached to it – its *plirophoria*.

The *holarchy* and *holonic networks'* organizational structure become popular in constructing highly effective self-learning and self-organizing complex systems [9-10]. Indeed, by definition, such structures contain at each level *holons*, which in their turn, have a *holonic network* substructure. For example, the Universe's *holarchy* first level consists of gravitationally interacting gigantic superstructure of *galaxies superclusters* and *voids*. As the *first level holons*, they consist in their turn of smaller galaxies clusters and voids as the *second level galactic holons*, that consist in their turn of smaller gravitationally bounded groups of galaxies and intergalactic voids. From one side, from the big-scale cosmic structures point of view, galaxies represent some *basic level of holons*. However, this is at first glance. By analogy with the zooming fractality of *Mandelbrot's set*, each galaxy seemed as a small point at this level, can be zoomed up to the size of the Milky Way galaxy, for example. From the other side, one can clearly conceive the inner fractality of its holonic substructure by zooming in up to stars' planetary systems, gravitationally bounded together within our galaxy. This path can be zoomed down to the *holon of Earth* as a holarchy of the Tree of phenomenon of life, where each constituent holon as a living organism manifests itself as a fractal holarchy of its constituent holons, down to the living cells, as basic holons for the phenomenon of life on Earth. Nevertheless, it is not the end of the journey that begins at the Universe's first level of holarchy. A cell biologist can easily virtually dive into the inner universe as a holarchy of

a living cell, where billions of complex molecules do their meaningful work to sustain the cell's functionality, being bound not only by the gravity, but mostly by electromagnetic interaction. They will prolong this journey up to the level of atomic holons that in their turn possess inner holonic fractal structure, as the Standard Model predicts.

Definition_05. *A visible Universe's holon* or ***baryoholon,*** represents any self-assembling, self-organizing, self-replicating, self-preserving from the environment, self-learning, ***self-evolving, and environmentally dependent complex classical object composed of baryonic matter***, that has a distinct beginning point of its evolution. It links it with *one or a few holons-ancestors*, and by analogy, a ***baryoholon*** can possess an inner ***complex self-replicating structure*** by hosting evolutionarily formed inner ***baryoholons*** with the same properties.

One can take an individual animal, a star, a living cell, a planet, a galaxy, a brain neuron, a brain, and finally, the Universe itself as examples of such ***holarchy of baryoholons***.

Definition_05A. *A* ***Pliroholon*** represents the *plirophoria* of a *baryoholon* organized as a *holarchy* of quantum *holonic neural networks* of *plirostrands, plirobraides, and pliroknots* respectively, that resemble the inner structure of related *baryoholon*. A *pliroholon* exists in two different states – an ***emergent pliroholon*** and an ***unloaded pliroholon***.

Definition_05B. ***An emergent pliroholon*** represents a self-assembling *pliroholon* driven by its *hosting baryoholon's evolution* and resembles all its evolutionary information in the manner, that the lifespan evolutionary path of its *hosting baryoholon* can be fully restored or replicated. The information about *ancestors and descendent holons* holds together with its evolutionary information.

Definition_05C. ***An unloaded pliroholon*** represents the unloaded *plirophoria* of a given *baryoholon* that ceases to exist in its *homeostatic equilibrium state* and terminates its evolution as a homeostatic feedback control ignites a *continual singularity state* due to changed homeostatic conditions.

During this singularity, *the baryonic evolved structure* gets into active *self-disassembling states*, when all inner lower-level *baryoholons* get disassembled down to the lowest quantum objects *basic pliroholons* (molecules, atoms, particles) admitted by such *singularity* conditions. Such continual singularity conditions also can ignite the environmental conditions of *self-assembling complex atoms* for the next cycle of a hosting galaxy (as a holon) evolution. The *singularity* represents from one side, a clear endpoint for a given *baryoholon*'s evolution and, on the other side, a beginning point for a *new cycle of evolution* that can influence the evolution of both *ancestors* and *descendent holons*. Simultaneously, this s*ingularity* ignites the process of detaching *an unloaded pliroholon* from its *hosting baryoholon* at all its levels of inner *holarchy* structure.

When a classical object ceases to exist, its *pliroholon* detaches from the dead body and its space journey in the open Universe begins. As the environment temperature is almost ' **0** ' degree Kelvin, there are comfortable conditions for self-storage of quantum *Plirophoria's structure*.

That perfectly fits with the introduced definitions for the Universes' *plirophoria* structure as *holarchy* of respective *pliroholon* attached to it hosting *baryonic holon*, with one exception. A *pliroholon* contains the evolutionary information about all ever-formed combinations of baryonic matter in this *baryoholon* and the *plirophoria* of all inner *holons* during that holon lifespan. In addition, it fits with the scientific approaches to the *fractal structure of the Universe* [11], which resembles the *holonic* structure discussed above.

Postulate_01A. The Universe represents a unity and superposition of dynamically evolving *holarchy* of *holons* and *holonic multilayered networks* of two kinds. ***The first kind*** represents a constant by volume of matter/energy content ***evolving holarchy of layered holonic networks*** of *baryoholons* in superposition with its *evolving pliroholons*. That coincides with visible, expanding and detectable Universe by means of any of four known interactions. ***The second kind*** represents the increasing only by volume and mass, detached from any *baryonic*

matter, topologically connected holarchy of all *unloaded pliroholons*, which detaches from hosting baryoholons during the *continual singularity* at the end of their lifespan. That represents ongoing process of growth of the *volume and content of the unchangeable evolutionary information* of all ever evolved of any kinds of *baryoholons*. From cybernetic systems architecture point of view, the second kind of expanding *holarchy* of unchangeable by volume, topology, and *gravitationally encoded evolutional information* represents *the storage room for output data* of a visible calculating Universe.

Postulate_01B. The ***Universe's plirophoria*** represents an ongoing process of gaining new kinds of increased complexity evolutionary information from a given by *Big Bang event* the amount of baryonic matter and energy as input data for the visible Universe as a *self-evolving* and *self-assembling, self-learning holarchically* organized cybernetic system of increasing order by its evolution towards evolving complexity of its inner *baryoholons*.

Postulate_01C. The visible Universe possesses a *twofold nature*. Firstly, the Universe is ***fractal by nature*** and its evolution has an *identical* ***dialectical pattern of evolution*** at all levels of its inner *baryoholons*. The evolutional changes that occur in each *baryoholon* towards increasing complexity of embedded in its baryonic matter information from its beginning (***thesis***) towards the end of its lifespan (***antithesis***) ignite the continual singularity (***synthesis***). Via ***synthesis's singularity***, a new cycle of increasing of complexity of the *embedded information in its baryonic matter* begins. *Secondly*, the ***Density of Emerging Information (DEI)*** (See the Postulate_04A) at the quantum level represents the *fractal dimension of the Universe*. Following that, as the next step in the evolution path of each *baryoholon*, will be chosen the one, which will guarantee a higher ***DEI*** for the next step of evolution of this *baryoholon* that in its turn, will imply an *increasing complexity of embedded information in evolving matter*.

Arriving here, one can ask "What about conservation laws in the proposed ontology of computable Universe? I was promised that discussed framework would obey the laws of physics, as well as the laws

of complex systems theory. All baryonic matter is well counted by the Standard Model, no mass left for the proposed account of Plirophoria?" Here we will address the famous Landauer's principle, which surprisingly bridges the Second Law of Thermodynamics with the energy cost of erasing a bit of information during the calculations on a classical or quantum computer as

$$E = Kb * T * \ln 2, \tag{1}$$

where ***Kb*** *is the Boltzmann constant and T is the temperature of a "reservoir" with which the bit exchanges heat* [12-13]. For the purpose of the presented theory, we will reformulate this principle to the quantum computations by natural kinds. Here we will take into account the above *pliroknot's* definition and admit as elementary quantum calculation by natural kind any quantum effect that makes a *pliroknot* to be encoded on relevant *plirostrand*.

Postulate_02. ***Landauer's Principle for Quantum Calculations by natural kinds.***

For the purpose of quantum calculations by natural kinds the Landauer's principle can be reformulated, based on the relativistic matter/energy/information equivalence principle initially formulated by Albert Einstein in 1905 [14]. A *pliroknot* is topologically encoded into the relevant *plirostrand* when the related quantum effect occurs. The *pliroknot* rest mass is given by the formula

$$Mpknot = (Kb * T * \ln 2)/c^2, \tag{2}$$

where ***c*** is the speed of light, or

$$Mpknot = 3.19 \times 10^{-38}\text{kg}. \tag{3}$$

The volume and spatial dimensions of the newly created mass are smaller than Plank's length, following that they cannot be described by quantum mechanics, limited by Plank constant.

Postulate_2A. The newly created mass according to **Postulate_02** possesses few fundamental properties. Being of gravitational and quantum nature simultaneously, it has a ***constant volume*** and different

topological structure to embed with early created *plirophoria* in the atom's *plirostrand*, forming a two-dimensional structure with ***constant volume and different topology***. Each time a new emerged *pliroknot* increases the volume of its *plirostrand* with the same quantum of mass and volume of special topology. Two *plirostrands* occupy *separate volumes* in the Universe's space.

Postulate_2B. The *plirophoria* within a *baryoholon* possesses indestructible discreet mass and three-dimensional volume in space that cannot be occupied by another *plirostrand*. However, any *baryonic matter* and its *plirophoria* are in superposition in the same volume of space.

Postulate_2C. The organized *plirophoria* possesses a self-organizing property in a given volume of space that by analogy with the magnetic effect in the self-assembled volume of atoms increases the gravitational effect, which we name ***gravitational momentum (GM)*** in that given volume, occupied by a given ***baryoholon***.

At first glance, the rest mass is of the same order of magnitude as Plank's constant. However, one should have in mind the quantum amount of quantum calculations by natural kinds on the level of galaxies and of the Universe. In [15] the author presented the following amounts of encoding *pliroknots* per second in the Sun and in the entire biosphere of Earth, respectively $\boldsymbol{N}\mathbf{sun} = \mathbf{1.08} \times \mathbf{10}^{39}$ and $\boldsymbol{N}\mathbf{earth} = \mathbf{3.0\ x10}^{38}$. Incredibly, the Sun that is millions larger than Earth, due to complex systems evolution within Earth's biosphere, the amount of produced *plirophoria* is of the same order of magnitude.

The *pliroholon*'s definition and the above considerations bring us to the idea that the Universe exists not for the purpose of life or complex evolving matter, but for the purpose to yield up much evolutionary quantum information, *Plirophoria*, of the highest complexity. If so, the problem of how the *Plirophoria* is stored appears. It cannot mix up due to its structure introduced in this section of the *holons'* network. It is like separate memories in one's brain, which can be forgotten, but not mixed up. The main conclusion thereby follows that *plirophoria* represents the "grains of baryonic soil in the Universe". We have some

emerging plirophoria that is still in the garden; but another, unloaded pliroforia are already stored in an inner galactic 3D space. As galaxy rotates, it makes *unloaded plirophoria* to run out from the edge of the galaxy and it goes into outer intergalactic space. However, *there are unloaded plirophoria* from other neighboring galaxies. Whether unloaded plirophoria from different galaxies should mix up or repel each other then? By definition, the volume of *plirophoria* is indestructible, even if the mass is very small. Follows, that the volume of the empty space itself will begin growing to store all *unloaded plirophoria* from the neighboring galaxies. One can imagine a clear analogy with a picture of a "blowing Swiss cheese". Therefore – here we are – the possible ontology of black matter and black energy is constructed.

Conclusion. In this section, we introduce an account that resembles how the quantum evolutionary information *Plirophoria*, as the second state of Information, gets its self-assembling, self-organizing, and autodidactic properties. This account is based on the formulated Postulate of the dual nature of Information in the Universe. That is a key for further description of the architecture of the Universe as a computer. To do that, we first need to introduce the account of quantum calculations by natural kinds and its ontology.

4 The Ontology of Quantum Computation by Natural Kinds

Definition_07. *Any stable atom within the Universe*, together with its *attached plirophoria*, represents a ***Stable Elementary Quantum Computing Unit (SEQCU)*** of the Universe by natural kinds that can become part of a bigger self-assembling quantum automaton only under the action of two interactions – gravity and electromagnetism.

Examples of *SEQCU* in nature are any stable atom, any stable molecule, any organic molecule, or proteins. As we mentioned before, Universe's quantum computation consists of qudits, because they can hold all quantum states of such complicated quantum objects as proteins or heavy atoms in superposition.

Postulate_03. The Universe represents a dynamical self-propelling, self-assembling, and self-computing holonic network holarchy acting simultaneously in both realms – quantum and classical. The computation process consists of an ongoing emergent superposition of two phenomena – *collapsing and activating quantum states of* ***SEQCU*** *natural kinds in all Universe's holons simultaneously.* Collapsing quantum states makes ***SEQCU*** bounded within its quantum system, which changes the classical encoded information in that system. At the same time, a *pliroknot* is encoded in a relevant *plirostrand* of *SEQCU holon's* quantum evolutionary information.

Postulate_03A. The Universe represents by its fractal nature a dynamically evolving self-organizing and autodidactic cybernetic system of evolving order organized as a *holarchy of holonic quantum networks. Holons* on the basic level represent *quantum cybernetic systems of the first order.*

To make it more clear, we can compare computing by a SEQCU as one headed quantum Turing machine with two bands that simultaneously evolve in two opposite directions – forward and backward in time. One band represents a quantum system from the real world, to which this SEQCU represents an inner part; the other band is a SEQCU's *plirostrand*, where the head simultaneously encoded respective evolutionary information as of the collapsed quantum event. The first band represents the only one present state "the NOW", and the second band, "the PAST" – the SEQCU's *plirostrand* with the history of all encoded NOW states, which passed away from now.

Postulate_04. *The computing process* in the Universe as a *holarchy* aims to produce maximum complex *plirophoria* from a given finite amount of baryonic matter in each of its *holons* by combining all contained baryonic matter in them in all possible manner that goes under the supervision of the four Universe's interactions.

Postulate_04A. *Two main kinds of computation* processes by natural kinds occur simultaneously in each *Universe's holon.* The ***First kind of computation*** aims to increase the complexity of the Universe's atoms from a given amount of protons and neutrons after Bing Bang's

baryogenesis epoch from the simplest hydrogen, helium, and lithium atoms to the most complex atoms evolved by natural kinds in such events in a cosmic scale as an evolution of higher stars generation, a collision of neutron stars and black holes. This list can be prolonged. Important is that the *First kinds of computations* are subject to all four known interactions for their proper functioning. The type of *Universe's holons* that holds the *First type of calculations by natural kinds* we define as ***baryoholons***. The ***Second type of computation by natural kinds*** aims to increase the complexity of computed complex matter composed of molecules from a given amount of stable atoms. We define this type of holons as ***cyberholons***, by analogy with cybernetic self-organizing complex systems of first and higher orders. As the first kind of process is governed by all four interactions – gravity, electromagnetism, weak and strong interactions, and the laws of the quantum realm, taking place in stars and other similar cosmic objects, the second kind of computation takes place under the supervision of gravity, electromagnetism, and laws of quantum mechanics.

Definition_07A. We will name the matter that is basic for the first type of calculations by natural kinds in *baryoholons* governed by all four types of interactions to form the next generation of complex matter as ***baryomatter***. The matter that is basic for the second type of calculations by natural kinds in *cyberholons governed by gravity and electromagnetism to form the next generations of complex matter* will be named ***cybermatter***. Any *Universe's holon* can consist of inner *baryoholons and cyberholons,* following that it can contain both kinds of matter defined above.

For example, our Solar system being an inner *holon* for the Milky Way galaxy contains inner *holons* like the Sun, a *baryoholon*, and its planets, where Earth represents a typical *cyberholon*.

Postulate_04B. The goal of both kinds of calculations by natural kinds is to achieve a maximum ***Density of Emerging Information (DEI)*** that corresponds with quantum effects in a given amount of baryonic matter in a given Universe's holon. From one side, these quantum effects embed new information in that matter. On the other

side, these increase the *holon*'s *plirophoria* with the same amount of *priroknots* on the respective *plirostrand*.

Postulate_04C. The Universe computation process constructs all real objects of the visible Universe and the Universe's plirophoria by ongoing computing process of these two kinds of computations at all its instances and levels from the quantum realm up to main galactic structures. This ongoing unstoppable process of quantum decoherence implies the emergence of the next generations of active quantum states on the edge of each SEQCU and UEQCU that embeds new emergent information in given instances of classical matter.

Postulate_04D. Electrons, photons, and neutrons play the role of mediators of quantum computing processes in neighboring or distant baryoholons and cyberholons by carrying mass, energy, or information that mediates the calculation processes taking place as a cyclic quantum process, governed by electromagnetism or gravity. Such events produce ***reversible calculations*** in a given Universe's *holon* by natural kinds. Two calculation processes stated in *Postulate_04A*, represent the ***irreversible universal calculations*** by natural kinds.

As an example of reversible calculations, one can take the energy transportation described by an atom's nucleus photon capturing and emitting, by changing its electron's energy levels. Same with nutrition or oxygen transportation chain for a living cell. Even they occur as quantum effects at the covalent bonds level, which possess a cyclic character and do not play a direct role in the inner structure of a living cell. By contrast, the copy-making DNA strands change the structure and inner content of the cell that will disappear only with the cell's division or death.

Postulate_05. Each of the two kinds of computational processes by natural kinds produces different types of Plirophoria and cosmic matter with different properties, relevant to the processes and environments that produce them.

Definition_09. Any unstable molecule, atom, or quantum particle of the Standard Model, together with their attached primordial plirophoria, represents an Unstable Elementary Quantum Computing

Unit (UEQCU) of the Universe by natural kinds, which can become parts of stable protons or neutrons under the action of weak and strong interactions.

As the stable atoms play the role of main computational qudits, we need to have an explanation of how they evolved to their stable state and what environment governs such an evolution from primordial Universe's plasma.

Definition_10. The dynamical evolution of the Universe possesses fractal nature in all its *holons* driven by calculation processes that occur in their SEQCU and/or UEQCU respectively. The *reversible calculations* represent chains of informational or energy supply or communication within a given holon or between neighbor *holons* in the *hosting holon*. The *irreversible calculation* drives the evolution of structure within each *holon* that holds the lifespan of a given *holon*.

Postulate_5A. Any of two kinds of Universe's *holons* forms within a hosting holon in the Universe's holarchy under the cumulative effect of the gravity and electromagnetic interactions for *cybermatter* and gravity and any from other three interactions for *baryomatter*, contained respectively in the environment of a *hosting holon*. The formation of a *new holon* within *a given holon* is called ***holon synthesis*** or ***holosynthesis.*** The ***holosynthesis*** occurs when the *Density of Emergent Information (**DEI**)* in a given volume of *baryomatter* or *cybermatter* and its environment reaches the *threshold* for the ***formation of a new holon***. At this *threshold*, a ***process of continual singularity*** self-ignites due to feedback loop control of a continual homeostatic equilibrium over the involved matter towards a new state of equilibrium with a higher *Density of Emergent Information* (***DEI***).

Postulate_5B. The ***holosynthesis*** in ***baryoholons*** ignites the processes of different generations of stars formation, from white dwarf up to neutron stars and black holes. That kind of ***holosynthesis*** takes place under the governance of all four fundamental interactions and quantum mechanical effects towards increasing ***DEI*** in a given matter. The ***holosynthesis*** in ***cyberholons*** ignites within a given amount of baryonic matter when environmental conditions under the governance

of gravity and electromagnetism pass the complexity threshold for self-replicating and self-assembling of a SEQCU and a UEQCU in a given environment, which leads to its increased ***DEI***.

Postulate_5C. Because both types of calculation are performed by quantum effects, it leads to the connection between hosting holon plirophoria and plirophoria of new complex *SEQCU* in a *new holon.* In other words, the new *SEQCU* were connected with the ***main operational memory of*** a *hosting holon* that biases also to all memories of *neighboring holons.* As *plirophoria* is quantum by definition, follows that the *plirostrand* of a *newly formed SEQCU* is connected with *plirophoria* of the entire Universe. *That connection gives to respective* ***SEQCU*** *knowledge how to calculate with the purpose to increase the* ***DEI***.

Postulate_05D. The lifespan of a *baryoholon* ends when the highest value of *DEI* is reached during its active life decrease below the given threshold for such *kind of holons*, controlled by the *hosting holon's plirophoria* or its ***control center***. That ignites a *self-destruction continual singularity process of a given holon.* During this process, new types of SEQCU occur and UEQCU are created and baryonic matter of the vanishing holon becomes ready for a new lifespan *baryosynthesis* cycle. The *self-destruction continual singularity* ignites the detaching process from the given *baryoholon* of its *plirophoria* that begins its own journey in neighboring empty space.

Postulate_05E. All Universe's black holes represent an *evolving continual singularity* in a given *baryoholon* from the moment of the Big Bang until nowadays. There are two types of black holes in the Universe. First type – *the primordial black holes*, which represent all galaxies' central black hole. Second types - the *evolved black holes* that appeared due to evolved big-sized stars within a given galactic *baryoholon.* As continual existing singularities in a given *baryoholon*, the black hole plays a role of a ***gatherer of baryoholon's plirophoria*** in a given volume of space by detaching under the singular conditions the *pliroholon* from its hosting *baryoholon. The baryonic matter of a baryoholon falls into the black holes, where its pliroholon remains detached*

and follows its path within the hosting galaxy.

Postulate_06. As evolution represents an ongoing process, and many of both kinds of *plirophoria* are created from the baryonic matter in the same neighborhood of space within the *hosting holon*, there is a possibility that many copies of *plirophoria* quantumly stick together unifying and increasing their memory and knowledge strength. Nevertheless, from ***Postulate_02F*** and ***Postulate_02B*** follows that this new structure possesses a volume, equal to the linear sum of each involved copy.

Postulate_06A. From the *postulates 2A, 2B and the mass of a pliroknot* (3) it follows that the mass of ***an evolving pliroholon*** in comparison with the mass of its ***hosting baryoholon*** differs in many orders of magnitude. However, they both occupy the same volume of 3D space, being in quantum superposition. That is why in two gravitational interactions of two baryoholons, their *pliroholons* follow the known laws of nature of interactions between this pair of baryoholons and their gravitational effects are so small, that do not really influence the output of that interactions. However, the evolving *pliroholon* occupies the same volume as the *hosting holon*.

Postulate_06B. The lifespan of a ***cyberholon*** ends when the average value of ***DEI*** in a given volume of baryonic matter decreases dramatically below a given threshold. As the ***DEI*** is much higher in *cyberholons* than in ***baryoholons*** due to the higher complexity of produced SEQCU (as differ in complexity enzymes and oxygen molecules in living cells, or the living cell itself, for example), their size/mass are by orders of magnitude lower size/mass for ***baryoholons***.

Postulate_06C. For ***cyberholons*** their control center has a distributed structure by analogy with cybernetic systems (or *holarchies*) of the first and higher orders, with distributed *self-adapting, self-learning, and self-improving properties*. An important and common part of knowledge is stored in the main control center of *such cybersystems*. The often-used knowledge and subprograms are distributed and hosted within that *holons*. The ***holosynthesis in cyberholons*** by contrast with ***baryoholons***, possesses a ***two-steps recursive level of self-***

learning and fractal self-evolving towards increasing DEI as their fractal dimension. *From one side*, the plirophoria from a given hosting ***cyberholon*** takes a path towards its evolution and places a specially formed code named ***pliroDNA*** that embeds the main rules on how to reach that specific point in ***cyberholon*** evolution; and many generations of *inner cyberholons* are created, beginning from this point in their evolution. *On the other side*, each new inner ***cyberholon*** enriches the *plirophoria* of its hosting ***cyberholon*** that improves the decision power of its control center and brings new knowledge to the entire system. From its turn, that implies an improvement in inner holon's ***pliroDNA*** by the feedback control loop.

Postulate_06D. *Baryoholons and cyberholons* possess discreet outer border, where their *plirophoria* resides named ***pliroshell***. That permits them to keep organized their inner recursive fractal holarchy structure of any holon at all levels of the Universe's holarchy.

5 The Quaternionic Nature of a Wave Function

As a matter of fact and from the above definitions, a *cybermatter* of a living cell possesses different properties and different DEI in comparison with the same volume of a baryomatter from the Sun, for example. That is because both represent the output data of the two different computational processes producing them. However, for the Standard Model, a collapse of the wave function for any quantum particle is described by Schrodinger's equation

$$\frac{-\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} = i\hbar\frac{\partial\psi}{\partial t}. \tag{4}$$

We will not explain in detail each term from this seminal and well-known formula. For the sake of this work, we will take only one term – the imaginary unity "i". From one side, it is well known that imaginary unity $i = \sqrt{-1}$, or $i^2 = -1$ and it represents the imaginary part of complex numbers, where the classical representation of a complex

number is $a + ib$, where a and b are real numbers. From the comparison presented above follows that the Standard Model and Schrodinger's equation do not make difference between an electron and a quark or a photon as at quantum scale all of them can be approximated by one or two-dimensional wave. Nevertheless, for purpose of our Universe's computational ontology, we need to make this differentiation between *cybermatter* and *baryomatter*. Moreover, during the last decade, there were many proofs that a wave function is not only a theoretical construct but also necessarily a real entity [16]. Spiking in the way of theoretical physicist, for the sake of our work we need to color the wave function of a quantum particle or a system in the "color" of respective interaction, which dominates the quantum particle's computing process stated above. Finally, we need only three different imaginary units, one for each from strong, weak, and electromagnetic interactions. We do not consider gravity for this purpose as the gravity effect at the quantum level does not collapse from quantum states. We have the possibility of such differentiation thanks to the genius of Irish mathematician William Hamilton who proposed the *quaternions theory.*

Quaternions are generally represented in the form

$$a + bi + cj + dk, \tag{5}$$

where $a, b, c,$ and d are real numbers; and $i, j,$ and k are the basic imaginary quaternions with the following properties $i^2 = j^2 = k^2 = ijk = -1$, and $ij = k$; $ji = -k$. From the above, it follows that the given quantum particle will have the equation in its respective wave function (1) instead of only one imaginary number i, the one which corresponds to the governing the holon's particle interaction. *Let us say, i – for electromagnetism, j – for weak interaction, and respectively, k – in case of strong interaction.* One quaternionic approach for quantum calculations is presented in [17]. Saying that we can formulate the following.

Postulate_07. The wave functions for quantum particles governed by electromagnetism, weak and strong interactions are described by the Schrodinger equation in general form as

$$\frac{-\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} = q\hbar\frac{\partial\psi}{\partial t}, \tag{6}$$

where, instead of i, we have $q = a + bi + cj + dk$;

Follows that respective wave functions for different interactions are described in three orthogonal imaginary subspaces and all three are real, as follows from [16].

Following that Universe's computation takes place simultaneously in four subspaces, as follows:

$$\{X, Y, Z\}, \{X, Y, Q\}, \{X, Q, Z\}, \{Q, Y, Z\} \tag{7}$$

as one Euclidian 3D subspace and three quaternionic subspaces, each governed by one of *gravity, electromagnetic, weak, and strong interactions*, respectively.

Postulate_07A. From the Postulate_07 follows that the Universe's computing holonic networks of SEQCU and UEQCU, simultaneously compute in four above-mentioned subspaces. As the Universe is real, only one possibility follows, that the Universe is evolving within the superposition of described four three-dimensional subspaces (7). From the simultaneous superposition of four spaces it follows, that any point (x, y, z) in the three-dimensional Euclidian space of our reality does represent a superposition of four points as

$$\{(x, y, z), (x, y, i), (x, j, z), (k, y, z)\} \tag{8}$$

that finally represents a point in the four-dimensional space $\{X, Y, Z, Q\}$. That superposition represents the main reason for the quantumness nature of the quantum realm.

The **Postulate_07A** will be needed for our explanation of the Big Bang's ontology very beginning *epochs*, including the inflation epoch up to the end of the baryogenesis epoch, when the Universe got all its *SEQCU* and *UEQCU* ready to begin the computation processes at all levels of its holonic network, at the very beginning of their evolution.

Deep and concise introductions in the *epochs* of the dynamically evolving Universe from the Big Bang up to big cosmic structures formation can be found in [18 – 20].

6 The Ontology of the Universe's Feedback Homeostatic Control Center

Definition_11. The Universe's *holon control center* is of dual nature and represents from one side a quantum superposition of *plirophoria* contained by all *inner baryo or cyber holons*, as well as of the *plirophoria* of its *ancestors hosting evolving holons*. We name it an ***active evolution control center***. On the other side, it represents a quantum by nature structured detached *pliropholons* of all previous stages of the evolution of related generations of *baryoholons*, named as ***evolutionary control center***.

Postulate_08. The evolution of the Universe occurs under the supervision of the ***Universe's Control Center or UCC***, which represents a quantum superposition of entangled states of ***holarhic networks of plirophoria*** of all kinds of the Universe's actually *evolving holons* or of *detached pliroholons* of ever evolved *baryoholons or cyberholons.*

Postulate_08A. The ***Universe's Control Center (UCC)*** resembles the holarchic structure of the Universe and represents a holarchy of ***Control Centers (CC)*** from the ***UCC*** at first level and as a *holarchic* network of CC at each level down to ***basic pliroholder (Definition_02A)***.

Postulate_08B. *The* ***UCC*** *and the* ***CC*** *at all Universe's levels* ***by natural kinds*** represent a ***self-assembling, self-organizing, self-didactic, self-sampling, and self-sufficient evolving multilayered holarhic networks of CC***, with only one *layer of visible knots* that coincide with all the Unverses' *pliroholders* (Definition_2A) and of a multilayered *holonic* networks of a hidden *pliroknots* as quantum connected *pliroholons.*

A computer science researcher can easily detect in the above *postulate* the description of ***deep learning multilayered Boltzmann***

quantum machines that are of increasing popularity nowadays in all advanced Artificial Intelligence implementations.

Postulate_08C. The Autodidactic Universe and the Laws of Nature. The Universe represents by its nature defined and postulated above an *autodidactic holarchy of multilayered holonic networks of pliroholons* that evolved from the Big Bang event. The autodidactic process goes simultaneously with the evolution of all Universe as fractal *baryoholons* and *cyberholons* towards the Universe's fractal dimension – maximization of DEI in a given matter. *The learned paths represent the all known laws of nature and laws of physics.*

Postulate_09. ***The Observer Postulate.*** The collapsing active state of each plirohead of each *Universe's holon* is aimed to maximize the *DEI* in involved *baryomatter* and *cybermatter* with the purpose to produce new material objects of increased complexity, that guarantee *new knowledge and data* within its *evolving and stored plirophoria.*

For this purpose, any forward incoherence state of each *active plirohead* within its quantum states of a given matter occurs with the probability that increases the future DEI by embedding a relevant quantum state in a given matter. By its essence, the learned laws of nature represent the ***Observer*** within each ***active plirohead*** that is connected with the ***Universe's Control Center***, where all ***laws of nature nest.***

That postulates the clear conditional probability of the future evolution of the quantum system towards increasing its total DEI, not only in the given matter in that quantum volume of space, occupied by this system, but also towards increasing DEI of entire *baryohlon* or *cyberholon* that hosts given quantum system.

Postulate_09A. The Postulate_09 represents by its essence the *Bayesian theorem* formulated for each quantum system in the Universe, under the governance of the CC of its hosting *baryoholon or cyberholon.*

Postulate_09B. All statistically differentiated rules, self-learned and self-memorized by UCC as the Universe's holarchic network of quantum CC all knowledge that optimizes the collapsing quantum states towards increasing DEI represents all known laws of nature and laws of physics

of the Universe.

Definition_12. The collapsing of a quantum system postulated in postulate_09 represents an observation of that system, defined in quantum mechanics as an Observer problem paradox.

Definition_13. By Wiener's Oracle, we name a Control Center of any level of complexity from the basic plirohead up to UCC that forms its decisions (or observations) by activating for any *plirohead* the relevant law of nature, that guarantees that the holon's evolution goes towards increased DEI, as the Universe's fractal dimension requires, as postulated in postulates 9A, 9B and 9C. The Universe's UCC are organized as a web of Wiener's oracles that act as a Universe's consciousness.

Postulate_10. The Transactional Interpretation of Quantum Mechanics. The Universe's Quantum Computer Control Center acts in two realms simultaneously, in 3D real world and in the quantum realm, by fulfilling the *requirement of time-reversal symmetry* as described by Transactional Interpretation of Quantum Mechanics [21-22]. Where the retarded waves interfere with the visible plironodes of a given pliroholon, as postulates 08, 08A, 08B, and 08C describe to get by the feedback loop control the relevant law of nature according to which the wave function will collapse, imposed by an advanced wave from the *plirohead.*

7 Possible UQC's Architecture

The purpose of this work is to present a possible architecture of the Universe under the hypothesis that it represents a computing machine. In previous sections, we gave the main definitions and postulates of the ontology, which allow us to conclude the following.

Definition_14. The Universe represents the holarchy of cybernetic systems of a different order of complexity by natural kinds organized as holonic multilayered autodidactic and self-assembling networks evolving by quantum computation of all Universe's baryoholons and cyberholons under the laws of nature governed by the holarchy of a Wiener's oracles as a control center of Universe's feedback control, driven by

time-reversal symmetry.

From this definition and from the above sections we can describe the architecture of the ***dynamically evolving Universe's quantum computer (UQC)*** as follows.

1. The baryonic matter of all visible Universe at all its instances of their level of complexity represents from one side the 3D volumes of input data for all Universe's SEQCUs and UEQCUs (Definitions 7 and 8). On the other side, it represents 3D volumes of output data of the ongoing evolving dynamical process of embedding new information by the process of continual incoherence in this matter on the quantum level.

2. The calculation process itself, by natural kind, represents an ongoing recursive quantum decoherence process of quantum states of active *pliroheads* (Definition 2A and 2B) and of simultaneous emergence of a new active quantum state of a *plirohead*, with a purpose for future decoherence.

3. The calculation process also continually produces *gravitationally preserved and topologically stored* evolutionary information as *plirophoria* of all quantum events ever occurred.

4. The Universe's calculation process is oriented toward producing the most complex baryonic matter of two kinds – textitbaryoholons and *cyberholons*, as described above.

5. The Universe consists of matter that possesses dual nature. From one side it produces a classical object of the surrounding environment of all levels of complexity. On the other side, it continually produces the *evolving and preserved plirophoria* (Definition 5A and 5C), which can be the part of explanation ontology for the Universe's dark matter and dark energy. Where the *evolving plirophoria* represents the Universe's dark matter and the *preserved plirophoria* represents its dark energy.

6. The architecture of the Universe's control center coincides with a multilayered deep Boltzmann machine as it consists only of one visible input/output layer of pliroheads.

7. The calculating rules represent all laws of nature that the Universe learned as statistical rules of the conditional probability of incoherence of active quantum states with the purpose of DEI maximization.

8. The Universe represents a dynamically evolving fractal of *baryoholons* and *cyberholons*, with maximization of the DEI as its *fractal dimension*.

As models of abstractions for real calculations in all holons of the Universe's holarchy described in this work in accordance with the presented ontology, one can take the models proposed by Stephen Wolfram in 2020 [23].

8 Conclusions and Further research

This work introduces an account of ontology for the Universe as a computer. This account enables us to answer the questions formulated in the first sections. Nevertheless, the other queries; for example, "*How did Big Bang happened*?" remains unanswered as a goal for future research.

Acknowledgments. The author wishes to thank Professor **Averkin A. N.**, for his time and productive discussions on the above topics and for his advice of introducing the Greek word ***pliropho-ria*** to denominate the second dual state of information, postulated by the Principle of dual nature of Information in the Universe. The research was partially supported by project 20.80009.5007.22. "Intelligent information systems for solving ill-structured problems, processing knowledge and big data".

References

[1] Stephon Alexander, W. J. Cunningham, J. Lanier, Lee Smolin, St. Stanojevic, M. W. Toomey, and D. Wecker. *The Autodidactic Universe*. arXiv:2104.03902v2 [hep-th], 2 Sep 2021.

[2] R. Penrose and H. Zenil (Eds). *A Computable Universe. Understanding and Exploring Nature as Computation*. World Scientific Publishing Co., 2013.

[3] V. Vanchurin. *The world as a neural network*, arXiv:2008.01540v1 [physics.gen-ph], 4 Aug 2020.

[4] Irun R. Cohen and Assaf Marron. *The biosphere computes evolution by autoencoding interacting organisms into species and decoding species into ecosystems*. arXiv:2203.11891v3, [cs.NE], 12 Apr 2022.

[5] N. Wiener. *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, 2nd revised ed., 1961.

[6] V. Albu. *The Plirophoria: The Missing Puzzle of the Ultimate Picture of the Universal Information*, Proceedings of the Workshop on Intelligent Information Systems WIIS2021, October 14-15, 2021, Chisinau, Republic of Moldova pp. 26–33.

[7] Yuchen Wang, Zixuan Hu, Barry C. Sanders, and Sabre Kais. *Qudits and high-dimensional quantum computing*, arXiv:2008.00959v4 [quant-ph], 11 Nov 2020.

[8] Y. Chi, J. Huang, Z. Zhang et al. *A programmable qudit-based quantum processor*. Nature Communications, 13, 1166 (2022). https://doi.org/10.1038/s41467-022-28767-x.

[9] E. Pournaras, S. Yadhunathan, and A. Diaconescu. *Holarchic structures for decentralized deep learning: a performance analysis*, Cluster Computing (2020) 23, pp. 219–240, https://doi.org/10.1007/s10586-019-02906-4.

[10] Piero Mella. *The holonic revolution: holons, holarchies and holonic networks: the ghost in the production machine*. Pavia University Press, 2009.

[11] Jose Gaite. *The Fractal Geometry of the Cosmic Web and its Formation*, Published in: Adv.Astron. 2019, (2019) 6587138.

[12] R. Landauer. *Irreversibility and Heat Generation in the Computing Process*, IBM Journal of Research and Development, Volume 5, Issue 3, July 1961 pp 183–191, https://doi.org/10.1147/rd.53.0183.

[13] L.L. Yan, T.P. Xiong, K. Rehan, F. Zhou, D.F. Liang, L. Chen, J.Q. Zhang, W.L. Yang, Z.H. Ma, and M. Feng. *Single-Atom Demonstration of the Quantum Landauer Principle*, Phys. Rev. Lett. 120, 210601 (2018).

[14] A. Einstein. *Does the inertia of a body depend upon its energy content?* Annalen der Physik, 18, 1905, pp. 639–641 (https://www.astro.puc.cl/~rparra/tools/PAPERS/e_mc2.pdf).

[15] V. Albu. *A Road Map for an Informational Ontology of Information/Energy/Matter' Homeostatic Unity Inspired by Information Catastrophe*. Proceedings of the Workshop on Intelligent Information Systems WIIS2020, December 04-05, 2020, Chisinau, Republic of Moldova, pp. 15–24.

[16] R. Colbeck and R. Renner. *Is a System's Wave Function in One-to-One Correspondence with Its Elements of Reality?* Phys. Rev. Lett., 108, 150402, DOI: https://doi.org/10.1103/PhysRevLett.108.150402, Published 11 April 2012.

[17] J. Batle, A. R. Plastino, M. Casas, and A. Plastino. *Understanding Quantum Entanglement: Qubits, Rebits and the Quaternionic Approach, Optics and Spectroscopy*. Vol. 94, No. 5, 2003, pp. 700–705.

[18] R. Allahverdi, M. A. Amin et al.. *The First Three Seconds: a Review of Possible Expansion Histories of the Early Universe*. FERMILAB-PUB-20-242-A, arXiv:2006.16182v2 [astro-ph.CO], 27 Jan 2021.

[19] Peter Anninos. *Computational Cosmology: From the Early Universe to the Large Scale Structure*. Published in *Living Reviews in Relativity by the Max Planck Institute for Gravitational Physics & Albert Einstein Institute, Germany, 2001.*

[20] Jay M. Pasachov. *Big Bang Theory*. Hopkins Observatory, Williams College, Williamstown, Massachusetts, https://www.accessscience.com/content/big-bang-theory/080900, DOI:https://doi.org/10.1036/1097-8542.080900, 2018.

[21] John G. Cramer. *The transactional interpretation of quantum mechanics*. Rev. Mod. Phys. 58, 647, published 1 July 1986, DOI: https://doi.org/10.1103/RevModPhys.58.647 .

[22] Ruth E. Kastner. *The Transactional Interpretation of Quantum Mechanics: The Reality of Possibility*. Cambridge University Press, 2022.

[23] Stephen Wolfram. *A Class of Models with the Potential to Represent Fundamental Physics*. Complex Systems, 29(2), 2020, pp. 107–536.

Veaceslav Albu
Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: vaalbu@gmail.com

Mooding - Emotion Detection and Recommendation System

Paula-Luiza Balint, Adrian Iftene

Abstract

In our daily life, we face all kinds of situations and problems, which directly affect us and alter our moods. One of the most common things that have become almost completely integrated into our lives without us even realizing it, is music. On the street, in the car, at a restaurant, or at home, it is present in our lives. Music has become for many people a method of therapy, a way to temporarily escape from the space in which we are captive, and a way to block out any outside sound in order to find our inner peace and be able to have a few moments of introspection, undisturbed by events external to us. Music listeners often find themselves in a situation where they open a music app and receive a list of recommendations that didn't match their mood. Thus, we considered that an application that at least tries to provide recommendations that consider this aspect could be helpful. For this, we built a music application that has a component to detect the emotion experienced by the user in order to offer suggestions as personalized as possible.

Keywords: face recognition, sentiment identification, recommendation system, microservices.

1. Introduction

Throughout time, people have always strived for the better, they have constantly tried to improve their condition, to ease and automate their everyday life. This need for improvement has resulted in the building of tools and functionalities that, although they have become part of our daily lives, being indispensable, a few decades ago were just a dream, an idea. The field of informatics has grown in recent times, so we have to adapt

and evolve along with it, especially in the context where access to information is becoming more and more accessible.

This paper presents the development of a WEB application that uses a song recommendation system based on the user's listening history. This is done in accordance with the emotion experienced by him at the time, thus integrating a real-time recording and detection component of emotions. In making the component responsible for emotion detection, we performed a series of experiments on the constructed neural network, adjusting the hyperparameters, to be able to obtain and opt for the variant that offers the best results, the one with the best accuracy, similar to [1, 2, 3].

In the case of the WEB application, we designed the application in the architectural style based on microservices, having within it four microservices and an API Gateway responsible for managing them. In three of the four microservices, we used the Onion architectural style to ensure a good organization of the code and the links that are established between the functionalities found within it.

To make improvements, we made a form containing an app interaction test, which we gave to the participants. Following their responses and feedback, we were able to identify both the predominantly positive aspects of the app and possible features that can be added to improve the user experience.

2. Music Recommendation Systems

2.1 Mood Beats - Music Player

Mood Beats[1] is an Android application that aims to help the user organize the songs available on their mobile device into categories corresponding to the moods they represent. The application provides a number of predefined states: *fear, wonder, anger, anxiety, bored, calm, disgusted, distracted, happy, angry, and satisfied*. Audio sources present on the user's device are categorized as they are analyzed. Within this application, the emotions experienced by the user are not recorded, the recommendations are made based on existing audio sources. The audio sources classified in this way are then distributed in playlists that

[1] https://play.google.com/store/apps/details?id=com.dreamsin.fl.moodbeatsmp

correspond to the state they represent. Thus, the recommendations are not personalized and are not provided in real-time, based on the fact that the user already has audio sources on his device that can be distributed in the respective categories, which can later be accessed by the user.

2.2 Youtube Music

Youtube Music[2] is a music application owned by YouTube, Google, and offers various functionalities in terms of actions that the user can perform within the application. Among the app's features is exploration, which provides a section where playlists are suggested based on moods and moments. These statuses are predefined and include generic playlists that contain songs that fit into that category.

3. Proposed Solution

3.1 The Viewer's Emotion Recognition Component

Recognition of Facial Expressions

Emotions contribute to establishing a good relationship between individuals, being one of the most important aspects of interpersonal relationships. Sometimes, perceiving and understanding them can be a complex and difficult process. These can be identified by a number of non-verbal cues such as tone of voice, body movement, posture, changes in heart rate, hyperventilation, or increased sweat gland activity. But one of the simplest methods to identify emotions is to recognize facial expressions.

This is a subject of interest in the field of informatics as it has various uses, such as: in the field of marketing, for various measurements and statistics carried out for products or brands; in the field of video games, especially in the testing stage, where the player's reactions represent an important part of their assessment. To date, seven types of emotions have been identified, of which six are basic: *happiness, sadness, wonder, anger, disgust, and fear/fear*, and a seventh type that was added later: *neutral*. Recognizing emotions based on images is, as mentioned above, a more difficult process, due to the fact that there are not enough diverse

[2] https://music.youtube.com/

images to train the model, and the conditions in which they are located influence it: *lighting conditions*, *rotations*, *states of transition*, *clarity*.

Dataset

The data set used in the experiments was FER2013[3], being the most used data set in the context of emotion recognition, because it is the most varied and difficult, the best models developed in the competitions on this data set have an accuracy that falls between 70% and 75%. The data set is divided into two parts: data for training and data for testing. Test data is also divided into two categories: *private* and *public*, respectively, those for validation and testing. The validation set is used in the training process, to obtain the validation error metric, respectively, validation accuracy, and the test set, used after training the model, to test its generalization capacity. Figure 1 shows the distribution of the entire data set over the seven emotion classes.

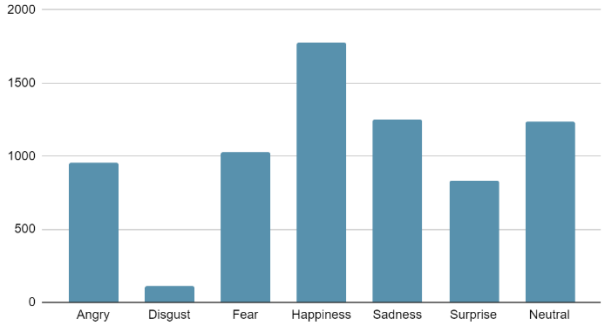


Figure 1. Distribution of the FER2013 dataset over the seven emotion classes

The reason this data set is difficult is that, in addition to being the most diverse, which is positive, it is not balanced: the seven classes are not balanced, which can lead to poor generalization of the model, it not having enough data to clearly identify each type of emotion separately. Thus, the model may find it harder to detect emotions in classes for which there is insufficient data.

Model Development

[3] https://www.kaggle.com/datasets/msambare/fer2013

The model was made by means of a convolutional neural network, which is based on deep learning, this being supported by the studies and results obtained by such networks in the field of similar problems, in the field of Computer Vision, working on large data sets. In making it, we carried out various experiments to make comparisons between different configurations, after which we could opt for the model that gave us the best results.

The network consists of seven components, the first is represented by a *spatial transformer* [4], the next two are responsible for the convolution operation, the next for the *Dropout* operation, followed by two fully connected layers, followed by the *Softmax* function.

The first component, which in turn consists of two convolutional layers followed by a fully connected layer, allows the network to learn how to perform spatial transformations on the input data to increase the degree of geometric invariance of the model. For example, it can crop a region of interest, scale, and correct the orientation of an image, thus helping the neural network as they are not invariant to affine transformations (e.g. *rotation*, *scaling*, etc.) [5]. Flattening is also performed between the two convolutional layers and the fully connected layer. The next two components each consist of two convolutional layers, followed by Max Pooling and activation. Within the third component, between the two layers, normalization is also performed.

To start the series of experiments, it was necessary to establish one of the most important hyperparameters in the development of the model: the learning rate. For this, we considered two values: 0.001 and 0.005, which are usually used in networks of this type. To compare model performance across the two, we used the ReLU activation function and batch size 32, which is a common starting point in model development. With these details established, we trained the model for a number of epochs equal to 10 and obtained the results in Table 1.

Table 1. Comparisons of results obtained from training the network for learning rate, on dataset size = 32 and number of epochs = 10 in the case of the ReLU function

Learning rate	Accuracy	Error
0.001	50.8725%	~1.2792
0.005	43.8678%	~1.4487

We can see that the model performed much better on the learning rate = 0.001, so further experiments will be with this hyperparameter set to this value. Next, to establish the batch size, we trained the model on a number of epochs = 10 and the previously established learning rate and obtained the results shown in Table 2.

Table 2. Comparisons of results obtained after training the network for batch size, on learning rate=0.001 and number of epochs=10 in the case of the ReLU activation function

Batch size	Accuracy	Error
8	50.5591%	~1.2788
16	51.6702%	~1.2553
32	50.8725%	~1.2791

Once the values of the parameters learning rate = 0.001, batch size = 16, and the number of epochs = 10 were set, we performed a series of experiments to determine which activation function is more suitable. Considering the fact that the most used and indicated activation function for this type of network is ReLU, we sought to perform experiments on similar functions, or that propose improvements, to avoid the problem of the ReLU function, namely "dying relu". This problem refers to the fact that if the value of a neuron becomes negative, it will be canceled, never being able to recover.

Haar-Cascade Classifier

In the emotion recognition process, one of the key steps is to identify the person's face. For this, we used a Haar-Cascade type classifier, which is part of the OpenCV library. Each image that wants to be processed is passed through a series of operations that include applying a black and white filter over them and detecting the faces in the image using the *detectMultiScale* function from the OpenCV library. Following these operations, the program will have at its disposal, in case of success, the coordinates corresponding to the area of interest, which are then resized and offered to the model for prediction.

3.2 Recommendation System

Technologies that were used within the application, for the backend part are as follows: (1) Spring Boot 2.4.1, (2) Flask 2.1.2, (3) Netflix Eureka,

(4) JWT, (5) Selenium, (6) Deezer API, (7) PyTorch; and for the frontend: (1) React JS.

On the *backend side*, the application components, which are represented by microservices, are Spring Boot-type projects, suitable for complex REST-type applications. A REST API is an architectural style for a programming interface that uses HTTP requests to access and use data. These types of HTTP requests can be GET (get or read a resource), POST (create a resource), PUT (update a resource) and DELETE (delete a resource). These requests are processed by the server, and, following the call to the API that provides access to the resources, the client (most often a web application), receives as a response, the data, in a standard format. Most of the time, the format used in applications for data exchange is JSON.

The *client*, which in this case is represented by the React JS web application, sends requests to the server through HTTP verbs, and the server, as a result, will process the received request and provide an appropriate response.

3.3 Application Architecture

The song recommendation application is built on the model of a system based on microservices, which are managed by an API Gateway [6]. This way of structuring architecture on microservices is a way to develop an application as a collection of autonomous services that communicate with each other through various mechanisms. Each microservice is built to provide a single functionality, and thus, can be delivered independently, and reused within other applications that would need such a component. Thus, this architectural style encourages code reuse. The microservices each have their own database, excluding the emotion detection microservice, which does not require one, and can be developed using different technologies. In this way, the functionalities made available by a certain platform or technology can be exploited, depending on the needs.

The application consists of four microservices:

1. *User management microservice* is responsible for creating a new account for a user and authenticating it. In its database, information related to the user entity is stored, whose properties are: *id, email, user_role* (predefined as a generic one), *first_name, last_name, password, and username*;

2. *The songs microservice* is responsible for songs and information about them, such as artists and their characteristic albums and songs; it has a database schema consisting of several tables that store information characteristic of each entity it maps. The song-specific table contains the following characteristic properties: *id, title, the duration of the song in seconds, and the URL to its audio source*. The tables specific to artists and respectively albums, each contain two properties, namely the name and the photo, respectively the title, and the album image. The source from which all this data originates and is taken is the API made available by Deezer, used in accordance with a top 1000 artists taken from a website that represents a rating tool in the music industry, which offers rankings based on certain criteria[4]. The top 1000 artists taken over is created by them with the help of a ranking taken from Spotify;
3. *Recommendation microservice* is responsible for providing song recommendations to users who have listened to at least one song by an artist within the application. This is used to save future recommendations and get recommendations based on an identified emotion. Recommendations are made based on emotion and are based on what artists the user is listening to when they are feeling a certain way. These artists are uniquely identified by an *id*, as well as the *users* and the *emotions* they can experience. So, this microservice has the database schema consisting of a single table, of recommendations, which has as its unique property and alone, the primary key which is composed of 3 elements: *artist_id, mood, and user_id*. In this way, we can uniquely identify a recommendation to give to a user;
4. *The emotion detection microservice* is responsible for providing the emotion detected based on an incoming photo. It uses the developed model. This microservice is of Flask API type and exposes an endpoint to the outside, so that it can be called from within the WEB client, through the API Gateway. To get the emotion based on a photo, the "/emotion-recognition/detect"

[4] https://chartmasters.org/2020/05/most-streamed-artists-ever-on-spotify/

endpoint can be called, which has in the body the photo for which the detection is desired, in *base64* format.

Netflix Eureka, a technology developed on the Java platform by Netflix, was used to manage microservices. They communicate with each other and are orchestrated by the API Gateway, which is the intermediate layer between the client and each microservice.

3.4 The Visual Component of the Application

In creating the visual component of the application, i.e. the front-end part, the React JS framework was used as a technology, which provides various functionalities and libraries to facilitate the process of creating and maintaining this component. When a new, unauthenticated user accesses the application's web page, he is greeted by a login form where he must enter his credentials to log in.

After successful authentication, the user is redirected to the main song page, where they can search for a song by artist, title, or album, and receive the results in real-time. Here he can play a song, adjust the volume, and pause it (see Figure 2). When the user taps on a song, the app starts the process of capturing the emotion with a camera that is on the page but not visible and sends a request to the recommendation endpoint to save a future recommendation based on the artist name of the song listened to and of the emotion identified.

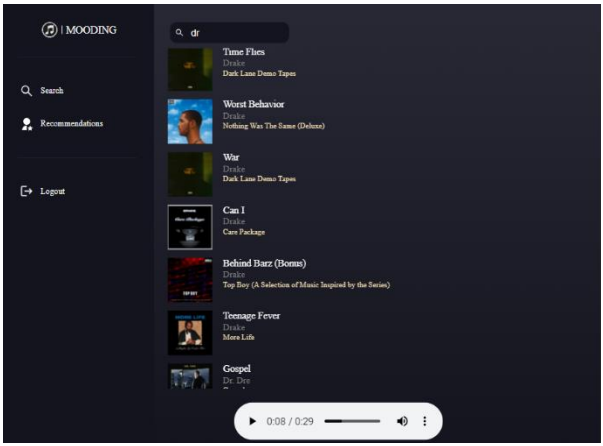


Figure 2. Song search page, where user emotions are analyzed

If the user accesses the recommendations page, a button is provided to load the recommendations based on the emotion experienced by him at that moment. If there is no data yet to provide recommendations, a message is displayed to inform them of this fact. If, however, there are records for the detected emotion, it will receive as a result a list of songs of the artists that it listens to when it has the identified state. The user has the option to navigate between pages to search for songs and receive new recommendations. He has the option to disconnect from the application, using the "Logout" option from the left menu.

5. Conclusion

The presented work aims to provide personalized recommendations to the user, thus facilitating their search process. When the user is using the app, if he doesn't want to search for songs, he can use the recommendation component that will give him songs that fold on the status he has. The architecture of the backend component is based on microservices.

The presented application has the potential to be further developed, by adding new functionalities, but also by improving existing functionalities. Some examples would be the addition of options for creating playlists and favorites, diversifying the playback options of songs, and the ability to access lyrics and details for each of them. Another possible interesting functionality would be the option to create shared playlists so that users who have the same musical tendencies can discover new songs similar to the ones they already listen to.

References

[1] L. G. Coca, C.G. Cușmuliuc, and A. Iftene. *CheckThat! 2019 UAICS*. Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, (2019).

[2] C.C. Ciubotariu, M.V. Hrișca, M. Gliga, D. Darabană, D. Trandabăț, and A. Iftene. *Minions at SemEval-2016 Task 4: or how to build a sentiment analyzer using off-the-shelf resources*? Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics. San Diego, California, June, pp. 247-250, (2016).

[3] C. Florean, O. Bejenaru, E. Apostol, O. Ciobanu, A. Iftene, and D. Trandabăț. *SentimentalITsts at SemEval-2016 Task 4: building a Twitter sentiment analyzer in your backyard*. Proceedings of the 10th International

Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics. San Diego, California. June, pp. 243-246, (2016).

[4] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. *Spatial Transformer Networks*. Arxiv Computer Science, Computer Vision and Pattern Recognition, arXiv:1506.02025, (2016).

[5] G. Hamrouni. *Spatial Transformer Networks Tutorial*. PyTorch, Tutorials, (2021).

[6] M. Baboi, A. Iftene, and D. Gîfu. *Dynamic Microservices to Create Scalable and Fault Tolerance Architecture*. 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. Procedia Computer Science, 4-6 September, Budapest, Hungary, vol. 159, pp. 1035-1044, (2019).

Paula-Luiza Balint, Adrian Iftene

"Alexandru Ioan Cuza" University of Iasi,
Faculty of Computer Science, Romania
E-mail: paula.balint@info.uaic.ro, adiftene@info.uaic.ro

Distributed computing infrastructure for complex applications development

Petru Bogatencov, Grigore Secrieru, Radu Buzatu, Nichita Degteariov

Abstract

Implementation and use of heterogeneous Multi-zone Cloud infrastructure that integrates various types of computing resources are described. It is shown the necessity of the development of computer infrastructures and services that are focused on supporting Open Science initiatives and offering conditions for solving complex problems with high demands of computing resources. Approaches to the deployment of complex cloud infrastructures, their configuration, and administration are presented. The described computing infrastructure has an important role for the research community of Moldova in using the performances of the creating European Open Science Cloud resources and services.

Keywords: cloud computing, e-infrastructure & services, Open Science support tools.

1. Introduction

The importance of the development of computing infrastructure and services for support of open research data accumulation, storage, and processing is permanently increasing. These e-Infrastructures became more and more universal and provide various types of services for operation with research data, including high-performance resources for complex data processing applications development and porting.

Work on the implementation of distributed computing infrastructure in Moldova started in 2007 when the first Agreement on the creation of the MD-GIRD Joint Research Unit Consortium and accompanying Memorandums of Understanding were signed by seven universities and research institutes of Moldova. Since this time, the works started on the

deployment of the first national distributed computing infrastructure that included integration of computing clusters and servers installed in the Technical University, State University, State University of Medicine and Pharmaceutics, Vladimir Andrunachievich Institute of Mathematics and Computer Science, RENAM, and State Hydrometeorological Service.

Since 2013 in the Vladimir Andrunachievich Institute of Mathematics and Computer Science (VA IMCS), the works started on the deployment of virtualized cloud-based computing infrastructure and transferring of the existing in Moldova distributed computing infrastructure to a distributed cloud environment. Participation of VA IMCS and RENAM specialists in the regional project "Experimental Deployment of an Integrated Grid and Cloud-Enabled Environment in BSEC Countries on the Base of gEclipse (BSEC gEclipseGrid)" supported by the Black Sea Economic Cooperation Programme (http://www.blacksea-Cloud.net) in 2014-2015 had a significant impact on support of these works [1]. In continuation (2014-2022), deployment of the national cloud computing infrastructure was supported by the bilateral project "Instrumental support for complex applications porting to the regional HPC infrastructure" funded by the Science and Technology Center in Ukraine and the Academy of Sciences of Moldova [2] and by EU funded projects: Eastern Partnership Connect (EaPConnect), EU4Digital - Connecting Research and Education Communities (EaPConnect2). These projects significantly contributed to the procurement of new computing equipment and cloud computing infrastructure extension. Now, these works are supported by the national project "Investigation and elaboration of the integrated infrastructure of the unified environment "cloud computing" to support open science" funded by the National Agency for Research and Development [3].

2. Approaches for the Distributed Computing Infrastructure Deployment

In the first stage, it is planned to deploy a multi-zone IaaS Cloud infrastructure that combines the resources of VA IMCS, the State University of Moldova (SUM), and RENAM into distributed computing network for processing scientific data, performing intensive scientific calculations, as well as storing and archiving research data and results of computational experiments.

Works on a new Scientific multi-zone IaaS Cloud Infrastructure that is based on OpenStack Ussuri have begun in 2021 and are progressing now. As a result, today in VA IMCS and RENAM, in parallel are available previously deployed and operating resources via cloud.renam.md, which runs on an outdated version of OpenStack Mitaka and a new Cloud platform, based on the latest OpenStack Ussuri version, offering more features, more processing power, and flexibility of operation. During the pilot cloud infrastructure testing and subsequent work on the outdated version of OpenStack Mitaka, many bottlenecks were identified and several ideas were proposed for performance enhancements. There are several improvements to the new platform accessible via openstack.math.md comparing with the old one. In the new platform, now 4 servers are used instead of 1 used previously as a host system with the following parameters: 2 servers with 24 vcpu and 2 servers with 16 vcpu; 48 Gb RAM on each server and 1 TB HDD space for creating virtual machines.

A new important component has been added - block storage, which allows the creation of volumes for organizing persistent storage. In general, in OpenStack, as in other modern Cloud systems, several concepts exist for providing storage resources. When creating a virtual machine, you can choose a predefined flavor, with a predefined number of CPU, RAM, and HDD space; but previously, when you delete a virtual machine, all data stored on the machine instantly disappears. The new storage component, used in the created multi-zone IaaS Cloud Infrastructure, is deployed on a separate storage server and allows you to create block storage devices and mount them on a virtual machine through special drivers over the network. This is a kind of network flash drive that can be mounted to any virtual machine associated with the project, unmounted and remounted to another, etc., and most importantly, this type of volume is persistent storage that can be reused when the virtual machines are deleted. Thus, you can no longer worry about data safety and easily move data from one virtual machine to another, or quickly scale up VM performance by creating a virtual machine with larger resources and simply mount volumes to it with all scientific data available for further processing.

Now, for guest systems, two separate subnets with 32 IP addresses each have been created, as opposed to the one subnet with 16 available IP addresses in the previous version.

A more advanced and flexible model of interaction with the network has been implemented. In the new cloud infrastructure (openstack.math.md), in addition to the usual "provider network" model, which allocates one real IP address from the pool of provider network addresses to each virtual machine, a self-service network is also available. A self-service network allows each project to create its own local network with Internet access via NAT (Network Address Translation). For a Self-service network, the user creates a virtual router for the project with its own address space for the local network. Virtual eXtensible Local Area Network (VXLAN) traffic tagging is used to create such overlay networks that prevent the occurrence of address conflicts between projects in case several projects will use network addresses from the same range. To ensure the functioning of NAT, one IP address from the provider network is allocated to the external interface of the virtual router, which serves as a gateway for virtual machines within the project. Also, when using the self-service model, the floating IP technology becomes available, which allows you to temporarily bind the IP address from the provider network to any of the virtual machines in the project, and at any time detach it and reassign it to any other virtual machine of the project. Moreover, the replacement occurs seamlessly, that is, the address does not change inside the machine, but remains the same - the address is from the internal network of the project, but the changes occur at the level of the virtual router. Incoming to the external address packets are forwarded by the virtual router to the internal interface of the selected virtual machine. This allows you efficiently to use IP addresses and not allocate an external address to each virtual machine. The external IP address remains assigned to the project and can be reused by other machines within the project.

For the deployment of new computing infrastructure, the process of transition to a 10G network has started according to the elaborated plan. The New Juniper switch already has been installed and all storage servers with 10G cards on board have been connected to this switch. We have a plan to switch all remaining servers to 10G interfaces this year. To increase the bandwidth and improve the reliability of the existing 1G

network, Linux bond technology has been applied to the existing network, which allows aggregating two or more network interfaces into one logical device by selecting one of seven possible modes of operation [4]. We use the balance-rr mode, which balances traffic by distributing network packets sequentially from the first interface to the last. This allows getting a twofold or more increase in throughput when combining two or more network interfaces into a bond. We use this technology by combining two 1G interfaces into one 2G interface for connecting compute and storage nodes for faster and more stable operation of a guest OS with persistent volumes (see Fig. 1). As you can see, data is transferred between the two servers with the speed 1.94 Gbps for sending and 1.93 Gbps for receiving.

This temporary solution will remain operational until the existing 1G network infrastructure will be completely switched to 10G.

```
root@pve002:~# iperf3 -c 192.168.0.125
Connecting to host 192.168.0.125, port 5201
[  5] local 192.168.0.127 port 48862 connected to 192.168.0.125 port 5201
[ ID] Interval           Transfer     Bitrate         Retr  Cwnd
[  5]   0.00-1.00   sec   226 MBytes  1.90 Gbits/sec  278    760 KBytes
[  5]   1.00-2.00   sec   234 MBytes  1.97 Gbits/sec  315    690 KBytes
[  5]   2.00-3.00   sec   227 MBytes  1.90 Gbits/sec  640    638 KBytes
[  5]   3.00-4.00   sec   234 MBytes  1.97 Gbits/sec  158    725 KBytes
[  5]   4.00-5.00   sec   224 MBytes  1.88 Gbits/sec  635    166 KBytes
[  5]   5.00-6.00   sec   230 MBytes  1.93 Gbits/sec  251    743 KBytes
[  5]   6.00-7.00   sec   230 MBytes  1.93 Gbits/sec  398    664 KBytes
[  5]   7.00-8.00   sec   236 MBytes  1.98 Gbits/sec   11    909 KBytes
[  5]   8.00-9.00   sec   235 MBytes  1.97 Gbits/sec   19    979 KBytes
[  5]   9.00-10.00  sec   231 MBytes  1.94 Gbits/sec  528    839 KBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bitrate         Retr
[  5]   0.00-10.00  sec  2.25 GBytes  1.94 Gbits/sec  3233             sender
[  5]   0.00-10.00  sec  2.25 GBytes  1.93 Gbits/sec                  receiver

iperf Done.
```

Figure 1. Speed measurement on the bond interface

Another significant improvement is the ability to configure guest machines using cloud-config as shown in Fig. 2. Various parameters and commands can be passed to the guest virtual machine at the boot stage to configure it, which allows you to fully automate OS pre-configuration, package installation, starting services, etc.

In this example, we illustrate creating a virtual machine with pre-configuration parameters passed to the VM that, for example, allows creating new user, defining a password for the user, appending a ssh key for passwordless access, and installing new packages and updates. The last command sends an email with the subject "Your Openstack VM is

ready" and the body text "Now you can ssh to it, cheers!" to the predefined user mail address.

```
      #cloud-config
      users:
        - name: nikita
          groups: [adm, audio, cdrom, dialout, floppy, video, plugdev, dip, netdev, sudo]
          passwd:
$6$1rj3YMu0hbgxIR$Sm4QjQdN0jYFcD/HcCvP9k1KCZsJ3eiTPEJ6aF7ZoTWAtaG6apsQNz0BT3afhlUDZQDL0Uj.fi0ySBKYn
aGPS1
          ssh_authorized_keys:
            -                                                                                "ssh-rsa
AAAAB3NzaC1yc2EAAAADAQABAAABAQCj8YYTA+pcq7rZzQd7r8ClOqbkGHABkbAwBrNy2QG+BFwMStp9dg8Ynf9xlJVdqwh8KAX
9kCiJPxSFFH97HqCjfjET1kOBpTI99Bp2R0NfmIH2NALKJgTzQo4mWFLd0AgO82M0vmANQXpF2slRfbPjqGWkJQQRzwm0/YiKAg
kzQU/+Es98i03g9JWyvQXoToqt9NZoLGMLiwh/savlEl163YYf9T+vRmzo2bmHARh5bkGa8RE0Gf6PhK3Z1HdFcOTUdVvtb2Hhx
1XSKFIvj7S7/PqIexjBaU1WWIR59SCuzVnlYQzR+XEh8xEXaiFegBj3Wud9ZFA5t6wAfjOt30O3 nikita@cloud"
      package_upgrade: true
      packages:
        - fail2ban
        - sendmail
        - mailutils
      runcmd:
        - systemctl restart sendmail
        - systemctl restart fail2ban

        -  echo  "Now  you  can  ssh  to  it,  cheers!"  |  mail  -s  "Your  Openstack  VM  is  ready"

nichita.degteariov@math.md
```

Figure 2. Configuration example using cloud-config.

Created robust, performant, and secure computing infrastructure has become a platform for expansion and improvement of services provided for different research teams.

3. Use of the Computing Infrastructure for Complex Applications and Algorithms Realization

Distributed computing infrastructure supports the adaptive execution framework that can be configurated and tested for the solution of different complex applications. The research teams from VA IMCS and SUM developed and continue developing several applications that require resources of multiprocessor clusters and distributed computing infrastructure. One practical example of the application for solving complex decision-making problems elaborated for porting on the created computing infrastructure was described in [3]. Other examples of complex applications and services deployment by using resources of the created computing infrastructure are presented below.

3.1 Optimal Partition of the State into Economic Territorial Units

The main algorithm for determining the optimal territorial partition of the state into coherent economic territorial units (ETUs) was proposed and realized [5]. The algorithm allows solving the problem that involves merging a set of localities having strong infrastructure communication, which comprises telephone lines, gas pipes, road, power, and water systems. So, the resulting ETUs have to meet multiple criteria, which need to be balanced. The criterion, which is hardest to satisfy is the territorial contiguity of an ETU; it essentially means that it is possible to travel by roads between any two localities within the ETU without having to visit locality from other ETUs.

The solution process of the described problem consists of the following three steps:

1) Elaboration of two appropriate integer linear programming models (ILP models), which can be easily adjusted to special restrictions and criteria;
2) Determination, by using the first ILP model, the optimal number of ETUs needed to partition the state, for which the imposed restrictions are satisfied;
3) Determination, by using the second ILP model, the most balanced partition into the optimal number of ETUs established at the previous step.

The contiguity of ETUs in ILP models, used in steps 2) and 3), is expressed by the shortest roads (paths) between the ETUs centers and any other locality assigned to these centers. This formulation substantially reduces the number of variables and constraints used in the ILP models and, as a result, a balance between the quality of the solution and the computational effort is achieved. We consider that the maximum allowable travel distance D from the center of an ETU to any commune assigned to it is the main restriction of the elaborated IPL models. Depending on the choice of D, the algorithm will determine the optimal partition.

The developed algorithms, essentially consisting of steps 2) and 3), will be implemented and tested on the created distributed computing infrastructure that includes clusters with multicore processors by means of Python programming language and CPLEX Optimization Studio.

3.2 DICOM Network - Distributed System for Medical Images Pre-processing and Archiving

The "DICOM Network" project was launched in Moldova in 2012, whose goal is to provide access to the collected imagistic data for medical staff with the appropriate access rights and for patients - to the personal radiography investigations. Today the system implemented in many hospitals in Moldova, collects and processes more than 5TB of data per month gathered from different types of medical equipment [6].

"DICOM Network" realization based on the national scientific cloud platform opens many possibilities for using this application for various types of activities. DICOM investigations could be added to some other datasets, that are collected and available from cloud infrastructure. But additional functionalities require the realization of supplementary solutions for Imagistic Data anonymizing [7]. Cloud technologies and services allow optimization and making collected medical imagistic data compatible with Open Science principles; they become widely accessible by means of mobile devices.

The diagram below (see Fig. 3) shows different data processing options that make it possible to store collected data in cloud storage.

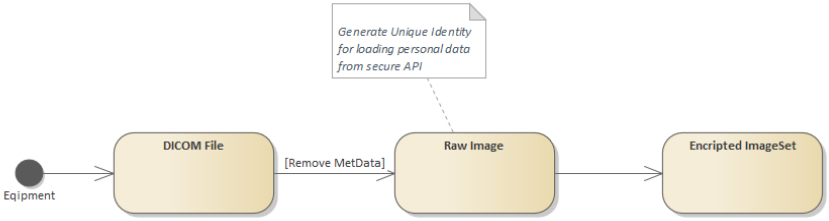


Figure 3. Data processing options for storing data in the cloud storage.

Initially, all data is collected in DICOM (.dcm) format that contains raw images and XML data with personal information. The proposed data format optimizations could be divided into two steps:

The first is the removal of XML with personal data from the image file. The extracted data will be sent as metadata and stored under a unique identifier for this image set study. Anonymized this way data can be stored in the local network of the data owner.

The second is encrypting the full image set, that will completely eliminate the possibility of restoring the personal data based on the image. This option will make it possible to store encrypted images using external cloud storage facilities.

This will require additional image processing for data storage, data accessing, and data visualization, but it makes all personal data protected. At the same time, this approach makes it possible to use various cloud-based APIs and services for data processing and exchange.

3.3 Integrated system for distant learning and video-conferences support

Since 2019, the created Cloud infrastructure started to use for a new service – support of online lectures organization for the State University and the Technical University of Moldova. These realizations allowed to intensify the use of distributed computing resources. One of the representative examples is launched in 2020 and actively used - a multi-node distributed video-conferencing system, that provided facilities for the organization of online classes since the beginning of the lockdowns, caused by the COVID-19 pandemic back in 2020. The video-conferencing system is powered by the open-source project BigBlueButton [8]. The system is integrated with Moodle, creating a self-sufficient distant e-learning platform and it is actively used for distant learning by the main universities of Moldova: the Moldova State University, the Academy of Economic Studies of Moldova, as well as by some smaller institutions in Chisinau and the regions (e.g., in Comrat and Taraclia cities). It hosts roughly 1 – 1,2k concurrent users daily with peaks up to nearly 2k users in about 60 separate virtual rooms distributed among the servers' cluster. As an example, the statistics for the first half of September 2021 are presented in Fig. 4.

The effective use of the VC system has been achieved by uniting distributed BBB nodes in a cluster using the Scalelite project. Scalelite is an open-source load balancer that manages a pool of BBB servers. It makes the pool of servers appear as a single (very scalable) BigBlueButton server. A front-end, such as Moodle or Greenlight platforms, sends standard BBB API requests to the Scalelite server which, in turn, distributes those requests to the least loaded BigBlueButton server in the pool. We also use Greenlight as a meeting managing plugin and a

pool of three Traversal Using Relay NAT (TURN) servers for relaying the traffic between peers behind the NAT.

Figure 4. Number of VC system users

5. Conclusion

With the support of several national and international projects, the distributed computing infrastructure in Moldova was created, jointly operating by interested institutions, and permanently upgraded by the installation of new computing equipment and new versions of middleware. The created computing infrastructure ensures reliable operation and wide access to its resources, deployment of various tools and platforms for support of Open Science that corresponds to the current needs of researchers in Moldova. Deployment of modern architectural solutions, tools, and platforms, and installation of new high-performance servers allowed intensifying the use of the distributed computing resources and contributed to the qualitative development of IT services for R&E in Moldova.

Acknowledgments. Grants from the National Agency for Science and Development (grant No. 20.80009.5007.22 and grant No. 21.70105.9ŞD) and EU funded "EU4Digital: Connecting Research and Education Communities (EaPConnect2)" project (grant contract ENI/2019/407-452) have supported part of the research for this paper.

References

[1] Nichita Degteariov, Petru Bogatencov, Nicolai Iliuha, and Grigore Secrieru. *Upgrading Cloud Infrastructure for Research Activities Support*. Workshop on Intelligent Information Systems (WIIS2020) Proceedings, Chisinau, IMI, 2020, pp. 69-74.

[2] P. Bogatencov, N. Iliuha, E. Calmis, B. Hancu, V. Patiuc, and G. Secrieru. *Computational Infrastructure for Porting and Execution Complex Applications in Moldova.* Proceeding of the 5th International Conference "Telecommunications, Electronics and Informatics", May 20 – 23, 2015, Chisinau, UTM, 2015, pp. 21-26.

[3] Petru Bogatencov, Grigore Secrieru, Boris Hîncu, and Nichita Degteariov. *Development of computing infrastructure for support of Open Science in Moldova.* Workshop on Intelligent Information Systems (WIIS2021) Proceedings, Chisinau, IMI, 2020, pp. 34-45.

[4] *Link aggregation*, https://en.wikipedia.org/wiki/Link_aggregation.

[5] Radu Buzatu and Sergiu Cataranciuc. *On Nontrivial Covers and Partitions of Graphs by Convex Sets.* Comput. Sci. J. Moldova 26(1), IMI, 2018, pp. 3-14.

[6] Peter Bogatencov, Nicolai Iliuha, Grigore Secrieru, and Alexandr Golubev. *DICOM Network for Medical Imagistic Investigations Storage, Access and Processing.* "Networking in Education and Research", Proceedings of the 11th RoEduNet IEEE International Conference, Sinaia, Romania, January 17-19, 2013, pp. 38-42.

[7] A. Anagnostaki, S. Pavlopoulos, E. Kyriakou, and D. Koutsouris. *A Novel Codification Scheme Based on the VITAL and DICOM Standards for Telemedicine Applications*. IEEE Transactions on Biomedical Engineering, vol. 49, no. 12, pp. 1399–1411, 2002.

[8] Grigore Secrieru, Peter Bogatencov, and Nichita Degteariov. *Development of Effective Access to the Distributed Scientific and Educational e-Infrastructure*. Proceedings of the 9th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2021). Dubna, Russia, 5-9 July 2021; Vol-3041, urn:nbn:de:0074-3041-3, pp. 503-507; DOI:10.54546/MLIT.2021.21.73.001.

Petru Bogatencov[1,2], Grigore Secrieru[1], Radu Buzatu[3], Nichita Degteariov[1,2]

[1]Vladimir Andrunachievici Institute of Mathematics and Computer Science. Chisinau, Moldova.
E-mail: bogatencov@renam.md, secrieru@renam.md, nichita.degteariov@math.md

[2]RENAM Association. Chisinau, Moldova.
E-mail: bogatencov@renam.md, nichita.degteariov@math.md

[3]State University of Moldova. Chisinau, Moldova.
E-mail: radubuzatu@gmail.com

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Monitor city-wide sewage systems using the Internet of Things and eXplainable Artificial Intelligence

Mustapha Bourahla

Abstract

This paper presents a real-time monitoring of sewer systems to prevent overflow using the Internet of Things and eXplainable Artificial Intelligence. The Internet of Things is used to offer continuous data to be used in two steps. In the first step, we use historical data to construct Artificial Intelligence-based prediction models to forecast future system states and then, in the second step, we use the real-time data for monitoring and control of sewer systems using the prediction models combined with eXplainable Artificial Intelligence (XIA) technique.

Keywords: Sewer Systems Control, Internet of Things, Artificial Intelligence, eXplainable Artificial Intelligence.

1 Introduction

The sewer system is designed as an integral aspect of the infrastructure of cities to carry wastewater and storm water away from city centers in order to protect public health [1]. The polluted water should be treated before being discharged into the environment, with the sewage systems and treatment plants focusing not only on public health but also on the environment and resource recovery [2]. If the water flow in some sewage canals is measured to be not at the desired level, there will be problems when heavy rainstorms happened, causing infrastructure damage by causing floods and wastewater overflows due to sewer system

surcharges in certain locations [3]. Sewer overflows are becoming more common as a result of increased urbanization and significant rainfall events, resulting in water pollution and negative environmental, health, and fiscal consequences. Chemical and microbiological contamination, dangerous compounds, and other micro-pollutants in sewer overflows can degrade the water quality [4].

The major goal of this work is to examine the application of real-time control on sewer systems to prevent overflow using the Internet of Things (IoT) and eXplainable Artificial Intelligence (XAI). The Internet of Things [5] is used to offer continuous monitoring and control of wastewater systems in real time. The data acquired by the Internet of Things can also be used to construct and calibrate models that are useful for simulation and prediction. Models that are suited for prediction [6] are frequently used to forecast future system states. They rely on system measurement data and must operate at a high computational speed. The model prediction can be explained using the XAI technique [7, 8] to help diagnosis of the sewer system behavior.

A case study was undertaken for this project in M'Sila, Algeria, to demonstrate how real-time control can be used to manage sewer systems. A hydraulic model was utilized in this study to examine the viability of various control systems. The study looked into the prospect of employing artificial intelligence as a model for prediction, given recent breakthroughs in artificial intelligence and the huge amount of data collected through the Internet of Things.

This paper is organized as follows. In Section 2, we present the method to collect the data using the Internet of Things. Section 3 presents the construction of real-time control models. In Section 4, experimental results are presented to show how we use these real-time control models. Section 5 presents conclusions and perspectives.

2 Collecting data using the Internet of Things

To build a sewer model, we use the Internet of Things (IoT) technology to track process factors like water level and flow rate. Furthermore,

the Internet of Things (IoT) not only connects items, but also generates massive amounts of data from those things. Sensors, wireless communication protocols, actuators, and web interfaces are among the technologies that enable things to interact and collaborate with one another to achieve common goals. A sensing layer, a network layer, and an application layer make up the IoT architecture. Users can access all of the system's features through the application layer, which is located at the top of the architecture. The sensing layer is at the bottom of the IoT architecture, and it uses sensors to track the state of devices. The network layer sits between the sensing and application layers, transmitting data from sensors via wireless communication protocols and linking the physical and digital worlds.

To construct a city-wide IoT, the sensor nodes can be placed across the city and the data can be kept in a centralized storage facility. Sensor nodes collect data on rainfall amount, and other variables of water and wastewater. Using the tool ArcMap of ArcGIS, a spatial database has been created from the sewer map of the city of M'Sila to manage the transferred data. The sewer system has 2,113 pipes, 2,094 manholes, 29 weirs, 18 pumps, and 11 outlets; it is a completely detailed hydraulic model. The sensor network technique is based on using different sensors that send the data in real-time to the database where it is stored. The different sensors in the system could be water-level sensors in waste-water pipes, groundwater sensors, rain-level sensors, and water-level sensors in drainage systems.

The data from the water-level sensors in waste-water pipes could be used to calculate the flow rate; the data from the groundwater sensors could be used to calculate the movement of groundwater; the data collected from rain-level sensors could help track any correlation with rainfall; and data collected from water-level sensors in drainage systems help predict any potential future flooding. In addition to the main table in the database containing information about sensors as their identifiers and locations, each sensor is associated with a table in the spatial database to record its measures sent to server. The structure of all the sensors tables is the same. There is a field to capture the time

stamp of the received measure and another field to capture its value.

3 Building models for real-time control

Models are required for real-time control, which are sets of equations that describe the evolution of state variables in space and time. Models in the domain of sewer systems are data-driven models (derived directly from data). The data collected by the IoT supports in model building to be used for predicting future system conditions. An IoT-based sewage surveillance web portal has been built as part of this project to monitor the water level and velocity of the sewer system's main pipes, as well as rainfall. The information gathered can be utilized to construct artificial intelligence models. The sensing layer comprises of ultrasonic water level sensors, Doppler velocity sensors, and rain gauges. Wireless telemetry is used to send the data collected by the sensors and rain gauges to a remote server.

The goal is to develop an effective and efficient technique to monitor sewage systems to increase their capacity. We had access to all of the installed sensors in the M'Sila region for this investigation, and we were able to acquire information on precipitation, groundwater levels, and sewage water volumetric flow using these sensors. All of the sensors are connected to the database via a wireless network, which can be viewed on the municipality's IoT data portal. Scripts can be used to get both historical and real-time data from the database.

There are 12 sensors (we call them flow sensors) installed within the sewage system to measure how much in liters per hour is the flow of wastewater through the pipes. Each flow sensor is identified by the name fs_i, where $i = 1, \cdots, 12$, and each flow sensor fs_i is associated with a CSV file named $ff_i.csv$, where its content is taken from the geographical database. The groundwater level is adjusted to a zero point and was calibrated to this point by the business that installed the groundwater sensor, which is identified by the name gws and it is associated with a CSV file named $gwf.csv$. The area's precipitation is calculated using an automatic rain gauge that calculates how many

millimeters of rain are gathered every ten minutes, and then we choose the maximum amount for that hour. It is associated with the CSV file *pf.csv*. The municipality offers us the water consumption every hour in liters. This information is saved in a CSV file named *wcf.csv*.

3.1 Processing the training data

We use Python [9], a high-level all-purpose programming language running on the Jupyter Notebook (a web-based interactive computing platform) to develop explainable prediction models. We import the Python library Pandas for data manipulation: to create and manipulate the sensor data into usable CSV files, to load the dataset as a data frame and arrange the raw data in the format of the desired array.

```
import pandas as pd
from warnings import simplefilter
import os
import sys
# Assign main directory to a variable
dir=os.path.dirname(sys.path[0])
# Ignore all future warnings
simplefilter(action='ignore', category=FutureWarning)
```

We used historical data to train our machine learning model. We had to first index all the values to the same time stamp intervals so that we had a matching index for all the variables when obtaining the data. We rounded all the time stamps to the nearest hour, where the hour is our index and chose the greatest value for each time stamp. After processing the data in this manner, we used append on the Pandas data frame to join all of the files together, yielding a CVS file with all values indexed in full rows.

```
def get_dataset():
    # Array to specify the csv files containing sensors data
    csvFiles = [dir+'/sewer/dataset/wuf.csv',
     dir+'/sewer/dataset/pf.csv', dir+'/sewer/dataset/gwf.csv',
     dir+'/sewer/dataset/ff1.csv', dir+'/sewer/dataset/ff2.csv',
     dir+'/sewer/dataset/ff3.csv', dir+'/sewer/dataset/ff4.csv',
```

```
     dir+'/sewer/dataset/ff5.csv', dir+'/sewer/dataset/ff6.csv',
     dir+'/sewer/dataset/ff7.csv', dir+'/sewer/dataset/ff8.csv',
     dir+'/sewer/dataset/ff9.csv', dir+'/sewer/dataset/ff10.csv',
     dir+'/sewer/dataset/ff11.csv', dir+'/sewer/dataset/ff12.csv']
    # Array to store the rounded files for combination
    l = []
    for f in csvFiles:
        # read the csv file,in this case they used the delimiter ','
        r = pd.read_csv(f, delimiter=",")
        # Round each time stamp to the nearest hour
        r['Time Stamp'] = pd.to_datetime(r['Time Stamp']).round('h')
        # Save this to a new csv file and set indexing to false
        r.to_csv(dir+'/sewer/dataset/r.csv', index=False)
        # read the file again but now set the index to column 0 (ST)
        fr = pd.read_csv(dir+'/sewer/dataset/r.csv', index_col=0)
        # pick the highest value for that index
        fr = fr[~fr.index.duplicated(keep='first')]
        # append to l to get stored
        l.append(fr)
    # Turn the array into a frame and save to file
    frame = pd.concat(l, axis = 1)
    frame = frame[~frame.index.duplicated(keep='first')]
    # Fill missing row data with the past data if it exists
    frame.fillna(method='pad', inplace=True)
    # Or with the next data if the past data doesn't exist
    frame.fillna(method='bfill', inplace=True)
    frame.to_csv(dir+'/sewer/dataset/com.csv')
    df_features = pd.read_csv(dir+'/sewer/dataset/com.csv',
                        index_col = 'Time Stamp', parse_dates=True)
    df_labels = pd.read_csv(dir+'/sewer/dataset/Labels.csv',
                        index_col = 'Time Stamp', parse_dates=True)
    return df_features, df_labels

import plotly.figure_factory as ff
def save_frame(df):
    fig =  ff.create_table(df.round(2),index=True)
    fig.update_layout(autosize=False,width=1000,height=200)
    fig.write_image(dir+'/sewer/dataset/table_plotly.png', scale=2)
```

We begin by creating the dataset "$X, y = get\_dataset()$" to get the training input X and the training output y. Figure 1, which is produced by the call "$save\_frame(X.iloc[: 5, :)$" depicts the first five

lines of the training data after it has been processed. The first column represents time stamps, the index for our experiment. The second column shows the water consumption over an hour period. The area's precipitation is shown in the third column, while the fourth column represents the groundwater level. The last twelve columns represent the different flows of wastewater through the pipes, which can measure how much wastewater flows through the pipes in liters per hour in different locations.

	Water Consumption	Precipitation	Ground Water Level	Flow 1	Flow 2	Flow 3	Flow 4	Flow 5	Flow 6	Flow 7	Flow 8	Flow 9	Flow 10	Flow 11	Flow 12
2021-01-01 00:00:00	462.54	0.16	4.2	341.25	304.12	329.13	330.12	272.95	289.12	325.38	287.02	298.69	249.25	288.35	222.81
2021-01-01 01:00:00	408.53	0.0	3.46	387.04	317.16	247.02	380.81	282.62	313.51	264.85	258.93	340.66	281.51	293.53	237.18
2021-01-01 02:00:00	483.12	0.0	3.9	270.67	363.98	328.71	244.54	367.29	359.11	262.75	376.86	409.99	314.99	282.19	267.16
2021-01-01 03:00:00	392.06	0.63	3.88	243.39	314.46	387.5	311.55	382.35	318.92	376.72	279.88	293.57	400.94	293.23	348.14
2021-01-01 04:00:00	416.18	0.0	3.88	359.0	233.58	409.62	271.56	392.15	330.17	253.22	243.49	353.37	356.3	307.68	313.21

Figure 1. Data frame of proceeded data

3.2 Definition of the prediction model

The prediction model is a Keras sequential model composed of two dense layers. A dense layer is the regular deeply connected neural network. It returns the result of activation(dot(input, kernel) + bias) as its output. If the activation is none, then the output layer of the network is performing regression, which should be naturally linear.

```
from keras.models import Sequential
from keras.layers import Dense
def define_model(n_inputs, n_outputs):
    model = Sequential()
    model.add(Dense(32, input_dim=n_inputs,
              kernel_initializer='he_uniform', activation='relu'))
    model.add(Dense(n_outputs, kernel_initializer='he_uniform'))
    model.compile(loss='mae', optimizer='adam', metrics=['accuracy'])
    return model
```

The first dense layer has $n\_inputs = 12 + 3 = 15$ inputs and 32 outputs with weights initialized by $heUniform()$ and activation function $relu()$. The second dense layer has 32 inputs and 12 outputs ($n\_outputs = 12$) and without activation function (linear regression

model). The compilation uses the loss function *mae*() and the optimizer *adam*().

```
def create_model():
    # Get the number of inputs and outputs from the dataset
    n_inputs, n_outputs = X.shape[1], y.shape[1]
    model = define_model(n_inputs, n_outputs)
    history = model.fit(X, y, verbose=1, batch_size=10,
                        epochs=100, validation_split=0.2)
    summarize_diagnostics(history)
    model.evaluate(x = X, y = y)
    return model
```

We create an object of our model "*model* = *create_model*()". Figure 2 shows its accuracy, which is more than 75%. These metrics about the model are visualized by the call to the function below that generates a "png" file to plot curves for loss cross entropy and classification accuracy.

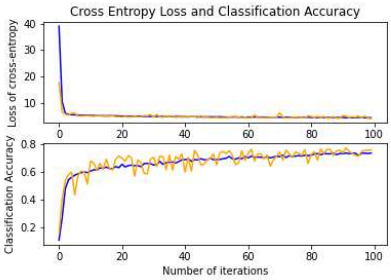


Figure 2. Cross entropy and classification accuracy

```
# This import is for visualizations
from matplotlib import pyplot
def summarize_diagnostics(history):
    pyplot.subplot(211)
    pyplot.title('Cross Entropy Loss and Classification Accuracy')
    pyplot.ylabel('Loss of cross-entropy')
    pyplot.plot(history.history['loss'], color='blue', label='train')
```

```
    pyplot.plot(history.history['val_loss'],color='orange',label='test')
    pyplot.subplot(212)
    pyplot.xlabel('Number of iterations')
    pyplot.ylabel('Classification Accuracy')
    pyplot.plot(history.history['accuracy'],color='blue',label='train')
    pyplot.plot(history.history['val_accuracy'],color='orange',
                label='test')
    pyplot.savefig(dir+'/sewer/dataset/sewer_plot.png')
    pyplot.close()
```

For prediction explanation, we use the package SHAP (SHapley Additive exPlanations), which is a game theoretic approach to explain the output of any machine learning model. It connects model prediction with local explanations using the classic Shapley values from game theory and their related extensions [10].

```
import shap
def create_explainer():
    # print the JS visualization code to the notebook
    shap.initjs()
    explainer = shap.KernelExplainer(model = model.predict,
                                     data = X, link = "identity")
    return explainer
explainer = create_explainer()
```

The jupyter-widgets (ipywidgets) are interactive HTML widgets for Jupyter notebooks and the IPython kernel. The interactive widgets are used in notebooks to gain control and visualize changes in the data. Widgets are eventful Python objects that have a representation in the browser as control. They can be used to build interactive GUIs for notebooks. We can also use widgets to synchronize stateful and stateless information between Python and JavaScript.

```
# This import is for feature selection
import ipywidgets as widgets
def create_labels():
    # Create the list of all labels for the drop down list
    list_of_labels = y.columns.to_list()
    # List of tuples: the index of the label is what is returned
```

```
    tuple_of_labels = list(zip(list_of_labels,
                                range(len(list_of_labels))))
    # Create a widget for the labels and then display the widget
    current_label = widgets.Dropdown(options=tuple_of_labels,
              value=0, description='Select Label:', disabled=False)
    return list_of_labels, current_label
list_of_labels, current_label = create_labels()
```

The call to the function *create_labels*() will create a list of labels and a drop down to select the current label to be used for prediction explication to take decisions. To explain the model prediction on a data frame (df) representing the observation values of the flow level sensors, we call the function below and we specify the name of the current label selected by the drop down selection. The explanation can be saved as a "png" file. This GUI contains a button to generate explanation for the model predication, which is developed with the package SHAP.

```
def explain_model_prediction(df):
    shap_value_single=explainer.shap_values(X=df,nsamples=100)
    return shap.force_plot(base_value =
       explainer.expected_value[current_label.value], shap_values =
       shap_value_single[current_label.value], features = df),
       current_label.value

def save_plot(df):
    shap_value_single=explainer.shap_values(X = df, nsamples = 100)
    shap.force_plot(base_value =
       explainer.expected_value[current_label.value], shap_values =
       shap_value_single[current_label.value], features = df,
       matplotlib = True, show = False)
    pyplot.savefig(dir+'/sewer/dataset/'+
               str(list_of_labels[current_label.value])+'.png')
    pyplot.close()
```

4 Experimental results

To use this model, we generate a data frame "df" from real-time information sent by sensors to predict the labels values, which are shown in Figure 3. Then, we select a label to explain its prediction.

```
df = pd.DataFrame({'Time Stamp' : ['2021-01-01 00:00:00'],
   'Water Consumption' : [462.54], 'Precipitation' : [0.16],
   'Ground Water Level':[4.20],'Flow 1':[341.25],'Flow 2':[304.12],
   'Flow 3' : [329.13], 'Flow 4' : [330.12], 'Flow 5' : [272.95],
   'Flow 6'  : [289.12], 'Flow 7' : [325.38], 'Flow 8' : [287.02],
   'Flow 9' : [298.69], 'Flow 10' : [249.25], 'Flow 11' : [288.35],
   'Flow 12' : [222.81]})
df = df.set_index('Time Stamp')
pr = model.predict(df)
# Display the prediction values of the labels
for i in range(len(list_of_labels)):
    print('Label '+str(i+1)+' output: ', pr[0,i], '%')
# Display the drop down list for selection of a label
current_label
```

```
Label 1 output:  44.55138 %
Label 2 output:  53.40689 %
Label 3 output:  52.180103 %
Label 4 output:  58.580154 %
Label 5 output:  50.863182 %
Label 6 output:  60.688377 %
Label 7 output:  51.26418 %
Label 8 output:  57.477436 %
Label 9 output:  51.55485 %
Label 10 output:  43.892826 %
Label 11 output:  53.034832 %
Label 12 output:  60.96122 %
Select Label: Label 8
```

Figure 3. Prediction of outputs and selection of label number 8

Figure 4 generated by the following explanation code, explains the model prediction on the label number 8, where its predicted value is 57.477436.

```
explain_result, current_label_value = explain_model_prediction(df)
print(f'Current label Shown: {list_of_labels[current_label_value]}')
explain_result
```

The explanation shows how the observations on other sensors have affected this value from higher effect on the left to the lower effect on the right. This explanation result helps us to diagnose our sewer system.

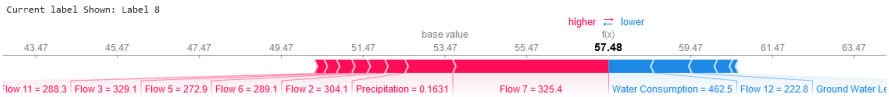


Figure 4. Prediction explanation

To explain prediction on other label output, it suffices to select that label and rerun the explanation code to have the prediction explication.

5 Conclusion

In this paper, a Python program running on the notebook of jupyter-lab is developed to create data-driven prediction model. This model is trained with data collected using the Internet of Things. After data processing and model training, the result model will be used to predict outputs of the real-time inputs. A corresponding explainer is developed with Shap to explain the prediction, which will help us to diagnose the behavior of our sewer system. Our perspective is to integrate this work with the simulator based on ArcGIS system used in the municipality of M'Sila.

References

[1] A. Ferriman. *BMJ readers choose the sanitary revolution as greatest medical advance since 1840*, BMJ 2007, pp. 334 – 111.

[2] Ane Loft Mollerup, Peter Steen Mikkelsen, and Gürkan Sin, *A methodological approach to the design of optimising control strategies for sewer systems*, Environmental Modelling & Software, Volume 83, 2016, pp. 103–115.

[3] Nadia Lund, Morten Borup, Henrik Madsen, Ole Mark, and Peter Mikkelsen. *CSO Reduction by Integrated Model Predictive Control*

of Stormwater Inflows: A Simulated Proof of Concept Using Linear Surrogate Models, Water Resources Research. volume 56, DOI: 10.1029/2019WR026272, 2020.

[4] K. Jenssen Sola, J.T. Bjerkholt, O.G. Lindholm, and H. Ratnaweera. *Infiltration and Inflow (I/I) to Wastewater Systems in Norway, Sweden, Denmark, and Finland*, Water 2018, 10, 1696. https://doi.org/10.3390/w10111696.

[5] S. Li, L. D. Xu, and S. Zhao. *The internet of things: a survey. Information systems frontiers*, 17(2), pp. 243–259, 2015.

[6] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*, 4th Edition. Pearson, 2020.

[7] N. Razavian, F. Knoll, and K. J. Geras. *Artificial intelligence explained for nonexperts*, In Seminars in musculoskeletal radiology, vol. 24, no. 01, pp. 003–011, Thieme Medical Publishers, February 2020.

[8] Z. K. Collier, H. Zhang, and L. Liu. *Artificial intelligence for propensity score estimation in multilevel educational settings*, Practical Assessment, Research, and Evaluation, 27(1), 3, 2022.

[9] V. R.Guido and F. L. Drake Jr. *Python reference manual*, Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica, 1995.

[10] S. M. Lundberg and S.-I. Lee. *A Unified Approach to Interpreting Model Predictions*, I. G. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett, Advances.

Mustapha Bourahla

Computer Science Department/University of M'Sila, Algeria
E–mail: mustapha.bourahla@univ-msila.dz

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Tools for Triaging in Mass Casualty Incidents

Olesea Caftanatov, Tudor Bumbu

Abstract

In this paper, we propose two tools for triaging process in mass casualty incidents. One tool is a mobile application that records, analyzes, and sets transportation order. The second tool is a web platform for management and storage purpose.

Keywords: mass casualty incidents, tools, triage processes, management platform.

1 Introduction

During any natural or man-made incident, which results in extraordinary levels of mass casualties, local and state medical personnel are often overwhelmed by the sheer magnitude of the situation. It was found that the biggest challenges to providing care are resources and communication restraints. Thus, the key to use the resources more efficiently is to keep patients moving toward definitive treatment through accurate triage of life threats.

The term triage originates from the French word trier, meaning *to sort*. In the medical context, this *"sorting"* is a method of selection and classification of patients, by priority, for initial treatment and subsequent transportation to a facility, where more definitive care is available. The classical triaging is made on paper forms, but completing them is a time consuming procedure. Additionally, classical triage has communication restraints. Due to modern technologies, we can digitize the triaging process, and this is our main purpose of this research.

2 Related Works

In the last decade, the development of tools for Mass Casualty Incidents (MCI) obtained great interest. According to [1], Montano et. al. reviewed and analyzed the state of the art regarding triage applications for health emergencies. Based on their systematic review of the literature in the scientific database from 2010 to early 2021, only 13 applications were identified from 26 relevant papers. Surprisingly, they also observed that despite the existence of much research, only 3 applications are accessible.

Usually, the triage are guidelines in a paper form; for example, clinical guidelines for major incidents and mass casualty events proposed by NHS England [2]. In the UK, triage at mass casualty incidents is performed in two steps. The first step takes place at the scene of the incident.

The primary triage assessment takes no more than 30 seconds per patient. Its objective is to rapidly identify those patients who need a life-saving intervention. According to [3], the first step is executed by using an algorithm such as the Modified Triage Sieve – MPTT-24.

The second step takes place in a more permissive environment and is performed by a more experienced clinician. There are more types of the secondary triage processes, such as MIMMS Triage Sort, Anatomical Triage, and Clinical Gestalt. For our research, we used the algorithm that was proposed by our colleagues [4].

3 Tools

The tools we focus in this paper are a mobile application designed for triaging in the field and a triage data management platform, where the data acquired in the field is stored and analyzed by specialists. Primarily, the mobile app stores and shares the data locally with all medics and paramedics in the field via Bluetooth LAN as there might be problems with internet connection (see Fig.1). When there is internet connection, the triage data is synchronized with the server and stored into the server database and further managed from the web platform. In the next subsections we describe these tools.

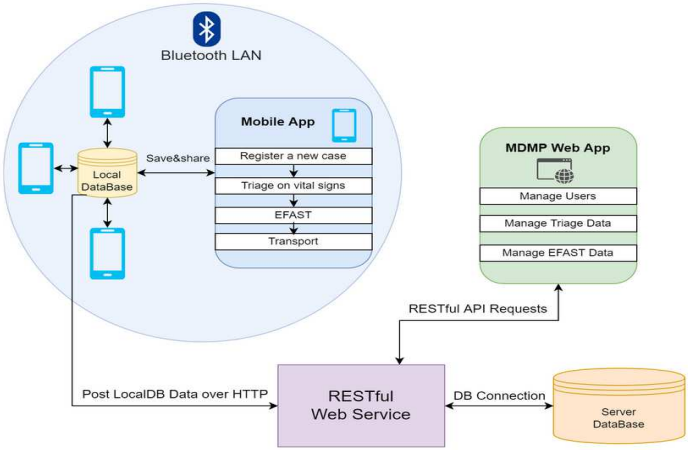


Figure 1. Architecture of communication between tools

4 Mobile Application

Mobile Application was developed by using Android Studio Arctic Fox v.2020.3.1. powered by the InteliJ Platform (see Fig.2 a.).

National Association of Emergency Medical Services Physicians (NAEMSP) proposed SALT [5] to triage and move Mass Casualty Incidents (MCI) patients forward to resources. SALT is a four-step process for first responders to manage mass casualty incidents, and stands for:

1. **Sort**;
2. **Assess**;
3. **Lifesaving interventions**;
4. **Treatment and/or transport**.

We define 3 steps in our triaging process:

1. **FAST**;
2. **EFAST**;
3. **Transportation**.

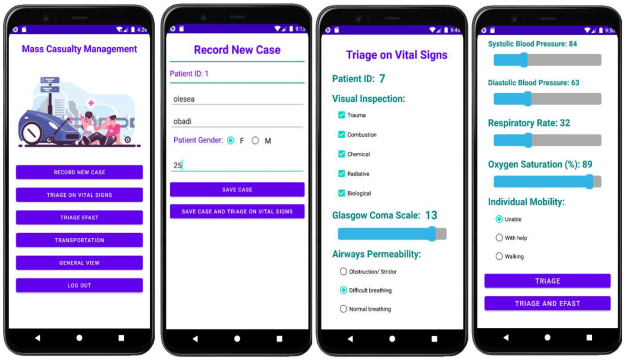
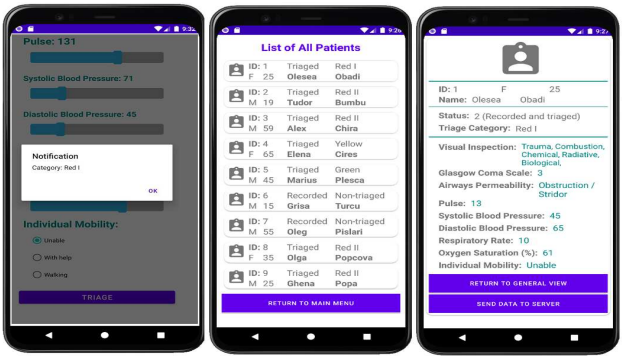


Figure 2. Mobile application home interface and FAST process interfaces

The FAST process, also called the triage on vital signs, has the feature to be completed in one go by one medical assistant (see Fig.2 b.-d.) or step-by-step by 2 people. For the cases, when they are completed by 2 people, one of them completes only **Record New Case** interface (see Fig.2 b.), the other one completes **Triage on Vital Signs** interface (see Fig.2 c., d.).

Our colleagues from Informational Systems Laboratory [6] developed a scoring system based on decision rules that allows re-assessment of triage priorities for casualties. Structurally, the score is in line with the well-known scoring systems, widely used in medical diagnostics. Based on a scoring system algorithm, our application at the end of completing the triage, on vital signs gives a quick categorization of casualties in Red I, Red II, Yellow, and Green, (see Fig.3 a.). The Red categories need immediate transportation, while the Green one is processed the last in order.

The added general view window allows users to visualize all basic information about injured persons: ID, First name, Last name, Gender, Age, Patient Status (Non-triaged, Triaged), and Triage category (see Fig.3 b.). Selecting any casualty from this list allows the user (with the



a. b. c.

Figure 3. Notification category based on system score algorithm and General View interfaces

corresponding rights) to proceed to the next stage of the management of mass casualty situations, or to monitor the overall performance (see Fig.3 c.).

The stage of triage EFAST (see Figure 4.a) allows to introduce information, using additional portable ultrasound equipment, regarding presence of free fluid in the following areas:

1. Right upper quadrant
2. Left upper quadrant
3. Pelvic view
4. Pneumothorax (left and right)
5. Hemothorax (left and right)
6. Cardiac tamponade.

The decision rules were created for EFAST, helping to clarify the triage category more accurately, for instance, to change it from Yellow into Red in case of free fluid detected (presumed to be blood under disaster conditions). Medical experts have validated these decision rules.

The stage of transportation allows medical assistants to select the corresponding type of ambulance (type C – intensive care ambu-

lance, type B – emergency ambulance, type A – non-emergency transportation) from the available ones, destination hospital, and the order of transportation (see Fig.4 b.,c.).

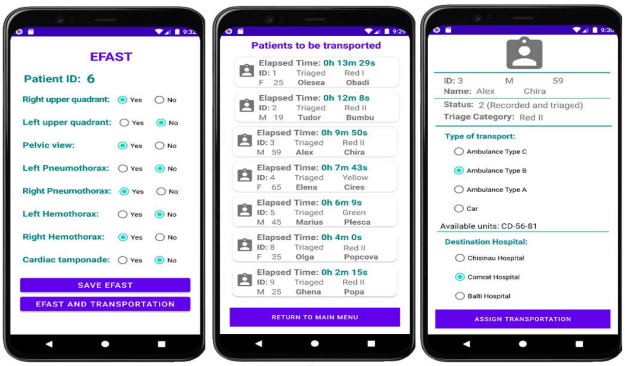


Figure 4. a. The EFAST process interface; b. and c. The transportation process interface

For every stage of the management of mass casualty situations, there was tested the stage's main functionality, modules integration, and if the developed software meets the needs of the user. While designing interfaces for mobile applications, the factor that medical assistants will use gloves was considered, so all buttons should be as big as possible. The main factor in designing was that the time consumption should be as short as possible. Therefore, the interface is intuitive and friendly.

5 Medical Data Management Platform

Data collected in the mobile application during the stage of triage of vital signs can be saved, stored, and managed from a web application developed within the Django framework. We call this project Medical Data Management Platform or MDMP (see Fig.5).

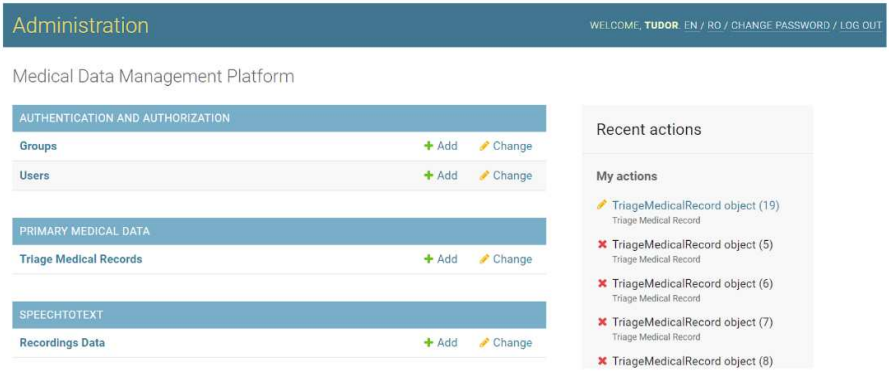


Figure 5. The MDMP homepage in English

The MDMP consists of 3 modules: the User Group Administration module; the Patient Medical Data Administration module; and a special module for audio processing and management, namely the Audio Recordings Administration module which is still in development and not described in this paper. The platform communicates with the mobile application via REST API by requesting the data of registered patients triaged as long as internet connection is established. The API allows GET and POST request methods.

In the User Group Administration module, the site administrators can manage the groups and users of the platform (see Fig.6). Also, the administrator can manage the permission of a user or a group of users, such as view, edit, or delete medical records. The following groups of users were added: Triage Members, the Medical Personal consisting of medical assistants and doctors which will review the data in the Primary Medical Data Administration module and prepare reports. Also, there is the group of Administrators including the managers and MDMP administrators. In this module, the Administrator can add new groups of users.

The next module is Patient Medical Data Administration (see Fig.7) for managing recordings of registered and triaged patients sent

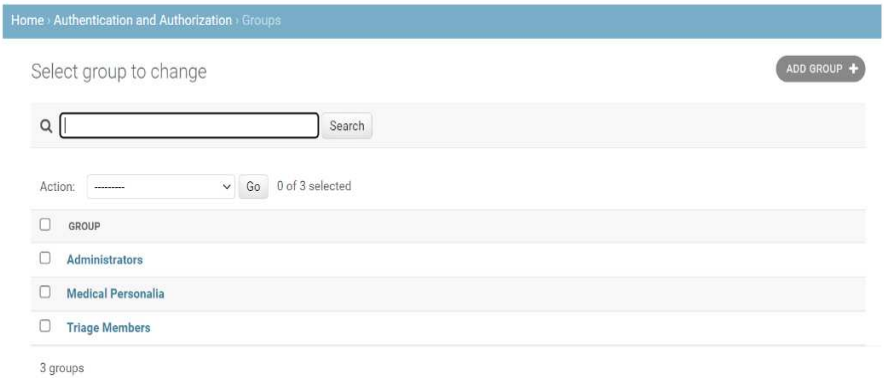


Figure 6. User groups management page

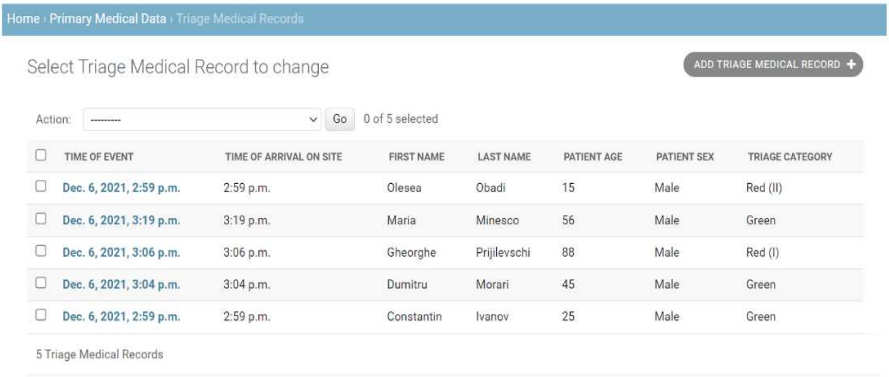


Figure 7. Patient medical data administration pag

from mobile application. The users within the Medical Personal group can view, edit, or delete these recordings. A recording from the table shown in the Figure 7 displays 7 attributes which include one of the most important attributes which is the Triage Category.

6 Conclusion

The developed tools, i.e., the mobile application for triage and the web platform for managing and analyzing casualties offer simple and user-friendly interface, allowing medical first aid personnel the following actions: to gather and organize the primary medical data of casualties; to perform triage based on vital signs and assign the triage priority for quick categorization of casualties **(Red I, Red II, Yellow, Green)**; to store and analyze the data acquired in the field.

It can be done for casualties with injuries and can be repeated during transportation; the priorities can be set for evacuation of injured persons from the disaster site and for routing them to the specialized medical centers (including the transmission of the casualty-related data).

The impact of these tools is crucial by doing fast registration and triage priority assessment; accurate casualty triage reassessment and more effective emergency therapy before further transportation, that will minimize over- and undertriage; coordinated evacuation of casualties will help in efficient distribution of the available resources; data analysis for improving the mass casualty management.

Acknowledgments. "Intelligent Information systems for solving ill structured problems, knowledge and Big Data processing project" Ref. Nr. 20.80009.5007.22, has supported part of the research for this paper.

References

[1] I.H. Montano, I.T. Diez, R.L. Izquierdo, M.A. Castro Villamor, and F.M. Rodriguez. *Mobile Triage Ap-*

plications: A Systematic Review in Literature and Play Store. J Med Syst, 2021; 45(9): 86. Published online 2021 Aug 13. doi: 10.1007/s10916-021-01763-2. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8361243/.

[2] NHS England (2019). *Clinical Guidelines for Major Incidents and Mass Casualty Events.* Guidelines online. Available: https://naru.org.uk/wp-content/uploads/2019/01/version1__Major_Incident_and_Mass_casualty_guidelines-Nov-2018.pdf. [Accessed 25 Jul 2022].

[3] R. Mersh and J. Vassallo. *Triage in Mass Casualty Situations* RCEM Learning. Published online: 06/01/2020. Available: https://www.rcemlearning.co.uk/reference/triage-in-mass-casualty-situations/#1572966037676-482ad5bd-c7b3. [Accessed 12 Aug 2022].

[4] S. Cojocaru, C. Gaindric, I. Secrieru, S. Puiu, and O. Popcova. *Multilayered Knowledge Base for Triage Task in Mass Casualty Situations*. Computer Science Journal of Moldova, vol. 24, no. 2(71), 2016, pp. 202–212.

[5] R. Duckworth. *How to use SALT to triage MCI patients*. Published on EMS 1 by LEXIPOL, 27 April, 2021. [Accessed 9 Aug 2022]. Available: https://www.ems1.com/mass-casualty-incidents-mci/articles/how-to-use-salt-to-triage-mci-patients-ioh8pD88282FDTdy/.

[6] Iulian Secrieru, Constantin Gaindric, Elena Guțuleac, Olga Popcova, Tudor Bumbu. *Formalization of decision knowledge and reasoning for casualty prioritizing*. Proceedings of the Workshop on Intelligent Information Systems (WIIS2022), to be published.

Olesea Caftanatov[1], Tudor Bumbu[2]

[1]Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: olesea.caftanatov@math.md

[2]Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: tudor.bumbu@math.md

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Developing augmented artifacts based on learning style approach

Olesea Caftanatov, Inga Titchiev, Veronica Iamandi,
Dan Talambuta, Daniela Caganovschi

Abstract

In this article, the possibilities offered by augmented reality in education are used based on learning style approach in order to increase the efficiency and quality of the study process.

Keywords: Augmented reality, artifacts, marker, learning style, Bartle's player taxonomy, Bloom verb taxonomy.

1 Introduction

Augmented Reality (AR) are technologies which enhance our perception and help us to see, hear, and feel our environments in new and enriched ways. The idea of augmented reality [3] is not new, these technologies have been developed and researched during the last years. Due to different possibilities that it offers, it has become a new trend including the education, but these opportunities are still unsettled and little applied. Thus, in this article, we will research the approach based on the development of augmented artifacts taking into account the learning preferences (learning styles) of the subject, types of students based on Bartle's characters theory. Moreover, for the development of the augmented artifacts, marker-based Augmented Reality approach was used [5], which allows the inclusion of different scenarios, being applicable for different types of learning styles. More about the principle used

in developing augmented artifacts is described in the methodology section. Additionally, while we get excited about various AR features that can be used in the Education field, the image trigger is fundamental to the quality experience. Therefore, we present a few recommendations in designing artifacts and markers in Section 3.1.

2 Background research

2.1 Augmented reality

AR technology has reshaped the way we interact with the real world. Augmented reality is the technology that expands our physical world, adding layers of digital information on it. AR is often mistaken for virtual reality (VR). While they do share pieces of development history, the two are not the same. Unlike VR, augmented reality does not create completely artificial environments to replace the real world with virtual one, it blends technology with the real world. It appears in direct view of an existing environment and adds multimedia elements to it, such as: video, sounds, graphics, etc. AR leaves a little to science and a lot to the imagination. Augmented reality is not a specific device or program, it is a type of human-computer interaction. Augmented reality tech was invented in 1968, with Ivan Sutherland's development of the first head-mounted display system. However, researcher Thomas Caudell coined the term "augmented reality" in 1990 [1].

In the specialized literature, there are various classifications of types of AR. For instance, in [4], the author highlights four types of AR (marker-based, markerless, projection-based and superimposition-based AR), the Wilson's team [15] members think about AR in terms of five distinct experiences types (video launch, 3D object, 360-degree surround, interactive game, information overlay). We identify AR in two major categories: marker-based AR [5] and markerless AR [6], because the other types can be a variation of these two. For our research, we used a marked-based augmentation reality approach to develop artifacts and implement various scenarios in them.

3 Methodology

The research approach develops an augmentation artifact based on five principles:

1. **Marker-based AR**. There are various types of augmented reality; but for our research, we implemented the most common type, which uses markers to trigger an augmentation experience. Due to its use of image recognition, this type of AR is sometimes also called a recognition-based augmented reality.

2. **Bartle's player taxonomy**. There is no field more experienced in engaging users in activity than players in the gaming industry. Every type of player, in our case – students, is unique and special, with their own motivations for engaging. We believe it's nearly impossible to assess and cater to each type of personality. Hence, understanding our audience is important and we need a taxonomy and some kind of assessment system. We decided to classify students based on character theory and player behavior presented in Bartle's taxonomy [18].

3. **VAK learning style**. Having our own preferences for learning, we are all different, this leads to distinct behavioral manifestation [7]. Depending on person, task, context, previous experience, education, etc., behavioral manifestations become over time constant, stable, and frequently applicable, turning into preferences (so-called learning styles) in the learning process. If the teaching material is designed to fit a learning style, then the student who characterizes that style understands the new information better and, consequently, has better results. The most frequently operated in educational practice are the typologies proposed by Barbe, Swassing, and Milone [2], who differentiate the following learning styles (VAK styles); each of these learning styles has particular characteristics [8] as well as learning strategies that can be applied on a case-by-case basis:

(a) **Visual Learners** – prefer viewing written information; transcribing it; using tools for study such as: illustrations, maps, tables, graphs, images, diagrams; emphasizing basic ideas.

(b) **Auditory Learners** – prefer reading aloud; explaining new information, expressing ideas verbally; learning with a tutor or in a group, where they can ask questions, provide answers, express how to understand information orally.

(c) **Kinesthetic Learners** – prefer handling the objects to be learned; arranging the tables and diagrams in the correct order; using movements, dancing, pantomime or role-playing; talking and walking while repeating the knowledge, and applying the learned knowledge in practice.

4. **Bloom verb taxonomy**. Given the fact that, after each lesson taught in schools, high schools, universities, students must obtain some skills, the ideology of the proposed personal learning scenarios concept can be represented by the formula below: S_g–$S_c = S_n$ (1), where S_g represents the general skills, S_c – the current skills, and S_n – the necessary skills to be obtained. It is important to note that the list of competencies is grouped according to Bloom's taxonomy [17].

5. **Geometry 5 grade curriculum**. The material used for augmentation experiences was prepared in accordance with the curriculum for 5 grade, for the geometry subject[16].

Below we present the interaction between artifacts with the mentioned principles (see Fig. 1).

3.1 Designing augmented artifacts and markers

An augmented artifact and marker has quite a few tasks to accomplish. Besides the fact that it has to capture student's attention, entice them to pick up their mobile and scan the image, it should have a high quality

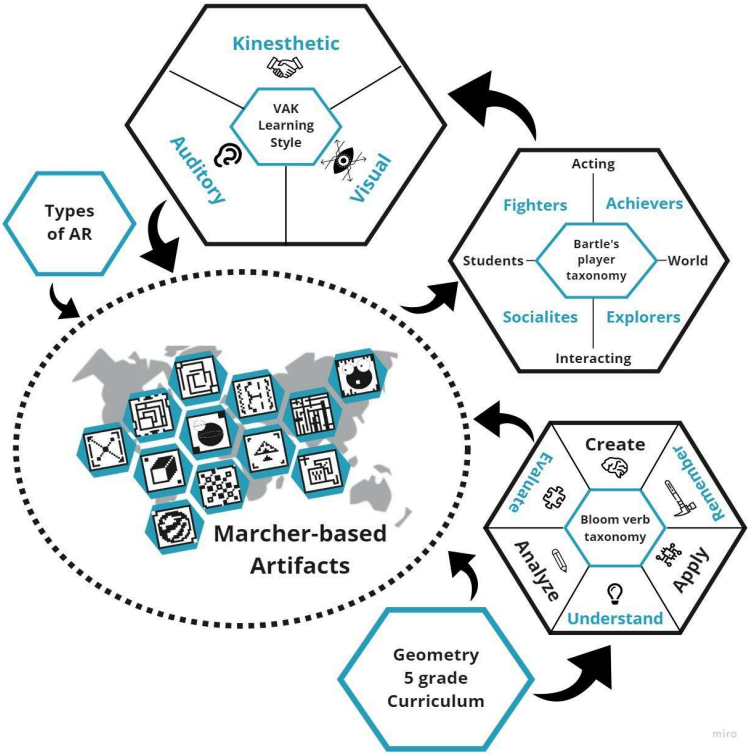


Figure 1. Interaction between artifacts and principles

to let the AR experience come to life. Therefore, in this section, we will describe the best practice to designing augmented artifacts and markers that we observed as a result of the working process and testing. Moreover, we will describe our experiences with low and high star rating image targets.

We consider that **"markers"** are the digital form of image targets that Vuforia Engine can detect and track by comparing extracted natural features from the camera image against a known image target resources database. Markers come in various forms: simple, flat image targets, curled targets in the form of cylindrical shapes, or multi-targets in the composition of a box.

We define **"artifacts"** as the physical form of markers. They can also come in various forms: cards, papers, newspapers, posters, objects, etc. In our cases, it is a laminated image with size 10 x 10 cm. The main purpose of the artifact is to trigger the augmentation content when it is scanned by camera.

When it comes to **designing artifacts**, there are few recommendations on obtaining the best performance from physical target images. Artifacts should be rigid, not flexible. A hard material such as card stock or plastic is better than a simple printed piece of paper, because the flexibility of the printed piece of paper can make it difficult for the object to stay in focus. However, paper artifacts are easily reproducible and widely available, so make sure to fix them to a non-flexible surface.

The size of artifacts varies based on the actual target rating and the augmentation experience. At least 10 or 12 cm (4 – 5 inches) are good for user manipulation. It also depends on the distance between the camera and the artifact; the larger the distance, the bigger your target should be. As an estimate, an artifact with size 20-30 cm wide should be detectable up to about 2-3 meter distance, which is about 10 times the target size. Another important factor is the flatness of the artifact; those that easily bend, coil up, and wrinkle degrade significantly the quality of the tracking. Last but not the least, for attractiveness purposes, the artifacts should be matte but not glossy. Printouts from modern laser printers can be very glossy. Although under ambient

lighting conditions, the glossy surface is not a problem, even so, under certain angles, light sources such as sun, lamp light, etc., can create a glossy reflection. Sometimes the reflection can cover up potentially large parts of the artifacts; as a result, tracking and detection issues can appear. One more recommendation is that when creating artifacts, we should let out our imagination, to make them more creative, relevant, and fun.

When it comes to **designing markers**, there is a range of factors that define how well it tracks when uploaded to Vuforia Target Manager. According to Vuforia Library guideline [14], markers that possess the attributes (see Table 1), will enable the best detection and tracking performance from the Vuforia Engine.

Table 1. Attributes of an ideal image target

Attribute	**Example**
Rich in detail	Street scene, group of people, collages and mixtures of items, and sport scenes are good examples.
Good contrast	Images with bright and dark regions and well-lit areas work well.
No repetitive patterns	Employ unique features and distinct graphics covering as much of the target as possible to avoid symmetry, repeated patterns, and feature-less areas.
Format	Must be 8- or 24-bit PNG and JPG formats; less than 2 MB in size; JPGs must be RGB or greyscale (no CMYK).

There are a few more recommendations to take in account when designing markers. The most important rule here is choosing a unique photo from your collection; however, we decided to create a design from scratch (see Fig. 2). The risk of using stock photos is that someone else may also pick the same picture; as a result, the augmentation experi-

ence may have some issues. For example, when markers are scanned by AR camera, both sides of the AR experience are distorted because the software will recognize the image but will not differentiate the content that has to be retrieved.

Figure 2. Examples of markers version 1.0

Vuforia Engine uses the grayscale version of markers to identify features that can be used for recognition and tracking. If the image has low overall contrast and the histogram of the image is narrow and spiky, it is not likely to be a good target image. Our first batch of markers was designed black-white colors, so after evaluation they were rated 2-3 stars.

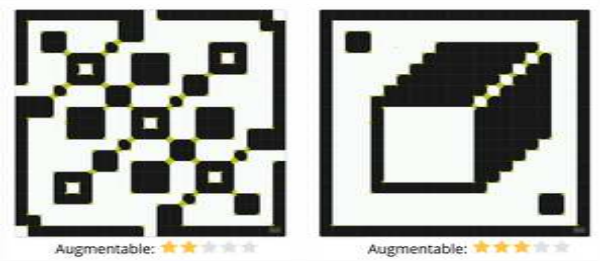


Figure 3. Markers with augmentable rating 2 and 3 stars

An augmentable rating defines how well an image can be detected and tracked using the Vuforia Engine. This rating is displayed in the Target Manager and is returned for each uploaded target via the web API, when using Cloud Reco Databases (see Fig. 3). The augmentable rating can range from 0 to 5 for any given image, where zero indicates that a target is not tracked at all by the AR system.

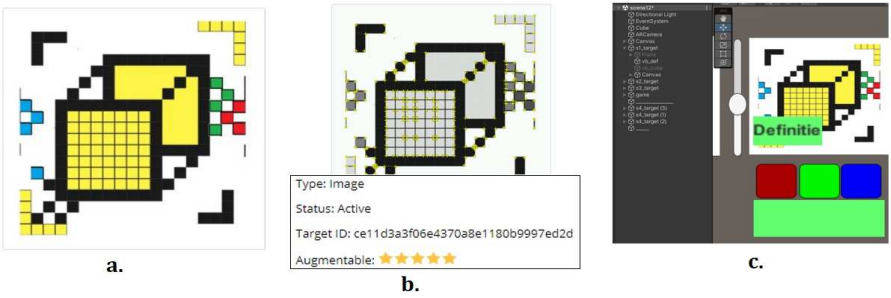


Figure 4. a) marker for square; b) marker evaluations 5 stars through Vuforia Target Manager; c) adding virtual button to marker through Unity platform.

On the other side, the rating of 5, indicates that the marker contains strong detection and tracking ability and is easily recognized by the AR system. Markers with 2-3 rate stars can still be used to trigger simple augmentation content such as visualization of a 3D object, video lessons, text, audio file; but the interaction is almost impossible. For the interaction content, 4-5 star rating is needed. For example, when adding virtual buttons to be easily detected and tracked, it needs to be placed in a zone with many features (see Fig. 4). A feature is a sharp, spiked, chiseled detail in the image, such as the ones present in textured objects. The image analyzer represents features as small yellow crosses; for example, square figure has 4 yellow crosses, circle figure has zero features. Other general features such as organic shapes, round details, blurred, or highly compressed images often do not provide enough richness and detail to be detected and tracked properly.

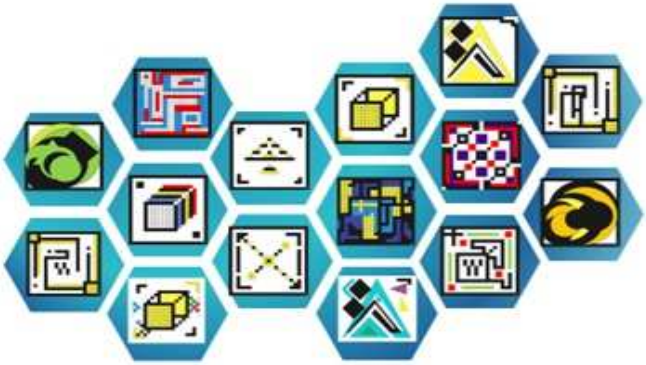


Figure 5. Example of markers version 2.0

After testing the first version of markers, we level up them by adding more color and adjusting contrast; the next batch was created (see Fig. 5). The second version after evaluation got 4-5 rating stars. All features extracted from these images are stored in a cloud database, of which the latter can be downloaded and packaged together with the application (see Fig. 6). The database can then be used by Vuforia Engine for runtime comparisons.

While working with markers, version 2.0, as we observed in some cases, gets a good recognition but a bad tracking during AR experiences. After a few tests, we came to the conclusion that repetitive patterns will confuse the computer vision and will be perceived as the same; it is even more consuming to detect in which direction the marker is placed and to retrieve it. Thus, choosing irregular shapes and photos that look different from all angles, so the computer vision knows if it is upside down per se.

3.2 Results

There are various platforms for creating Augmented Reality experiences, such as Wikitude Studio [9], Bear Go [10], PlugXR [11], etc.; but for our research, we used Vuforia Engine Developer Portal [12] for working with image-based triggers and Unity ver. 2021.3.1.1f1 [13]

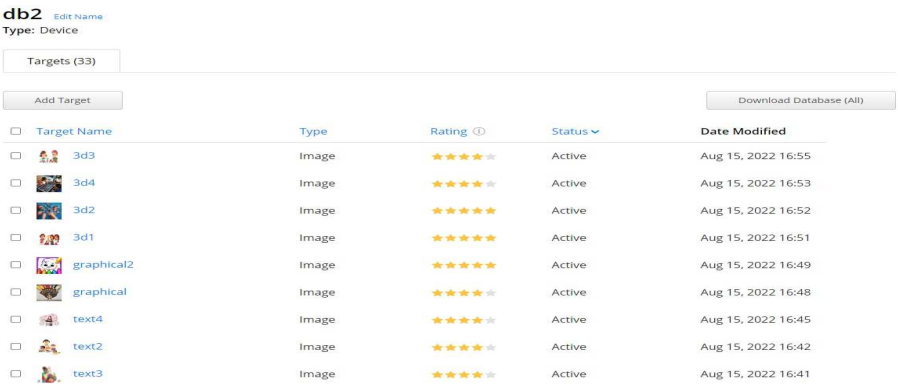


Figure 6. Markers database uploaded and evaluated by Vuforia Target Manager

platform for programming scenarios. In the previous section, we described our work with Vuforia Engine. In this section we will present a working process with Unity platform and a few results.

Using marker-based augmented artifacts, users can interact with the 3D information, objects, and events in a natural way. For example, in Fig. 7, there are presented features to change 3D object size, to rotate, and to change RGB color, or even changing color randomly by pressing a bigger green button.

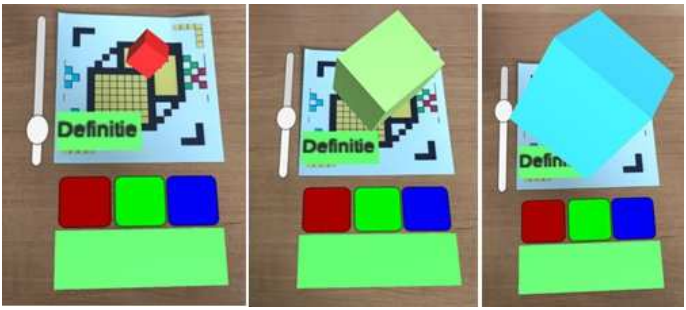


Figure 7. Example of interacting to 3D object

We created 30 types of AR artifacts. When scanned by mobile devices, markers trigger one of augmented experiences, such as 3D objects, video content, audio content, text, formulae, etc. From Bartle's classification, we realized 2 types of experiences for socializers and explorers. For killers and achievers they are still in process. Regarding Bloom's verb taxonomy, we used 30 words, one verb for each artifact (see Fig. 8) for artifacts classification.

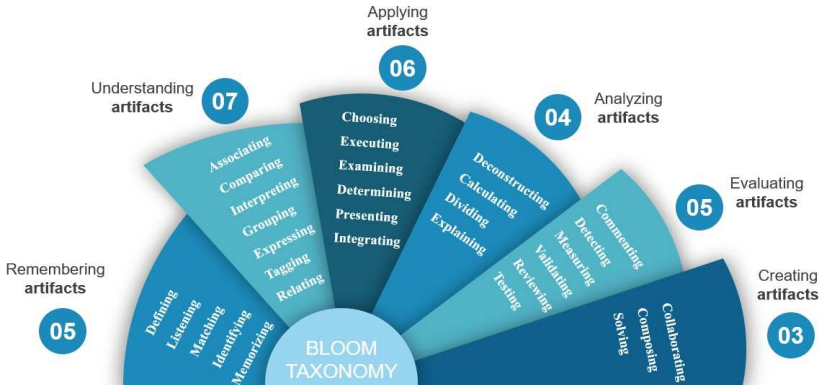


Figure 8. Classifying artifacts based on Bloom's verb used

When developing the scenarios (see Fig. 9) for the artifacts, learning styles based on sensory encoding methods was applied. This will deliver a positive impact by keeping pupils' high engagement and by enhancing their learning abilities like problem-solving, collaboration, imaginative thinking and spatial imagination.

Depending on the predominant sensory organ in receiving information and transmitting it to the brain, the performance obtained in the learning process will be higher.

4 Conclusion

In this article, the process of applying the marker-based Augmented Reality approach was described for the development of augmented artifacts based on learning preferences (learning styles) of the user, types

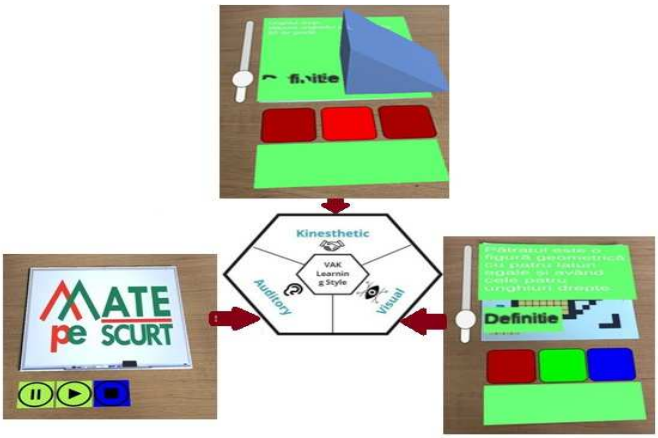


Figure 9. Developed scenarios based on learning style for Augmented artifacts

of students based on Bartle's characters theory, verbs from Blooms taxonomy for artifact classifications. For this, we used Vuforia Engine Developer Portal for working with image-based triggers and Unity platform for programming scenarios. In order to diversify the scenarios in the future work, the other approaches will be applied, such as the markerless one.

Acknowledgments. "Intelligent Information systems for solving ill structured problems, knowledge and Big Data processing" project Ref. Nr. 20.80009.5007.22, has supported part of the research for this paper.

References

[1] T. P. Caudell and D. Mizell. *Augmented reality: An application of heads-up display technology to manual manufacturing processes.* IEEE Xplore Conference: System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on., vol. 2, DOI:10.1109/HICSS.1992.183317, pp 659–669, 1992.

[2] W.B. Barbe, Swassing, and M.N. Milone. *Teaching through Modality Strenghts: Concepts and Practices*, Zaner-Bloser, Columbus, OH, 1979.

[3] D. Schafer and D. Kaufman. *Augmenting Reality with Intelligent Interfaces*, Artificial Intelligence – Emerging Trends and Applications, Chapter 11, pp. 221–242, 2018. DOI:10.5772/intechopen.75751.

[4] Yassir El Filali and Krit Salah-ddine. *Augmented reality types and popular use cases*, International Journal of Engineering, Science and Mathematics, vol. 8, no. 4, pp. 91–97, 2019.

[5] Osman Güler and Ibrahim Yucedag. *Developing an CNC lathe augmented reality application for industrial maintanance training*, Conference: 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1–6, 2018. DOI:10.1109/ISMSIT.2018.8567255.

[6] Veronica Teichrieb, Joao Paulo Silva do Monte Lima, Eduardo Lourenco Apolinario,Thiago Souto Maior Cordeiro de Farias Marcio Augusto Silva Bueno, Judith Kelner, and Ismael H. F. Santos. *A Survey of Online Monocular MarkerlessAugmented Reality*, International Journal of modeling and simulation for the petroleum industry, vol. 1, no. 1, pp.1–7, august, 2007.

[7] D. Kolb. *The Kolb Learning Style Inventory*. Version 3. Boston: Hay Group, 1999.

[8] S. Focsa-Semionov. *Invatarea autoreglata. Teorie. Strategii de învățare*, Editura Epigraf, pp.95–116, 2010.

[9] *Wikitude Studio for target images creation*. [Online]. Available: https://www.wikitude.com/products/studio/

[10] *ARGO Developer Portal*. [Online]. Available: https://developer.bear2b.com/

[11] *A Cloud-based Augmented Reality platform*. [Online]. Available: https://www.plugxr.com/

[12] *Vuforia engine, developer portal*. [Online]. Available: https://developer.vuforia.com/

[13] *Unity platform.* [Online]. Available: https://unity.com/

[14] *Vuforia Developer library.* [Online]. Available: https://library.vuforia.com/objects/best-practices-designing-and-developing-image-based-targets

[15] *Trekk a tech-driven creative services agency.* [Online]. Available: https://www.trekk.com/insights/five-types-augmented-reality-experiences-and-how-use-them-your-marketing

[16] *Grade 5 Curriculum.* [Online]. Available: https://www.mathsisfun.com/links/curriculum-year-5.html

[17] *Bloom's Taxonomy.* [Online]. Available: https://www.teachthought.com/learning/what-is-blooms-taxonomy/

[18] Richard Bartle. H*earts, Clubs, Diamonds, Spades: Players who suit muds.* [Online]. Available: https://mud.co.uk/richard/hcds.htm

Inga Titchiev[1,3], Olesea Caftanatov[1], Veronica Iamandi[1], Dan Talambuta[1], Daniela Caganovschi[2]

[1] Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: olesea.caftanatov@math.md
E–mail: inga.titchiev@math.md
E–mail: veronica.gisca@gmail.com
E–mail: dantalambuta@gmail.com

[2] State University of Moldova
E–mail: cda3721@gmail.com

[3] Tiraspol State University

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Development of a platform for heterogeneous document recognition using convergent technology

Alexandru Colesnicov, Ludmila Malahov,
Svetlana Cojocaru, Lyudmila Burtseva, Tudor Bumbu

Abstract

The paper describes architecture of a Web platform for recognition of heterogeneous documents. The platform provides comfortable working environment for users. The development of the platform uses the convergent technology that permits to integrate external applications easily, and maximally simplifies communication of the platform components.

Keywords: computer science, heterogeneous document recognition, framework architecture, integration of external applications.

1 Introduction

While digitizing texts, we often encounter documents with heterogeneous content. Each content type is recognized with its corresponding program. We need therefore a platform that combines all the necessary tools [1].

The proposed platform for digitization of heterogeneous documents supports all processing steps starting from electronic copies (scanned images). The platform supports image preprocessing, cutting of pages into homogeneous fragments, recognition of fragments, post-recognition processing, assembling of recognition results, and page reconstruction.

We discuss below architecture of the platform, the technology used at its development, and steps of digitization.

2 Platform architecture and development

The platform is a Web application consisting from user interface on the client side (frontend), and services implementation on the server side (backend). In our case, some functions are implemented on the client side that is described below.

The frontend is developing in Javascript using available standard packages and libraries (React, and more). Some necessary functionality is implemented in Javascript, for example, elementary image preprocessing. We use the corresponding packages, and implement some minor operations in the frontend instead of backend.

The backend is developing in Python using its rich libraries, and also calls external applications. The latter may be installed on the server, for example, ABBYY FineReader engine, and ImageMagick, or residing in the Web.

The convergent technology of the development [2] supposes smooth integration of the ready-made external applications. If the application has its documented API, this API is used. If the application doesn't have any API, it is executed in an isolated environment (sandbox). In both cases, exchange of data with the platform is performed through files.

The development is performed using Github version manager.

3 Digitization step-by-step

Steps of digitization supported by the platform are: uploading images and/or PDF files; image preprocessing; image fragmentation to parts with homogeneous content; recognition of fragments; post-recognition processing ot the results; assembling results; saving and downloading the results; restoring pages.

Comparison of restored and original pages is made manually but can be implemented lately.

Step 1. Uploading files. One or more files can be processed in a single digitization cycle. The following file types are supported:

PNG, JPEG, GIF, TIFF, and PDF. The total size and size of each file are restricted. File selection is performed through dialog, or by drag-and-drop.

2. Image preprocessing. This step is performed to obtain images with the quality suitable for recognition. Several preprocessing engines are available through submenu: Open CV, FineReader, ScanTailor, Gimp, ImageMagick. Open CV is available in Javascript and is executed on the client side. After selecting the engine, its specific options are offered.

Step 3. Fragmentation. This step permits to select image areas with homogeneous content, and detect the content type. The process is semi-automated. We use ABBYY FineReader engine on the server side to fragment images and preliminary detect fragment types. The fragmentation proposed by the program may be corrected manually. For the moment, the previewed types of content are: text, musical scores, mathematical formulae, chemical formulae and structures, chess diagrams. All unclassified content is marked as images. The platform is open for extensions; other content types may be added.

4. Recognition. Each type of the document content is recognized by its specific engine (FineReader, Mathpix, etc.) working on the server side, and the results are textual or script presentations of the content. It may be text in natural language for textual content, LaTeX script for mathematics, MusicXML for scores, etc.

5. Post-recognition processing. The obtained recognition results can be immediately checked and corrected. In some cases, transliteration of the text may be necessary, for example, from the old Romanian Cyrillic to the modern Latin alphabet. For the latter task, we implemented a dedicated application AAconv. The result of transliteration can be corrected manually.

6. Assembling recognition results. All original page images are collected in a PDF file. This is made even for pages of original PDF file because that file may be blocked for changes. The PDF standard permits to integrate inside a PDF other files as attachments. All results of recognition are attached for each page. Equally, page maps are

attached that contain coordinates of each page fragment.

7. Saving and downloading the results. The assembled file that integrates original images with recognition results and page maps is saved on the server side and can be downloaded by the user.

8. Restoring pages. Scripts permit restore graphical presentation of recognized fragments using applications that correspond to their content types.

4 Conclusion

The developing platform integrates in a unified comfortable environment all tools that are usually used at digitization of documents with heterogeneous content.

Acknowledgments. This work was prepared as part of the research project 20.80009.5007.22 "Intelligent information systems for solving ill-structured problems, processing knowledge and big data".

References

[1] A. Colesnicov, L. Malahov, S. Cojocaru, and L. Burtseva. *Semi-automated workow for recognition of printed documents with heterogeneous content.* Computer Science Journal of Moldova, vol. 28, Nr. 3, 2020, pp. 223–240.

[2] A. Colesnicov, L. Malahov, S. Cojocaru, and L. Burtseva. *On convergent technology in development of information systems for processing of documents with heterogeneous content.* In: Proceedings of the Workshop on Intelligent Information Systems, 04–05 December, 2020, Chișinău, Republic of Moldova, pp. 61–68, ISBN 978–9975–68–415–6.

Alexandru Colesnicov, Ludmila Malahov,
Svetlana Cojocaru, Lyudmila Burtseva, Tudor Bumbu

Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mails: alexandru.colesnicov@math.md, ludmila.malahov@math.md, svetlana.cojocaru@math.md, liudmila.burteva@math.md, tudor.bumbu@math.md

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Data in the technologies of modern society

Constantin Gaindric, Galina Magariu, Tatiana Verlan

1 Introduction

The world is going through the fourth industrial revolution, which is an amalgamation of physical, digital, and biological industrial technologies and is creating new opportunities and influences. Success depends primarily on a trained staff, on a coherent system of research and innovation, which will ensure technological and technical skills. No less important is an infrastructure that would ensure the collection, storage, processing, and security of information.

The problems to be solved are systemic and interdisciplinary, which means that the applied solutions must, in turn, be systemic and interdisciplinary. And the first issue, on which success will depend, is the citizen for reforms or reforms for the citizen.

At the Davos forum, among the most impressive achievements expected by 2025 were considered the following: 90% of the population with regular access to the Internet, the first robot with artificial intelligence (AI), a full member of the board of directors of some corporations, 30% of corporate audits will be carried out by AI, tax collection will be carried out through blockchain technologies.

The main problem lies not in these promising ideas, but in their implementation, which requires an information infrastructure, a system of professional education, and research institutions focused on societal problems. In the economy, the main factors of production were considered capital, labor, and raw material. The focus in the current economy, but especially in the future, is on data, knowledge, innovations, and technologies.

All information and analytical systems that are created to help specialists in various fields to make decisions use data. How correct or adequate the decision issued by a particular system will be, depends not only on the qualifications of the system developers and the quality of the resulting software product but to a large extent on the quality of the data used in the system.

And now, when the focus is put on artificial intelligence, and the expectations from an application of AI in many cases are too optimistic, the quality of the data and the trust in the source and the way of their processing become extremely pressing.

Therefore, the problem of data quality is becoming increasingly important in all areas of human activity, not only at the level of production or service units but also in the field of information technology.

2 Definitions for terms "Data" and "Information"

In the modern world including information technology, data is increasingly becoming a key component of the sequence of actions that leads to decision-making. Data is the starting point of the chain of notions *Data-Information-Knowledge*, without which any activity cannot be conceived, and philosophically it is also the foundation on which the *Data-Information-Knowledge-Wisdom* pyramid is built. The process of obtaining, processing, disseminating, and applying data has been critically accelerated; and their volume has also critically increased.

This article reflects some of the results of literature studies that address the problem of data quality and information quality. This problem has been studied for almost half a century. The authors of this article studied more than 30 articles from scientific and specialized literature on this topic written in different periods. It was interesting to analyze the approach that existed in the 20th century and the 21st and compare the changes if any. It should be noted that there is no consensus on the problem of data quality and information quality to this day. However, one can also observe general solutions both in terms

of time (approach to the problem in the past and today) and opinions of different researchers on this topic.

To begin with, it is necessary to define the terms "data" and "information". Very often these terms are used interchangeably, however, there are differences between them, which are explained in the scientific literature. For example, the Cambridge International Examinations manual [1] gives the following definitions:

Data *– a collection of text, numbers, or symbols in raw or unorganized form. They are with no meaning. Data, therefore, has to be processed or provided with a context, before it can have meaning.*

Information *– the result of processing data, usually by computer. This results in facts, which enables the processed data to be used in context and have meaning. Information is data that has meaning.*

Images and sounds can also be attributed to data, since they can be stored and used using a computer (played back and processed). In [2] *"Data are defined as simple facts, either quantitative or qualitative. Information is defined as organized data."*

Article [3] provides the following comment: *"Most definitions refer to a datum as the most basic descriptive element representing a perception or measurement about some object of interest. By itself, a datum's value typically lacks content, meaning, or intent. Information is more than just a set of data; it is the output of a process that interprets and manipulates data into some prescribed format."*

For practical use, it is useful to identify a classification of data into structured and unstructured data. Structured data is that which can be stored in a table, where each object has the same structure (i.e., a set of attributes). Structured data can be easily stored, searched, reordered, and combined with other structured data.

We encounter unstructured data more often than structured data. Unstructured one describes data, where each object in a set can have its own internal structure that is usually different for each object. Differences in structure between individual elements prevent the analysis of unstructured data in its raw form.

We can often extract structured data from unstructured data using artificial intelligence techniques. But the process of transforming unstructured data is expensive and generates significant expenses.

There are two main types of raw data in terms of how they are obtained: collected data and resulting data. The data collected is obtained by direct measurement or observation designed for this purpose. And the resulting data is a product of a process whose purpose is the explanation of the raw data.

One of the most well-known types of resulting data is metadata, that is, data that describes other data.

To better understand the terms "data" and "information" in the field of computer science, some authors explain[1] that data is input, *"or what you tell the computer to do or save"*. But, information can be regarded as output, *"or how the computer interprets your data and shows you the requested action or directive"*.

When describing the key difference between "data" and "information", the author of [6], among several points, gives the following:

- *"Data never depends on Information while Information is dependent on Data";*
- *"Data can be structured, tabular data, graph, data tree whereas Information is language, ideas, and thoughts based on the given data."*

3 About Data quality, Information quality, and criteria for them

The concepts of "data" and "information" are closely intertwined and, in a sense, interrelated (though only in one direction). *"When data are intelligently organized they convey information, and what information is conveyed depends upon just how the data are organized"* [2].

[1] https://examples.yourdictionary.com/difference-between-data-and-information-explained.html

This relationship affects the decision-making process by analysts since it implies not only the competent organization of data but also their quality. If you have bad data, then you will have bad information, which will lead to making wrong decisions, which, in turn, can lead to significant losses in production, business, finances, misdiagnosis, etc. The issue of data quality (DQ) is, to some extent, subjective: data that one user considers to be of high quality may be of low quality to another user. In addition, data that is considered high-quality today will be low-quality data tomorrow. Also, the opposite is possible: data that is considered low-quality data today, tomorrow we can learn how to process it better, and it will become high-quality data.

3.1 Data quality and criteria

There are many approaches to measuring the quality of data and quality of information and defining objective criteria – dimensions. The most popular approach is to answer two questions:

- How suitable is the data to be used for these purposes?
- How well do the data reflect the real situation they describe?

"To increase the trust in data-driven decisions, it is necessary to measure and to know the quality of the employed data with appropriate tools" [4].

Many studies provide some objective criteria for assessing data quality. However, there is still no consensus in the scientific literature on the issue of significant criteria for data quality. In a simplified sense, the quality of data is the degree of their suitability for use. So, for example, in the review [4], the following comment is given:

"DQ is most often associated with the "fitness for use" principle, which refers to the subjectivity and context-dependency of this topic. Data quality is typically referred to as a multi-dimensional concept, where single aspects are described by DQ dimensions (. . .) Our evaluation framework covers the four most frequently used dimensions, which are (. . .) accuracy, completeness, consistency, and timeliness."

ISO 9000:2015 defines the quality of data by the degree to which they meet requirements: needs or expectations, such as completeness, reliability, accuracy, consistency, availability, and timeliness.

Each of these criteria is defined differently in the literature and in different fields of application.

Completeness *"is very generally described as the 'breadth, depth, and scope of information contained in the data' and covers the condition for data to exist to be complete."* [4] Data completeness tells you whether your datasets contain everything you need for your research or management needs.

Reliability is the ability to trust in data used in the organization. The reliability of the data refers to their completeness and accuracy, as well as how consistently the measurements were made, and whether the same results are obtained with the same measurement under the same conditions several times, or by different people.

Accuracy *"can be described as the closeness between an information system and the part of the real world it is supposed to model".* [4] Accuracy means the usage of data that conforms the reality. Moreover, it is very important that the data values are correct not only in their value but in their form too; they should *"be represented in a consistent and unambiguous form"* [2].

Consistency *"captures the violation of semantic rules defined over data items, where items can be tuples of relational tables or records in a file."* [4] As WIKI tells, *"Data consistency refers to whether the same data kept at different places do or do not match."* Also, T.S. Adams in his article gives a very clear and explicative definition[3]: *"Data consistency is the process of keeping information uniform as it moves across a network and between various applications on a computer. There are typically three*

[2]http://etutorials.org/Misc/data+quality/Part+I+Understanding+Data+Accuracy/Chapter+2+Definition+of+Accurate+Data/2.3+Data+Accuracy+Defined/

[3]https://www.easytechjunkie.com/what-is-data-consistency.htm

types of data consistency: point in time consistency, transaction consistency, and application consistency. Ensuring that a computer network has all three elements of data consistency covered is the best way to ensure that data is not lost or corrupted as it travels throughout the system. In the absence of data consistency, there are no guarantees that any piece of information on the system is uniform across the breadth of the computer network."

Availability of data – the term that usually is used in the framework of some organization, institution, or company and means that all data related to this organization are available to its internal goals or its partners at any time 24/7/365. It is essential for the uninterrupted and stable work of this organization and its management without faults.

Timeliness describes *"'how current the data are for the task at hand' and is closely connected to the notions of currency (update frequency of data) and volatility (how fast data becomes irrelevant)"*. [4] Timeliness of the data means that the whole chain "data collecting-transfer-processing-presentation" is run in real-time.

3.2 Information quality and criteria

Similarly, for the question "What kind of information should be considered qualitative" – in different areas of information use, the answer to this question may be different. However, it is important to understand how information is perceived and used by its consumer. There are two necessary steps here [5]:

1. *Highlighting which attributes are important.*
2. *Determining how these attributes affect the customers in question.*

In the specialized literature, 10 attributes of information quality are most frequently used, which can affect the effectiveness of information systems and can contribute to the development of strategies to improve

the quality of information. We will highlight their definitions as follows (see [5], except for "Relevance" with the definition formulated by the authors of this article):

Relevance – how well it reflects the user's needs. If it does not reflect his needs, it still does not mean that the information is bad; it may be good for another class of users.

Accuracy – Accurate information reflects the underlying reality. Less well understood is that information can be too accurate when its degree of precision exceeds its customer's processing capability. This increases the cost of information systems.

Timeliness – The concept of what is timely is itself constantly changing and being redefined, because of changes in customer perceptions caused by technology and the competitive environment.

Completeness – Incomplete information can lead customers astray. However, complete information for one person may be incomplete for another. Information may also be too complete. The danger in business lies in information systems that generate so much information that customers cannot process it all in a timely fashion.

Coherence – is how well the information hangs together and is consistent with itself.

Format – refers to how the information is presented to the customer.

Accessibility – information that can be obtained when needed. It depends on the customer and even on the specific circumstances for that customer. For information quality to occur, timeliness and accessibility should complement each other.

Compatibility – Information quality lies not only in the quality of the information itself, but also in how it can be combined with other information and delivered to a customer.

Security – Two aspects of information security are protecting information from people (logical security) and protecting information from natural disasters (disaster recovery planning).

Validity – Information has validity when it can be verified as being true and satisfying appropriate standards related to' other dimensions such as accuracy, timeliness, completeness, and security.

This set of information quality attributes can change, even within the same domain of use.

Therefore, at each current moment, the user must answer the following questions [5]:

- *First, are yesterday's perceptions of quality needs still valid?*
- *Second, how do quality needs translate into technology requirements?*
- *Third, do internal information collection, dissemination, and verification procedures measure up to quality requirements?*

4 Conclusion

To confidently use data in decision-making in any field, we need not limit ourselves to the attributes described but also need confidence in: the reliability and identification of the supply chain and other attributes of the origin of the data; the recipient's assessment of the public data provider's business qualities as responsible and relatively independent.

Recently, the penetration of AI into all fields is increasingly discussed, but practically very little attention is drawn to the fact that the conclusions and recommendations proposed are based on the data collections used, the quality of which is not evaluated either in terms of veracity or about the sources from where they are obtained.

This gave us an impetus to present some considerations on the issue of data quality and information produced from data processing.

The quantitative evaluation of the whole set (or a smaller number) of significant indicators for a decision-making process allows us to know the degree of suitability for the proposed purpose and to be sure that it will be the expected one.

Acknowledgments. "Intelligent Information systems for solving ill structured problems, knowledge and Big Data processing project"

Ref. Nr. 20.80009.5007.22, has supported part of the research for this paper.

References

[1] *Topic support guide. Cambridge International AS & A Level, Information Technology 9626, For examination from 2017. Topic 1.1 Data, information and knowledge*, Cambridge International Examinations 2015, Version 1, https://www.cambridgeinternational.org/images/285017-data-information-and-knowledge.pdf.

[2] Sue P. Stafford. *Data, Information, Knowledge, and Wisdom*, in: Knowledge management, organizational intelligence and learning, and complexity, vol. III, Encyclopedia of Life Support Systems (EOLSS), 2011.

[3] Tuomo Uotila and Helina¨ Melkas. *Quality of data, information and knowledge in regional foresight processes*, Futures 39, 2007, pp. 1117–1130.

[4] Lisa Ehrlinger, Elisa Rusz, and Wolfram Wöß. *A survey of data quality measurement and monitoring tools*, Preprint, arXiv:1907.08138v1 [cs.DB], 18 Jul 2019.

[5] Holmes E. Miller. *The multiple dimensions of information quality*, Information Systems Management, vol. 13, no. 2, pp.79–82. DOI: 10.1080/10580539608906992.

[6] David Taylor. *Difference between Information and Data*, Updated July 9, 2022, Available: https://www.guru99.com/difference-information-data.html

Constantin Gaindric[1], Galina Magariu[2], Tatiana Verlan[3]

Vladimir Andrunachievici Institute of Mathematics and Computer Science
5, Academiei street, Chisinau, Republic of Moldova, MD 2028
E–mails: [1]constantin.gaindric@math.md
[2]galina.magariu@math.md
[3]tatiana.verlan@math.md

Application of Predictive Methods in the Analysis of Data Sets Regarding the Pandemic Caused by COVID 19

Adela Gorea, Alexandr Carpenco

Abstract

The purpose of this research is to investigate the relationship between the number of cases of Covid-19 and to develop an automatic learning model for predicting the number of illnesses for 2022. In the first part of the research, there was analyzed the situation in several countries that were affected by the pandemic. Then we analyzed the elaboration and training of the model of automatic learning for prediction[1].

Keywords: Regression analysis models, Machine learning, Deep learning, COVID-19.

1. Introduction

In the winter months of 2019-2020, another type of coronavirus, Covid-19, was detected in Wuhan [1]. This epidemic has spread worldwide and that led to an increase in everyday cases and the appearance of new forms of Covid-19. There are many techniques that analyze natural phenomena, including artificial intelligence, mathematical and statistical methods, such as optimization, analysis of time series, automatic learning, regression modeling, clustering, and numerical analysis [2], [3]. As Covid-19 has many impacts on the environment, health, society, and economy, the study of the spread rate of this disease and comparison of its rate in different countries is quite essential. There are some researches on the classification of Covid-19 data sets [4], [5], [6]. These researches are

[1] The work is done within the project 20.80009.5007.22, Intelligent information systems for solving ill-structured problems, processing knowledge and big data.

based on the analysis of the time series and the analysis of the main components.

2. Predictive modeling and statistical learning

Predictive modeling is an approach that focuses on the development of models to get the most accurate predictions on a data set [7]. It has associations with model recognition, Data Mining, and ML, which represent the full predictive modeling process. However, it easily ignores the broad scope of ML (establishing computer-based learning tools) and statistics (data understanding) and is concerned about getting the minimum error with a certain degree of modeling the model skills. On the other hand, the statistical equivalent of predictive modeling is called statistical learning, where the emphasis is put on the interpretability of the model and not on the behavior of the model skills. Statistical learning is a sub-domain of applied statistics in which the emphasis is placed on the application of statistical instruments to model and understand complex data sets [8].

3. General structure

The overall workflow of this study is summarized in the following steps: (1) data collection of COVID-19 and other pneumonia data; (2) statistical analysis and development, training, and evaluation of ML and deep learning models; (3) result analysis including the model evaluation and feature interpretation in Figure 1.

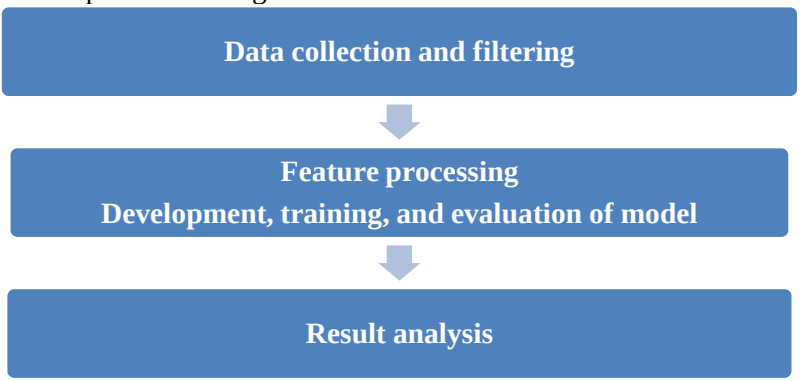


Figure 1. The overall workflow of this study

4. Collection and Analysis of dataset of COVID-19

The received data set is very sensitive to noise, lack, and inconsistency of data. In order to obtain useful results from the data, the following pre-processing techniques are used: (1) Data cleaning – to eliminate noise and correction of inconsistencies from the data; (2) Data aggregation – to merge data from multiple sources in an appropriate form, to combine data from several sources in a coherent data deposit: data warehouse or data cube; (3) selection of data – to select the information relevant to the current analysis process in order to simplify work in the actual stage of knowledge extraction; (4) Data transformation – to prepare data for analysis through the most appropriate representation. It consists of operations such as the normalization of data. Normalization can improve the accuracy and efficiency of the Data Mining Algorithms. Normalization of data represents escalating of all data in a predetermined field; (5) Data reduction – to reduce the size of data by aggregation, eliminating redundant characteristic features, or grouping the common features.

These techniques are applied a priori to the date of the date and can improve the quality of the process to find the templates and/or reduce the time required for the actual mining.

Data from the following countries were selected for analysis: Wuhan (source of the pandemic), Italy (heavily affected country), Romania (neighboring state), and Moldova. The datasets were taken from official sources such as WHO[2], Our World in Data[3], and ECDC[4] (European Centre for Disease Prevention and Control) in the period from 22.01.2020 to 07.05.2022.

To analyze the data, mean (along with standard deviation) or median (along with interquartile range) values were calculated as numeric measurements. The algorithms KNeighborsRegressor(), DecisionTreeRegressor(), ExtraTreeRegressor(), SVR(), LinearRegression() were used to examine if the sample data were

[2] https://www.who.int
[3] https://ourworldindata.org/coronavirus-source-data
[4] https://www.ecdc.europa.eu/en/covid-19/data

normally distributed. Additional statistical analysis methods including the T-test and Kruskal Wallis rank test were utilized to summarize and analyze COVID-19 patients. A p-value less than 0.05 was considered statistically significant. All these algorithms are included in various Python language libraries.

Projection of all 4 graphs in the period 30.01.2020-21.03.2020 is presented in Figure 2.

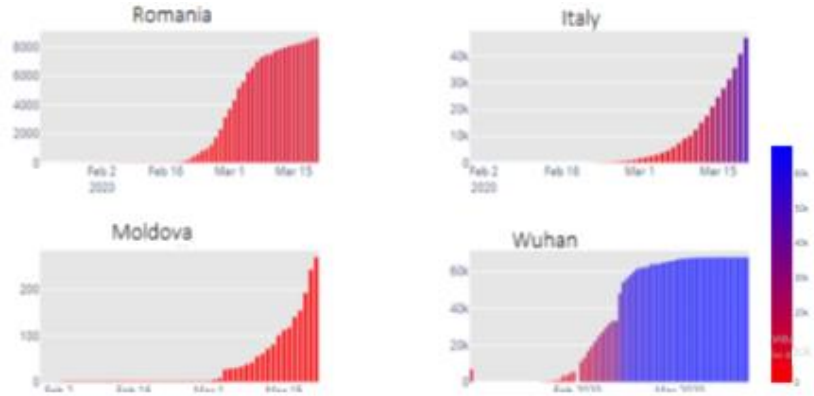


Figure 2. Cases of infection in 4 countries: Wuan, Italy, Romania, and Moldova

5. Prediction of COVID-19 for 2022

For prediction, there is used a new dataset for 2022. It contains statistics on the COVID-19 pandemic in the world: (1) statistics of cases of illness, recovery, and death from COVID-19; (2) statistics on the availability of vaccines and general vaccination of the population; (3) the number of received done tests and their effectiveness; (4) information about diseases associated with complications of COVID-19; (5) statistics on the smoking population.

The final result of the forecast is shown in the segment “Trend”, “Week”, and “Year”. Similar actions are repeated for the other two indicators. Thus, a forecast for possible outcomes is obtained based on the analyzed data from statistics. The forecast is displayed in the form of curves and in three segments of presentations in Figure 3.

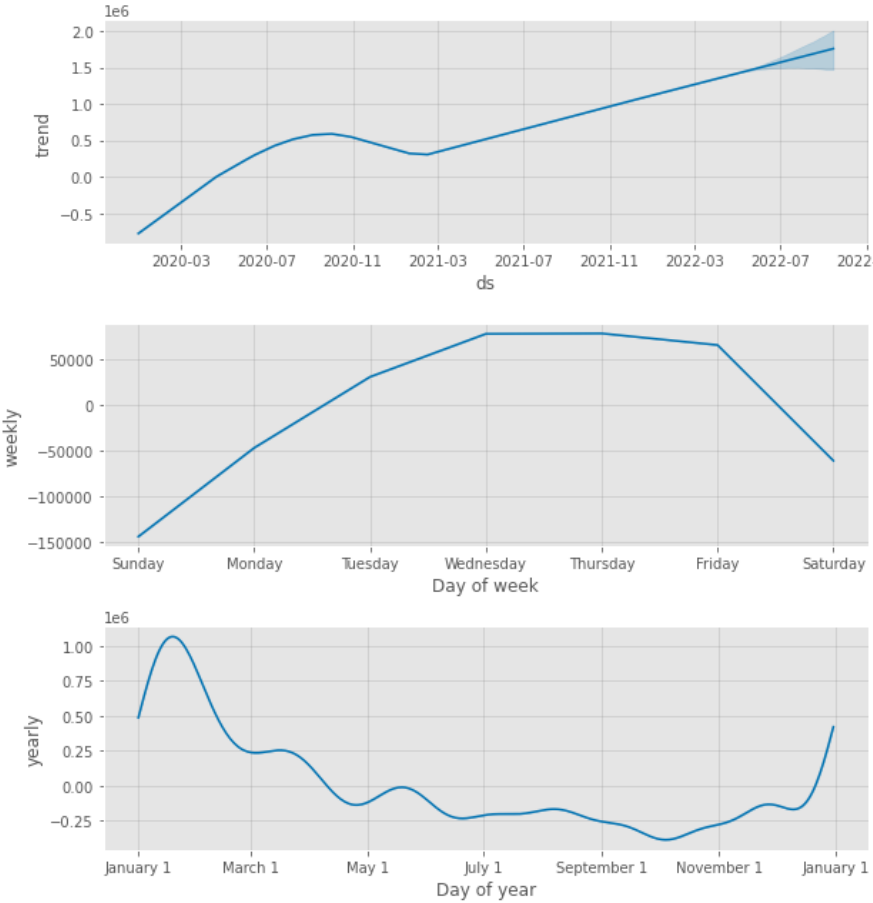


Figure 3. Result of the forecast, in the segments “Trend”, “Week”, “Year”

Any time can be indicated, depending on the requirements for the forecast.

5. Conclusion

The purpose of the article was to demonstrate the situation with the pandemic Covid-19. This task was successfully solved by using various big datasets. The data were taken from official sources such as WHO,

OUR World in Data, and ECDC. As the experiment was completed, data on the situation with the pandemic in Moldova were analyzed and visualized.

There was given the statistics of diseases. A comparison of the situation in Moldova with other countries such as China, Italy, and Romania was carried out.

The situation with the pandemic Covid-19 was demonstrated both in the world and in Moldova with the help of various graphs. As well it was possible to predict a possible outcome with the help of forecasting in the near future which in real life allows the relevant authorities to take the necessary measures in advance, with the help of various graphs, the situation with the pandemia Covid-19 was visualized both in the world and in Moldova. And with the help of forecasting, it was possible to predict a possible outcome in the near future, which in real life allows the relevant authorities to take special measures in advance.

References

[1] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu *et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding*. The Lancet, 395 (10224) (2020), pp. 565-574.

[2] M. R. Mahmoudi. *On comparing two dependent linear and nonlinear regression models*. J Test Eval, 47 (1) (2018), pp. 449-458.

[3] P. Ji-jun, M.R. Mahmoudi, D. Baleanu, and M. Maleki. *On comparing and classifying several independent linear and non-linear regression models with symmetric errors*. Symmetry, 11 (6), (2019), pp. 820.

[4] M. Maleki, M. R. Mahmoudi, M. H. Heydari, and K. H. Pho. *Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models Chaos Solitons Fractals*. (2020), pp. 140.

[5] M. Maleki, M. R. Mahmoudi, D. Wraith, and K. H. Pho. *Time series modelling to forecast the confirmed and recovered cases of COVID-19*. Travel Med Infect Dis, (2020), 101742.

[6] M. R. Mahmoudi, M. H. Heydari, S. N. Qasem, A. Mosavi, and S. S. Band. *Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries*. Alexandria Eng J, 60(1) (2021), pp. 457-464.

[7] M. Kuhn and K. JOHNSON. *Applied predictive modeling*. (Vol. 26), Springer, 2013.

[8] James Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An introduction to statistical learning: with applications in R*. Springer, 2013.

Adela Gorea, Alexandr Carpenco

Alecu Russo Balti State University
E-mail: adela.gorea@usarb.md

Alecu Russo Balti State University
E-mail: karpenko.aleksandr.96@mail.ru

Data parallelization on HPC systems for modeling decision-making problems

Boris Hâncu

Abstract

In the paper, a theoretical and practical study of the methods for data parallelization on HPC systems and a large class of decision-making problems in conditions of risk and uncertainty, based on game theory are presented. We will present a parallel algorithm for solving bimatrix games divided into blocks of submatrices using the 2D-cyclic algorithm. The proved theorems represent the basis on which the solution of the bimatrix game can be constructed using the solutions of the subgames generated by the 2D cyclic matrix distribution algorithm.

Keywords: computer science, information technologies, Cloud HPC, mathematical modeling, parallel algorithm, games.

1. Introduction

Decision-making is the process of making choices by identifying a decision, gathering information, and assessing alternative resolutions. Using a step-by-step decision-making process can help you make more deliberate, thoughtful decisions by organizing relevant information and defining alternatives.

The decision-making HPC support use cases will involve ensembles of jobs, data-staging to support workflows, and interactions with services/facilities external to HPC systems/centers. For effective HPC decision-making, we need flexible and interactive access to HPC resources, particularly in the context of dynamic workflows processing large datasets. This poses several technical and organizational challenges: short-notice secure access to HPC and data resources, dynamic resource allocation and scheduling, coordination of resource managers, support for data-intensive workflow (including data staging on node-local storage),

preemption of already running workloads, and interactive steering of simulations. Federation of services and resources across multiple sites will help to increase availability, provide elasticity for time-varying resource needs and enable leverage of data locality. In the Big Data domain, a huge amount of data is collected from numerous sources and the use of these data is crucial in an HPC decision-making environment. HPC systems are usually unique environments with specific interconnects and accelerators, as well as software sets customized to particular domains and use cases.

Running an optimized code requires machine-specific porting. Deployment can be expedited with the use of (hardware platform-specific) Cloud and container technologies that facilitate the handling of software dependencies and portability.

Decision-making processes are very complex and for the mathematical modelling of these problems, it is necessary to take into account the big data problems. Usually, data is too big to be stored and processed by a single machine and so data is divided into partitions that can be managed and accessed separately. In order to solve such problems in real time, parallel algorithms are built and then implemented on various types of HPC systems. For parallel data processing, we must use the ways of dividing, partitioning (sharing), and distributing data.

Largely, the structure of the parallel algorithm for models of this type is determined by the parallelization mode at the data level.

It is known that the basic parallel strategy consists of three main steps: the **first step** is to partition the input into several partitions of almost equal sizes, and so, the parallelization at the level of data and operations is actually achieved; the **second step** is to solve recursively the subproblem defined by each partition of the input, and so, the distribution and solution of subproblems on a parallel computing system is actually performed; the **third step** is to combine or merge the solutions of the different subproblems into a solution for the overall problem. The success of such a strategy depends on whether or not we can perform the first and third steps efficiently [1].

At the moment, in the opinion of the author, there are no algorithms that use the partitioning of matrices based on the 2D block-cyclic distribution algorithm for solving bimatrix games with very large matrices. In this article, the authors make an attempt to complete this gap.

So we will analyze the ways to build on HPC clusters parallel algorithms, and especially, data parallelization, for a class of noncooperative games and bimatrix games.

2. Generation of sets of bimatrix subgames as a result of the division into blocks of global matrices.

As an example of a decision problem, we will consider a mathematical model determined by a bimatrix game Γ in pure strategies, that is uniquely determined by the matrices $A = \|a_{ij}\|_{i\in I}^{j\in J}$, $B = \|b_{ij}\|_{i\in I}^{j\in J}$ that are the payoff matrices of player *1* and player *2*, respectively. Here *I={1,2,..,n}* is the line index set, *J={1,2,..,m}* is the column index set of the matrices. As the solution of this model, we will consider the set of all equilibrium profiles $NE(\Gamma)$.

The following sequential algorithm can be formulated to determine the Nash equilibrium profiles in pure strategies to construct $NE(\Gamma)$.

Basic Sequential Algorithm (BSA)

1. For any fixed column $j \in J$, the set $Br_2(j) = Argmax_{i\epsilon I} a_{ij}$ is determined. Under the algorithmic aspect, it can be as follows: for any column j of the matrix A all maximum elements of this column are highlighted.
2. For any fixed row $i \in I$, the set $Br_2(i) = Argmax_{j\epsilon J}$ is determined. Under the algorithmic aspect, it can be as follows: for any row i of the matrix B all maximum elements on this row are highlighted.
3. The graph $GrBr_1$ of the multilevel application Br_1 from step 1) and also the graph $GrBr_2$ of the multilevel application Br_2 from step 2) are built. The equilibrium profiles are all the profiles belonging to the intersection of the two given graphs: $NE = GrBr_1 \cap GrBr_2$. From an algorithmic point of view, it can be done as follows: we look for all highlighted elements in the matrices A and B, and the indices of the elements whose positions coincide both in matrix A and in matrix B will be the equilibrium profiles.

Suppose that matrices A and B are divided into submatrices (at the moment we do not analyze the case of distribution of submatrices). We

can consider the following cases: a) when both matrix A and matrix B are divided into submatrices using one and the same algorithm, and b) when both matrix A and matrix B are divided into submatrices using different algorithms. Thus, we can obtain a series of pairs of submatrices of the same size $\{(A_r, B_r)\}_{r=\overline{1,p}}$, where $A_r = \left\|a_{i_r j_r}^r\right\|_{i_r \in I_r}^{j_r \in J_r}$ and $B_r = \left\|b_{i_r j_r}^r\right\|_{i_r \in I_r}^{j_r \in J_r}$. We mention that we will analyze only the case when the submatrices, obtained as a result of the algorithms for dividing the matrices A and B, have the same dimension, and therefore bimatrix subgames can be constructed. Here the index r actually means "processor" which, as will be seen below, will obtain these submatrices. These submatrices will generate a series of games that are actually sub-games of the original game $\Gamma_r = \langle I_r, J_r, A_r, B_r \rangle$. We denote by $NE[\Gamma_r]$ the set of Nash equilibrium profiles in the problem Γ_r.

We introduce the following applications which determine the correspondence between the "local indices" of the elements of the local matrices A_r and B_r and the "global indices" of the elements of the global matrices A and B, specifically $\phi_r: I_r \to I$, $\psi_r: J_r \to J$. These functions are in fact determined by the algorithm of dividing the matrices A and B into submatrices. Obviously, these functions must verify the following conditions:

$$\forall i \in I, \exists r \text{ and } i_r \in I_r \text{ that } i = \phi_r(i_r), \quad (1)$$

$$\forall j \in J, \exists r \text{ and } j_r \in J_r \text{ that } j = \psi_r(j_r). \quad (2)$$

According to (1)-(2), the following condition is verified: for any strategy profile in the bimatrix games Γ, namely $(i, j) \in I \times J$, there exists a strategy profile (i_r, j_r) in the bimatrix games Γ_r, so that $i = \phi_r(i_r)$ and $j = \psi_r(j_r)$. As a result, we analyze only those divisions and distributions of the global matrices in local matrices for which there exist the applications ϕ_r and ψ_r such that conditions (1)-(2) are satisfied.

We formulate the following main problems.

Problem (**existence of solutions**). What are the conditions (necessary, sufficient, necessary and sufficient) for the equilibrium profiles in the game Γ_r to be the equilibrium situation in the game Γ. Here we can highlight the following two aspects:

1. which $(i_r^*, j_r^*) \in NE[\Gamma_r]$ there will also be $(\phi_r(i_r^*), \psi_r(j_r^*)) \in NE[\Gamma]$;

2. what properties must the algorithm of dividing the matrices into submatrices possess in order to ensure the following: if $(\mathrm{i}_\mathrm{r}^*, \mathrm{j}_\mathrm{r}^*) \in \mathrm{NE}[\Gamma_\mathrm{r}]$ then $\left(\phi_\mathrm{r}(\mathrm{i}_\mathrm{r}^*), \psi_\mathrm{r}(\mathrm{j}_\mathrm{r}^*)\right) \in \mathrm{NE}[\Gamma]$.

3. About a matrix division algorithm for which any solution of the subgames is also the solution of the initial game

We will first analyze those algorithms for constructing subgames based on the division of matrices for which the solutions of the subgames will be the solutions of the initial game. We will introduce the following definition.

Definition *If for any equilibrium profile* $(i_r^*, j_r^*) \in NE[\Gamma_r]$ *in the game* Γ_r *exists* $\phi_r: I_r \to I$, $\psi_r: J_r \to J$ *that is* $(\phi_r(i_r^*), \psi_r(j_r^*)) \in NE[\Gamma]$, *then we say that the algorithm of dividing the matrices into blocks of submatrices is perfect and is called "**p**erfect **m**atrix **d**ividing and **d**istribution **a**lgorithm" (PMDDA)*

To realise the data parallelization, we use the **two-dimensional block-cyclic data layout scheme** [2]. The P processes of an abstract parallel computer are often represented as a one-dimensional linear array of processes labelled $0,1,\dots,P$. It is often more convenient to map this one-dimensional array of processes into a two-dimensional rectangular grid, or process grid by using row-major order (the numbering of the processes increases sequentially across each row) or by using column-major order (the numbering of the processes proceeds down each column of the process grid). This grid will have l_{max} process rows (lines) and c_{max} process columns, where $lmax_{max}$. The process can now be referenced by its row and column coordinates, (l, c), within the grid $L \times C$, where $L = \{1,..,l,\dots,l_{max}\}$ is a set of row numbers and $C = \{1,..,c,\dots,c_{max}\}$ is a set of column numbers. These groupings of processes are of particular interest to the programmer since distributed data decomposition of a matrix tends to follow this process mapping. Viewing the rows/columns of the process grid as essentially autonomous subsystems provides the programmer with additional levels of parallelism.

For dense matrix computations, we assume the data to be distributed according to the two-dimensional block-cyclic data layout scheme. The block-cyclic data layout has been selected for the dense algorithms implemented in DMM parallel systems principally because of its

scalability, load balance, and **efficient** use of computation routines (data locality). The block-partitioned computations are processed in consecutive order just like a conventional serial algorithm.

According to the **2D b**lock-**c**yclic **m**atrix **d**ividing and **d**istribution (**2DBCMD&D**), an m by n dense matrix is first decomposed into m_A by n_A blocks starting at its upper left corner. These blocks are then uniformly distributed in each dimension of the Process Grid. Thus, every process owns a collection of blocks, which are locally and contiguously stored in a two-dimensional column-major array. We present below some examples of the distributed matrix based on **2DBCMD&D** algorithm.

Let $I_{(l,c)}$ (respectively $J_{(l,c)}$) denote the rows (respectively columns) of the local matrices. Denote by $A_{(c,l)} = \left\| a_{i_{(l,c)} j_{(l,c)}} \right\|_{i_{(l,c)}=\overline{1,|I_{(l,c)}|}}^{j_{(l,c)}=\overline{1,|J_{(l,c)}|}}$ and $B_{(c,l)} = \left\| b_{i_{(l,c)} j_{(l,c)}} \right\|_{i_{(l,c)}=\overline{1,|I_{(l,c)}|}}^{j_{(l,c)}=\overline{1,|J_{(l,c)}|}}$ submatrices formed from global matrices A and B that are distributed to the $(c,l) \in L \times C$ process grid. We call $i_{(l,c)}$ and $j_{(l,c)}$ local indices of the local elements of the local matrices $A_{(c,l)}$, $B_{(c,l)}$. According to (1)-(2), for any strategy profile $(i,j) \in I \times J$, there exist a process $(l,c) \in L \times C$ and the strategy profile $(i_{(l,c)}, j_{(l,c)})$ that $i = \phi_{(l,c)}(i_{(l,c)})$ and $j = \psi_{(l,c)}(j_{(l,c)})$.

It's easy to prove that for the **2DBCMD&D** algorithm, the $\phi_{(l,c)}$ and $\psi_{(l,c)}$ functions also verify the following conditions.

Proposition 1 *For the 2DBCMD&D algorithm, the* $\phi_{(l,c)}: I_{(l,c)} \to I$, $\psi_{(l,c)}: J_{(l,c)} \to J$ *functions verify the following conditions:*

a) for all fixed $l = \overline{1, l}_{max}$, *we have* $\phi_{(l,\hat{c})}(i_{(l,\hat{c})}) = \phi_{(l,\tilde{c})}(i_{(l,\tilde{c})})$ *for all* $\hat{c} \neq \tilde{c}$ *and* $i_{(l,\hat{c})} = i_{(l,\tilde{c})}$.

b) for all fixed $c = \overline{1, c}_{max}$, *we have* $\psi_{(\bar{l},c)}\left(j_{(\bar{l},c)}\right) = \psi_{(\tilde{l},c)}(j_{(\tilde{l},c)})$ *for all* $\bar{l} \neq \tilde{l}$ *and* $j_{(\bar{l},c)} = j_{(\tilde{l},c)}$.

Condition a) means the following: processes that are located on the same line of the process grid (for example line l) in its submatrices contain elements of the same line of the global matrix. Respectively, condition b) means the following: the processes which are located in the same column of the process grid (for example, column c) in its

submatrices, contain elements of the same column of the global matrix. This property is used to construct the equilibrium profiles in the bimatrix games.

4. Nash equilibrium profiles for bimatrix games with block-cyclic distributed matrices

Using the BSA algorithm, you can easily build the following parallel algorithm that can already be executed on mixed HPC systems (with distributed and shared memory).

Basic Best-response Parallel Algorithm (BBrPA)

1. *Data parallelization:* using some algorithm for dividing and distributing the matrices, each process $r \in R$ "gets" the matrix pair $A_r = \left\| a^r_{i_r j_r} \right\|^{j_r \in J_r}_{i_r \in I_r}, B_r = \left\| b^r_{i_r j_r} \right\|^{j_r \in J_r}_{i_r \in I_r}$.

2. Each $r \in R$ process independently using the BPA will build the following point-to-set applications $Br^r_1(j_r) = Arg\max\limits_{i_r \in Ir} a^r_{i_r j_r}$ for all $j_r \in J_r$, $Br^r_2(i_r) = Arg\max\limits_{j_r \in J_r} b^r_{i_r j_r}$ - for all $i_r \in I_r$ and the set of index pairs $GrBr^r_1 \cap GrBr^r_2$. is built. If, as a result of the division of matrices *A* and *B*, subgames $\Gamma_r = \langle I_r, J_r, A_r, B_r \rangle$ are generated, (the necessary condition is that the submatrices A_r and B_r are of the same size), then $NE[\Gamma_r] = GrBr^r_1 \cap GrBr^r_2$. After that, each process sends to the root process the set $GrBr^r_1 \cap GrBr^r_2$.
3. The root process from the sets of index pairs $\{GrBr^r_1 \cap GrBr^r_2\}_{r \in R}$ builds the set *NE[Γ].* Using the sets $\{GrBr^r_1\}_{r \in R}$, $\{GrBr^r_2\}_{r \in R}$, the graph $GrBr_1$ of the point-to-set application Br_1 and also the graph $GrBr_2$ of the point-to-set application Br_2 has built. The equilibrium profiles are all the profiles belonging to the intersection of the two given graphs: $NE = GrBr_1 \cap GrBr_2$.

We denote by $NE\left[\Gamma_{(c,l)}\right]$ or $NE\left[\left(A_{(l,c)}, B_{(l,c)}\right)\right]$ the set of all Nash equilibrium profiles of the bimatrix game (subgame) $\Gamma_{(c,l)} = \left\langle I_{(c,l)}, J_{(c,l)}, A_{(c,l)}, B_{(c,l)} \right\rangle$. Based on the above-mentioned, and, namely, according to basic parallel strategies, we can proceed to distribution on a parallel computing system the subproblems, which in our case consist in

determining the sets $NE\left[\Gamma_{(c,l)}\right]$ for any calculation process $(l,c) \in L \times C$. Based on the definition of the Nash equilibrium profiles, any process $(l,c) \in L \times C$ of a parallel computing system with the distributed memory, simultaneously and independently determines the equilibrium profiles, $\left(i^*_{(l,c)}, j^*_{(l,c)}\right) \in NE\left[\left(A_{(l,c)}B_{(l,c)}\right)\right]$ for each subgame $\Gamma_{(c,l)} = \left\langle I_{(c,l)}, J_{(c,l)}, A_{(c,l)}, B_{(c,l)}\right\rangle$ based on the BSA.

Here we are going to analyze the following problem**:** *if we use the 2DBCMD&D algorithm, and process with the coordinates* (l,c) *determined as* $\left(i^*_{(l,c)}, j^*_{(l,c)}\right) \in NE\left[\left(A_{(l,c)}, B_{(l,c)}\right)\right]$*, then which conditions should be checked so that* $\left(\phi_{(l,c)}\left(i^*_{(l,c)}\right), \psi_{(l,c)}\left(j^*_{(l,c)}\right)\right) \in NE[\Gamma]$.

Using Proposition 1 for the **2DBCMD&D** algorithm, we can easily prove the following

Proposition 2 *Let* $(i^*, j^*) \in NE[\Gamma]$ *and there are the following: a process* (l,c)*, applications* $\phi_{(l,c)}: I_{(l,c)} \to I$*,* $\psi_{(l,c)}: J_{(l,c)} \to J$ *for which (1)-(2) is verified, and* $\left(i^*_{(l,c)}, j^*_{(l,c)}\right) \in NE\left[\left(A_{(l,c)}, B_{(l,c)}\right)\right]$*. Then* $i^* = \phi_{(l,c)}\left(i^*_{(l,c)}\right)$ *and* $j^* = \psi_{(l,c)}\left(j^*_{(l,c)}\right)$.

Proposition 2 means the following: for any Nash equilibrium profile in the global matrix game, there is a subgame generated by the **2DBCMD&D** algorithm, for which this strategy profile is also the equilibrium profile. In the next theorems [3], sufficient conditions are formulated under which an equilibrium profile in the bimatrix subgame, generated by the **2DBCMD&D** algorithm, becomes an equilibrium profile in the initial game with the global matrices.

Theorem 1 *Let's assume that* $\left(i^*_{(l,c)}, j^*_{(l,c)}\right) \in NE\left[\left(A_{(l,c)}, B_{(l,c)}\right)\right]$ *is determined by the process (l,c). If for any process on the column* c*, namely* $(\tilde{l}, c)$ *for all* $\tilde{l} \neq l$*, the condition* $j^*_{(\tilde{l},c)} \neq j^*_{(l,c)}$ *is fulfilled, and for any process from the line* l *namely* $(l, \tilde{c})$*, for all* $\tilde{c} \neq c$*, the condition* $i^*_{(l,\tilde{c})} \neq i^*_{(l,c)}$ *is fulfilled, then* $\left(\phi_{(l,c)}\left(i^*_{(l,c)}\right), \psi_{(l,c)}\left(j^*_{(l,c)}\right)\right) \in NE[\Gamma]$.

This theorem states the following: if in the $A_{(l,c)}$ submatrices of the processes on the column c there are no marked elements that belong to the column $j^*_{(l,c)}$ and, at the same time, in the submatrices $B_{(l,c)}$ of the processes on the line l there are no marked elements that belong to the line

$i^*_{(l,c)}$, then the strategy profile $\left(\phi_{(l,c)}\left(i^*_{(l,c)}\right), \psi_{(l,c)}\left(j^*_{(l,c)}\right)\right)$ is a Nash equilibrium profile in the initial global matrix game.

Let's analyze the case when there are (l,c) processes in the process grid, so that $\left(i^*_{(l,c)}, j^*_{(l,c)}\right) \in NE\left[\left(A_{(l,c)}, B_{(l,c)}\right)\right]$ but $\left(\phi_{(l,c)}\left(i^*_{(l,c)}\right), \psi_{(l,c)}\left(j^*_{(l,c)}\right)\right) \notin NE[\Gamma]$. In other words, *not every equilibrium profile in the subgame is an equilibrium profile in the global matrix game*.

The following theorem can easily be proved for this case.

Theorem 2 *Supposing for a given* $(l,c) \in L \times C$ *process found strategy profile* $\left(i^*_{(l,c)}, j^*_{(l,c)}\right) \in NE\left[\left(A_{(l,c)}, B_{(l,c)}\right)\right]$. *If for fixed* c *and all* $\tilde{l} \neq l$, *such that* $(\tilde{l},c) \in L \times C$, *the conditions* $a_{i^*_{(l,c)} j^*_{(l,c)}} \geq a_{i^*_{(\tilde{l},c)} j^*_{(l,c)}}$ *are fulfilled, where* $i^*_{(\tilde{l},c)} = \arg \max_{i_{(\tilde{l},c)} \in I_{(\tilde{l},c)}} a_{i_{(\tilde{l},c)} j^*_{(l,c)}}$ *and for fixed* l *and all* $\tilde{c} \neq c$ *such that* $(l,\tilde{c}) \in L \times C$, *the conditions* $b_{i^*_{(l,c)} j^*_{(l,c)}} \geq b_{i^*_{(l,c)} j^*_{(l,\tilde{c})}}$ *are fulfilled, where* $j^*_{(l,\tilde{c})} = \arg \max_{j_{(l,\tilde{c})} \in J_{(l,\tilde{c})}} b_{i^*_{(l,c)} j_{(l,\tilde{c})}}$, *then* $\left(\phi_{(l,c)}\left(i^*_{(l,c)}\right), \psi_{(l,c)}\left(j^*_{(l,c)}\right)\right) \in NE[\Gamma]$.

Finally, Theorem 1 and Theorem 2 determine the conditions under which the algorithm **2DBCMD&D** becomes the **PMDDA**.

5. Conclusion

In this article, to solve the bimatrix games in complete and imperfect information over the sets of pure strategies, we elaborate the parallel algorithm for which partitioning the input into several partitions of almost equal sizes and distributing this data on a parallel computing system is achieved by using the 2DBCMD&D algorithm. To solve recursively the sub-problem defined by each partition of the input, the BBrPA is used; finally, combining or merging the solutions of the different sub-problems into a solution for the overall problem is enough to use Theorem 2.

Thus, the main results of this article can serve as a basis for data parallelization on HPC systems and a large class of decision-making problems decided under conditions of risk and uncertainty, based on the

mathematical apparatus of game theory solving bimatrix games with very large matrices.

Acknowledgments. The Project 21.70105.9ŞD *Investigarea şi elaborarea infrastructurii integrate a mediului unificat "cloud computing" pentru a sprijini ştiinţa deschisă* has supported part of the research for this paper.

References

[1] Joseph Jaja, *An Introduction to Parallel Algorithms*. Addison-Wesley Publishing Company, Inc., 1992.

[2] *ScaLAPACK - Scalable Linear Algebra PACKage*. [Online]. Available: http://www.netlib.org/scalapack/

[3] Boris Hâncu and Emil Cataranciuc. *Parallel algorithm to solving 2d block-cyclic partitioned bimatrix games*. Studia Universitatis Moldaviae, seria Stiinte exacte si economice, nr. 2(142), Chisinau 2021. p. 26-39.

Hâncu Boris[1]

[1]Moldova State University
E-mail: boris.hancu@gmail.com, boris.hincu@usm.md

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

OCR: Handwrite Recognition

Ionuț-Alexandru Hristea, Adrian Iftene

Abstract

Computer Vision aids the computer in comprehending what occurs in a visual representation of the real world, such as videos or photographs. A subfield of Computer Vision is Optical Character Recognition, which works with handwritten or printed material that is converted into machine-readable language. All of the world's paper-based documents could be translated into a more convenient format for storage without having to rewrite them. The purpose of this paper is to develop an algorithm that can convert a picture of handwritten text on paper into a machine-readable format, making it easier for individuals to keep papers.

Keywords: Computer Vision, OCR, Convolutional Network.

1 Introduction

Handwriting recognition is an ever-increasing challenge these days, with applicability in many fields, but in special in forensics and biometrics. In addition to the advantage that you quickly get what someone wrote, you can find out depending on the shape of the characters, the inclination of the writing, and other details, information about the person who wrote it.

Current handwriting recognition techniques are based on machine learning and neural networks. Thus, in [1], the author provides a way to train a model and use it to recognize handwritten characters. In [2], the authors present EMNIST: an extension of MNIST to handwritten letters about the EMNIST dataset which is a combination of multiple datasets: MNIST, CIFAR-10, CIFAR-100, STL-10, and SVHN. In

[3], the authors present BatchNorm, one of the most successful architectural innovations in deep learning. In [4], the author experimented multiple network architectures to find the best one for MNIST dataset. In these cases, in order to prevent the overfitting in [5], the authors propose a new method "Dropout in Dropout". Starting from these models and techniques and using our previous work from [6, 7, 8], we do something similar in this paper.

2 Proposed Solution

2.1 Datasets

The first two models were trained using two datasets: MNIST[1] (for digits) and A-Z Handwritten Alphabets[2] (for letters). The letters dataset had over 370,000 images of alphabet letters in 28×28 pixels. For digits, the MNIST dataset has been used, having 60,000 images of digits for training and 10,000 for testing.

The last model uses EMNIST[3] dataset, which is an extension of the previous models, containing 814,255 characters.

2.2 Optical Character Recognition

In order to recognize the characters in an image, some pre-processing needs to be done.

- **First step** is to transform the image's colors to black and white (see Figure 1). When detecting the characters, the colors don't help too much, a gray-scale image being enough.
- **Second step** is to blur the image (see Figure 2). Blurring means each pixel becomes the mean of it's neighbors. The blur helps reduce the noise.

[1] http://yann.lecun.com/exdb/mnist/

[2] https://www.kaggle.com/datasets/sachinpatel21/az-handwritten-alphabets-in-csv-format

[3] https://www.nist.gov/itl/products-and-services/emnist-dataset

- **Third step** is to find the contours (see Figure 3). The algorithm is pretty simple. If the pixel's neighbors are different, for example, the left one is white and the right one is black, that pixel is part of a contour.

Figure 1. Grayscale image

Figure 2. Blurred image

The algorithm will then iterate through each contour and try to predict them. This character recognition works only if the characters are not connected. Two connected characters will be only one contour, hence only one character.

2.3 The First Model

A model must initially identify tiny 28×28 representations of one character in order to perform well when recognizing letters in a picture.

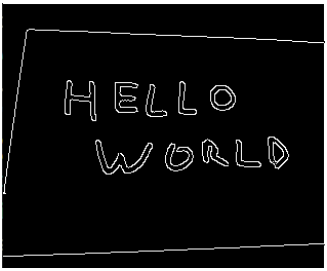


Figure 3. Contours

The model should learn as much as possible during training, but not too much, because else it will overfit, specializing on specific examples and failing to generalize. There are a variety of strategies that can help a model train more effectively while avoiding overfitting.

It's worth noting that this model is actually made up of two separate models: one for identifying letters and the other for classifying numbers, but they are almost identical under the hood.

Training

A convolutional layer is the network's first layer. A convolutional layer is a filter that extracts distinct features from an image. The following one is a regular layer called dense consisting of 128 neurons. The final layer is dense as well, but this time with the same number of neurons as the number of possible outputs. The number of neurons on the last layer is the only difference between the letters and the digits models.

The activation function is a mathematical function that is used to more or less activate a neuron in the final layer. The classification outcome is determined by the most active neuron. ReLU, or Rectified Linear Unit, is used on non-terminal layers, and Softmax is utilized on the last layer because it behaves like a probability function, with the best neuron having the highest probability.

There were also some optimization approaches applied. Dropout refers to the disabling of some neurons at each epoch in order to reduce

the reliance between them. Max Pooling is a technique for reducing image dimensionality by lowering the amount of pixels in the output from the previous convolutional layer. The last one, Flatten was used to convert the two-dimensional arrays, the images, to one-dimensional arrays. Finally, the model was trained for 10 epochs with a batch size of 256.

Performance

The performance metrics consist of testing the model on the test dataset and how well the model did in character recognition on some simple images. The charts below depict the training and testing accuracies for each epoch as well as the loss, which is a measurement of the difference between the expected and the model's delivered outcome.

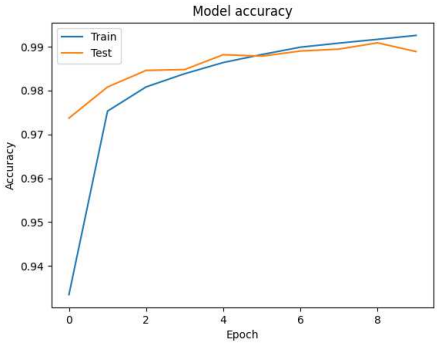


Figure 4. First model: Letters model accuracy for each epoch

We can see in Figure 4 how the accuracies change from epoch to epoch. Since the testing accuracy is lower than the train accuracy and decreasing in the last epochs, it means the model overfits. However, the accuracy in both testing and training is pretty good, hitting almost 99% in testing.

Experiments

In the first experiment, we will use a simpler image to show how good is the model. In Figure 5, we can see the actual recognition of the characters in the image. Since it's pretty hard to see each letter, below is a table with each recognized letter, the confidence, and the expected output. The average accuracy was 50%.

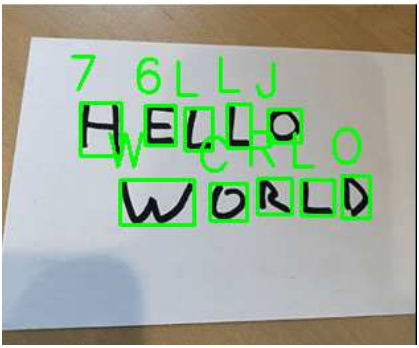


Figure 5. First model: Character recognition on first image

The second experiment has a more complex image, we can see how good the model performs on this one (see Figure 6).

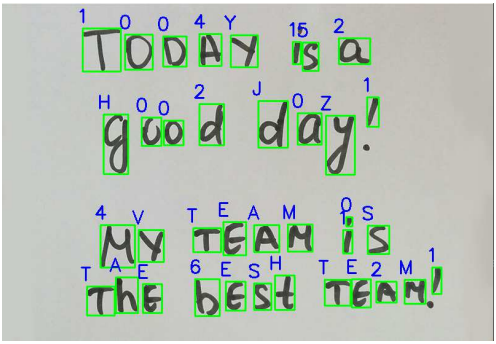


Figure 6. First model: Character Recognition on second image

2.4 The second model

In contrast to the first model, the second model is actually, ironically, only one model which is able to distinguish not only letters but also digits. The second difference is that this model works with 32×32 images of characters. It's worth mentioning that the datasets are the same, but scaled. Another improvement brought to this model is the data augmentation. This fancy word basically means that images are rotated, zoomed, shifted on height and width, and sheared.

Training

The neural network is pretty much the same as the one in the first model. The big differences here are the size, 32×32 for this model, compared to 28×28 for the first one, and the augmentation. The model was trained again for 10 epochs with a batch size of 256.

Performance

The results obtained are rather interesting and you will see exactly why. Here are the plots for the accuracy and loss on training and testing.

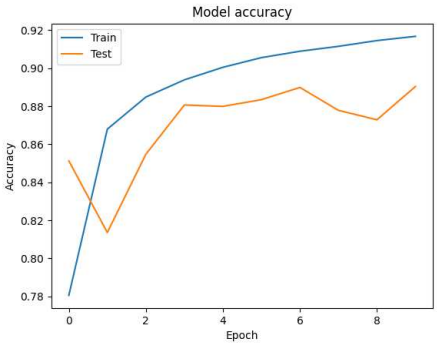


Figure 7. Second model: Accuracies for training and testing

In Figure 7, we can see lower accuracy in both training and testing and again the testing accuracy is lower.

Experiments

The first experiment has the same image from the first experiment for the first model. The accuracy is lower, but we'll see how it performs. As you can see in Figure 8, the results are way better than the ones we have seen for the first model (80% average accuracy).

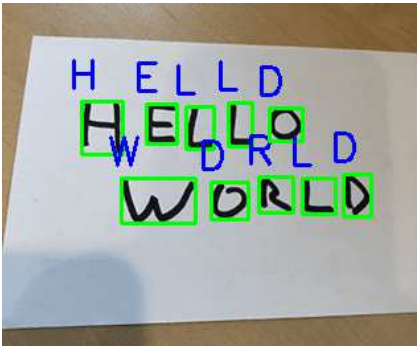


Figure 8. Second model: Character recognition on first image

The first experiment showed some improvements. Let's see how well it classifies the characters in the second image. Since the OCR algorithm is the same, it will still recognize a dot above the i as a character. This is caused by the fact that the OCR tries to find the contours. One whole contour is one character (see Figure 9).

This model managed to recognize 16 letters correctly, way better than the first one but there can be improvements.

3 Conclusion

From the beginning we have been looking to improve the model as much as we could. The recognition of the letters is a big problem and

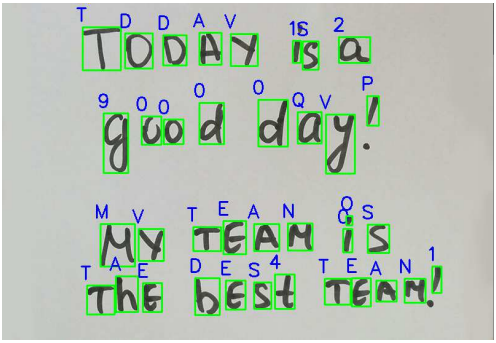


Figure 9. Second model: Character recognition on second image

the recognition of handwritten letters is a bigger one. The first model had the highest accuracy, but also performed the worst in experiments. One reason could be the fact that the dataset was smaller so the model could learn fast, but this is not optimal.

The second model had worse accuracy, but the performance was better. Although it used the same dataset, this model used data augmentation which created a better learning environment, but also a tougher one. The number of epochs wasn't that big, perhaps more epochs would create a better model.

In conclusion, the models have evolved slightly. You can say that each one is the result of training for an epoch and as we said, more epochs would create a better model.

References

[1] A. Rosebrock. *OCR: Handwriting recognition with OpenCV, Keras, and TensorFlow*. PyImageSearch, 2020.

[2] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik. *EM-NIST: an extension of MNIST to handwritten letters*. Arxiv:1702.05373, 2017.

[3] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry. *How Does Batch Normalization Help Optimization?* Arxiv:1805.11604, 2019.

[4] C. Deotte. *How to choose CNN Architecture MNIST*. Kaggle, 2018.

[5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. *Dropout in Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. Journal of Machine Learning Research, vol. 15, pp. 1929–1958, 2014.

[6] A. Iftene and A.M. Baboi. *Using semantic resources in image retrieval*. In 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, Elsevier. 5-7 September 2016, York, United Kingdom, vol. 96, pp. 436–445, 2016.

[7] G. L. Coca, S. C. Romanescu, S. M. Botez, and A. Iftene. *Crack detection system in AWS Cloud using Convolutional neural networks*. In 24rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. 16-18 September. Procedia Computer Science, vol. 176, pp. 400–409, 2020.

[8] L. G. Coca, A. Hanganu, C. G. Cușmuliuc, and A. Iftene. *UAIC2020: Lung Analysis for Tuberculosis Detection*. Working Notes of CLEF 2020 – Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020. vol. 2696, 2020.

Ionuț-Alexandru Hristea, Adrian Iftene

"Alexandru Ioan Cuza" University of Iasi,
Faculty of Computer Science, Romania
E–mail: ionut.hristea@info.uaic.ro, adiftene@info.uaic.ro

Proceedings of the Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Development of a plagiarism detection system in textual documents

Corina Negara, Ina Ciobanu, Anton Chiverciuc

Abstract

The purpose of this research[1] is to describe the phenomenon of plagiarism: definition and forms; to analyze existing plagiarism detection tools; to describe the plagiarism detection process and to develop the application for automatic plagiarism detection.

Keywords: Plagiarism, detection, PHP, framework Laravel, modelul MVC.

1. Introduction

Intellectual theft has always existed, and one of the first accusations of plagiarism comes from antiquity, from the poet Martial who accused other poets of copying his verses. He was also the first person to use the term "plagiarius" to refer to literary theft[2].

The epidemic of plagiarism, however, began suddenly, along with the development of computers and the Internet. Nowadays, plagiarism is exacerbated by the ease of taking PhDs, research papers, articles, ideas, images, etc. online.

Of course, there is the problem of monitoring and stopping the plagiarism process. Plagiarism detection can be done in different ways. One of the most traditional ways of plagiarism detection is human detection or expert detection. Although it has the highest degree of fidelity, this method is extremely time-consuming, in some cases the expert cannot identify the plagiarism (either he cannot find the source, or he does not know the language, etc.).

[1] The paper is developed within the project 20.80009.5007.22 "Intelligent computer systems for solving poorly structured problems, processing knowledge and large volumes of data", The State Program 2020-2023

[2] Kennedy A. A Short History Of Academic Plagiarism. [online]. Available on the Internet: https://www.quetext.com/blog/short-history-academic-plagiarism

Thus, there is a need to develop an automatic plagiarism detection system.

In this context, the optimal development of a plagiarism detection system involves the following aspects:

- permission to automatically scan documents and compare them with resources from databases to detect sources of plagiarism;
- generating reports according to the data collected regarding plagiarism.

However, in this article, the research problem lies in automating the process of detecting plagiarism from local documents and resources located on the Internet with the subsequent generation of the report based on the collected data.

2. Analysis of plagiarism detection tools

An effective understanding of the plagiarism detection process and the creation of methods to detect the given phenomenon requires direct knowledge of the definition and forms of the concept of plagiarism.

Plagiarism is defined, according to DEX, as a "literary, artistic or scientific work of someone else, appropriated (in whole or in part) and presented as a personal creation".

Forms of plagiarism are[3,4,5]:

1. Complete Plagiarism or full Copy&Paste – use the text of another author as own without any citation.
2. Direct Plagiarism or Partial Copy&Paste – use a part of the source text without any citation.

[3] Recommendations for Dealing with Academic Plagiarism in Student Texts. [online]. Available on the Internet: https://www.uni-goettingen.de/de/document/download/20a425e38a8b351a91c5d808b7ada5e7.pdf/Recommendations%20for%20Dealing%20with%20Plagiarism.pdf

[4] The 10 Types of Plagiarism Every Writer Should Know. [online]. Available on the Internet: https://inkforall.com/copy-editing/plagiarism-checker/10-types-of-plagiarism/

[5] Enago Academy. 8 Most Common Types of Plagiarism to Stay Away From! [online]. Available on the Internet: https://www.enago.com/academy/fraud-research-many-types-plagiarism/

3. Paraphrasing Plagiarism or Disguised Plagiarism - the source text is paraphrased or summarized without citing the source. It is similar to an Apparent Paraphrase.
4. Plagiarism by Translation – translation of a text in a foreign language without citation.
5. Inaccurate Authorship - a person contributes to a text but does not get credit for it or a person gets credit without contributing to the work.
6. Mosaic or Patchwork Plagiarism or Remix - use different parts of text from different source texts, as a direct quote or as a paraphrase, without citation.
7. Self-Plagiarism or Auto Plagiarism or Recycle – the person reuses his own text/parts of text from a previously published paper.

A multitude of text document plagiarism testing services can be found on the net, all of which have both advantages and disadvantages. These applications include: advego.ru, text.ru, content-watch.ru, etxt.ru, antiplagiat.ru, plagiarisma.net, plagtracker.com, copyscape.com, etc. Mostly, these applications work according to the principle of a "black box" providing results based on logic or algorithms that are not declared or cannot be analyzed. Applications provide results, usually in the form of originality grades, but there is no confidence that all cases of plagiarism are analyzed[6].

At the same time, the biggest problem common to all nominated services is the extremely small volume of work admissible for free daily verification up to 1000-1500 characters.

Thus, these services seem useless for testing large-volume text documents, but it is possible for a fee.

3. Plagiarism detection

A plagiarism detection system broadly contains 3 components: two inputs and one output[7]:

[6] Weber-Wulff D. Plagiarism detectors are a crutch, and a problem [online]. Available on the Internet: https://www.nature.com/articles/d41586-019-00893-5

[7] Subroto, I., Selamat, A. (2014). Plagiarism Detection through Internet using Hybrid Artificial Neural Network and Support Vectors Machine. TELKOMNIKA (Telecommunication Computing Electronics and Control). 12.

1. Entry into the system is represented by two sets of documents:
   a. Documents for verification for which the system must determine the degree of originality;
   b. The collection of reference documents against which the verified documents are compared.
2. Output from the system – represents a set of overlaps between the verified text and the documents in the reference document collection.

Reference documents can come from different sources: local or external. The local source of reference documents is a database containing various sources (theses, articles, monographs, etc.) that are stored locally on the server. The external source assumes that the reference documents are on the Web. The plagiarism detection system works faster with local reference documents because they are preprocessed, indexed, and processed locally.

One of the effective methods of detecting plagiarism is based on the notion of shingle. Shingle is the way of grouping the ordered words according to the predetermined length, usually, this being 3-5 words. As a result, groups of *n* words are obtained.

For example, the text to be split into shingles is: „Plagiarism is a fraudulent reproduction of copied text".

If the length of the groups is 5 words, then the groups will be:

Group 1 – „Plagiarism is a fraudulent reproduction";

Group 2 – „is a fraudulent reproduction of";

Group 3 – „a fraudulent reproduction of copied";

Group 4 – „fraudulent reproduction of copied text".

These groups are formed for the text to be checked and the text from the reference documents, and then compared to determine the common fragments. The result is determined by dividing the number of common shingles by the total number of shingles.

209. 10.12928/TELKOMNIKA.v12i1.648. [online]. Available on the Internet: https://www.researchgate.net/publication/270502073_Plagiarism_Detection_through_Internet_using_Hybrid_Artificial_Neural_Network_and_Support_Vectors_Machine

This method is effective in that it can quickly detect files that have no common fragments and files that have common elements with the processed text, respectively[8].

To determine the degree of similarity between two sets of shingles, the Jaccard similarity coefficient can be used to determine the degree of overlap between two sets.

The formula to find the Jaccard coefficient is:

$$\text{Jaccard coefficient} = \frac{\text{the number in both sets}}{\text{the number in either set}} \qquad (1)$$

The same formula in notation is[9]:

$$J(S,T) = |S \cap T| / |S \cup T|$$

4. Development of Plagiatus App

The Plagiatus application was developed using the Laravel framework and having the MVC (Model-View-Controller) model as its architectural model. The way the application works involves the following steps:

1. Preprocessing stage – at this stage, the text subject to verification is divided into shingles; spaces and punctuation marks are removed from the shingles; the text is moved to the lower case (lowercase letters). The number of words in a shingle can be set by the user, the minimum being 3 words.

[8] Ceglarek, Dariusz. (2013). Evaluation of the SHAPD2 Algorithm Efficiency in Plagiarism Detection Task Using PAN Plagiarism Corpus. Computer Science & Information Technology. 3. 10.5121/csit.2013.3312. [online]. Available on the Internet: https://www.researchgate.net/publication/307748601_Evaluation_of_the_SHAPD2_Algorithm_Efficiency_in_Plagiarism_Detection_Task_Using_PAN_Plagiarism_Corpus

[9] Fletcher, Sam & Islam, Md. (2018). Comparing sets of patterns with the Jaccard index. Australasian Journal of Information Systems. 22. 10.3127/ajis.v22i0.1538. [online]. Available on the Internet: https://www.researchgate.net/publication/323624308_Comparing_sets_of_patterns_with_the_Jaccard_index

2. Comparison stage – at this stage, the set of shingles in the text under verification is compared with the text of the reference documents. For each reference document, the set of fragments that are found in both documents (the document under verification and the reference document) is determined. The comparison will be made with each reference document, so the larger the database, the longer the check will be. To reduce processing time, reference documents are preprocessed when added to the database.
3. Jaccard coefficient calculation stage – at this stage, the similarity coefficient between the set of shingles is calculated according to equation (1).
4. The source detection stage – is the stage at which the reference document/documents that have common fragments with the document under verification are determined.
5. The report generation stage – is the last stage and involves the creation of the similarity report between the document subject to verification and the reference documents.

5. Plagiatus App Functionality

The Plagiatus app offers the following features:

- CRUD operations with user files;
- placing the documents in the waiting order for plagiarism checking;
- viewing the document's uniqueness percentage after local (with local files) and external (with Internet resources) verification;
- downloading the previously uploaded file;
- brief and extended visualization of the results of document uniqueness verification in local and external/online sources;
- exclusion of fragments from the document subject to verification;
- generation of the reduced and extended report.

The main page allows viewing and listing documents and their verification results (Figure 1).

This page offers the following possibilities: download the original file, upload the new document to the database, delete documents from the database, and go to the detailed pre-report pages.

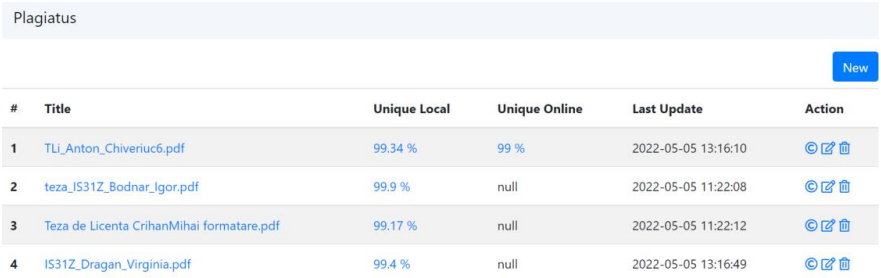


Figure 1. The main page of the Plagiatus App

The pre-report pages are roughly identical for local and external verification. The only difference between the results of the online check and the offline one is that in the online one, instead of naming the documents, links to online resources that are candidates as a source of inspiration are indicated.

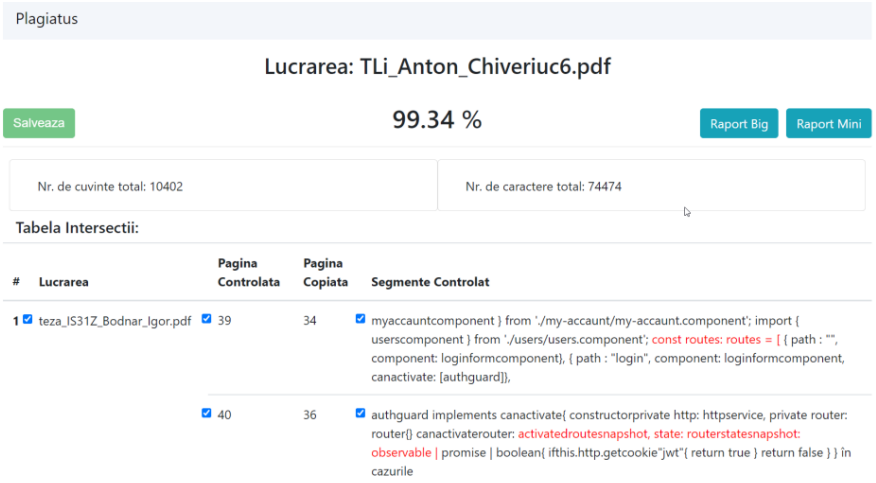


Figure 2. Plagiarism check results view page

This page allows you to exclude fragments from the report (Figure 2). They can be excluded according to the fragment, according to the page, where there can be several fragments; after that, it is necessary to press the save button for the changes to take effect. In addition to all this, there are

also 2 buttons that allow the generation of reports. Clicking the generate report button will download the report in PDF format.

The report has several compartments.

The first part of the report contains all the data related to the checked document: the name of the document, the name of the author, the date the document was uploaded, the date the report was generated, and the pages/fragments excluded from the check.

The second compartment contains the details of the work: length in characters, length in words, and the number of words with characters from different languages.

The next compartment contains the result of the similarity between the document under verification and the documents or inspiration sources.

The next section shows the details of the similarity in which the documents or online sources with which a match was found are listed. Depending on the type of report (extended or reduced), in the similarity details line, the details will be expanded; the source, the percentage, and the coinciding fragment will be indicated, while in the reduced report, the coincident textual fragment will be missing.

6. Conclusion

In this paper, the research on how to apply the PHP language and the Laravel framework in the development of a plagiarism detection system in textual documents located locally on the server and online resources, are given.

References

[1] B. Gilmore. *Plagiarism: A How-Not-to Guide*. Heinemann, 2009. 6p. ISBN: 978-0325026435.

[2] V. Bowman. *Stop Plagiarism: A Guide to Understanding and Prevention*. New York: Neal-Schuman Publishers, 2010. 96p. ISBN: 978-1555707163.

[3] T. Darr. *Combating Plagiarism: A Hands-On Guide for Librarians, Teachers, and Students*. Libraries Unlimited, 2019. 39p. ISBN: 978-1440865466.

[4] D. Correa and P. Vallejo. *Practical Laravel: Develop clean MVC web applications*. Independent publication, 2022. 74p. ISBN: 979-8416784195.

[5] M. Stauffer. *Laravel: Up and Running: A Framework for Building Modern PHP Apps*. California: O'Reilly Media, 2016. 23p. ISBN: 978-1491936085.

Corina Negara[1], Ina Ciobanu[2], Anton Chiverciuc[3]

[1]Institution: Alecu Russo Balti State University
E-mail: negara.corina@usarb.md

[2]Institution: Alecu Russo Balti State University
E-mail: ina.ciobanu@usarb.md

[3]Institution: Alecu Russo Balti State University
E-mail: kiverchuk.anton@gmail.com

Elearning content processing situations and their solutions

Alexandr Parahonco, Mircea Petic

Abstract

The article discusses processing approaches and their solutions for further usage of content generation in e-courses. It begins with an analysis of web-scraping techniques focused on fetching information from the web network. Then, it discusses their role in modern life and ways of application. Also, the scheme of the model for the dynamic creation of training courses is presented. Finally, the paper discusses content processing situations and their solutions.

Keywords: E-learning, content generation, web scraping, crawler, model.

1. Introduction

The twenty-first century has attended the emergence of groundbreaking information technologies that brought changes in our life. Since the mid-1990s, the Internet gave a start to methods, tools, and gadgets that covered all academic disciplines and business sectors. Soon afterward we witnessed a chain of web 2.0 technologies like E-commerce, which started social media platforms, E-Business, E-Learning, E-government, Cloud Computing, and more other in 2021 [1].

E-learning platforms require the elaboration of high-quality and relevant teaching resources and the constant updating of existing ones. This, in turn, is a complex process consisting of processing a variety of materials, their analysis, synthesis, creative development, and processing of all elements to build a single harmonious structure. Up to now, far too little attention has been paid to dynamic content generation for e-learning courses.

The aim of the article is to analyze web-scraping techniques focused on fetching information and propose a model for the dynamic creation of training courses. Especially it should be noted the discussed content processing situations and their solutions[1].

The article begins with an examination of web-scraping techniques for retrieving data from the Internet. It discusses their role in modern life as well as possible applications. The model's scheme for dynamically creating training courses is then presented. Finally, the paper discusses content processing problems and solutions.

2. State of the art

Udit Sajjanhar was one of the first who wrote an article about extracting information from the Internet in 2008. The author describes educational content mined from university websites in the form of course pages. His system tries to learn the navigation path by observing the user's clicks on as few example searches as possible, and then uses the learned model to automatically find the desired pages using as few redundant pages fetches as possible. Following that, H.W. Hijazi and J.A. Itmazi use keywords divided into two categories as a basis for launching web crawlers: included and excluded from the search query. The authors crawl websites of open educational resources (OER), mainly the Massachusetts Institute of Technology (MIT), which first announced plans to make all of its course materials freely available [2 - 4].

However, the idea of processing web content automatically, technically, came in 1993 with the World Wide Web Wanderer, the first web robot, the sole purpose of which was to gauge the size of the web. Though, it gave birth to the first concept of web-crawling. Soon afterward, the first crawler-based web search engine, JumpStation, was developed. It built a new milestone in web technologies — the prototype of Google, Bing, Yahoo, and other search engines on the web today. In 2004 it acquired a new concept — visual web scraping, provided by BeautifulSoup HTML parser and Web Integration Platform version 6.0

[1] 20.80009.5007.22 Intelligent information systems for solving ill-structured problems, processing knowledge and big data. https://www.math.md/projects/20.80009.5007.22/

[5]. This brought popularity to that technique. Many people, including researchers, started the use of web scraping in different domains.

During the ages, the concept of fetching information from the Internet has evolved into new technology — web scraping (web data extraction). It includes two categories of techniques, such as manual equipment (copy-paste) and automatic data scraping. Manual scraping involves copying and pasting web content, which takes a lot of effort and is highly repetitive in the way it is carried out. Automated scraping techniques shifts from HTML Parsing [6 - 7], DOM Parsing, and XPath to Google Sheets and Text Pattern Matching. Moreover, some semi-structured data query languages, such as XQuery and HTQL, can be used to parse HTML pages and retrieve and transform page content [7].

3. Program model for the dynamic creation of training courses

Our numerous studies guided us to the development of content generation applications. We took as a basis the concept of manual scraping and designed finally the program model given in Fig. 1.

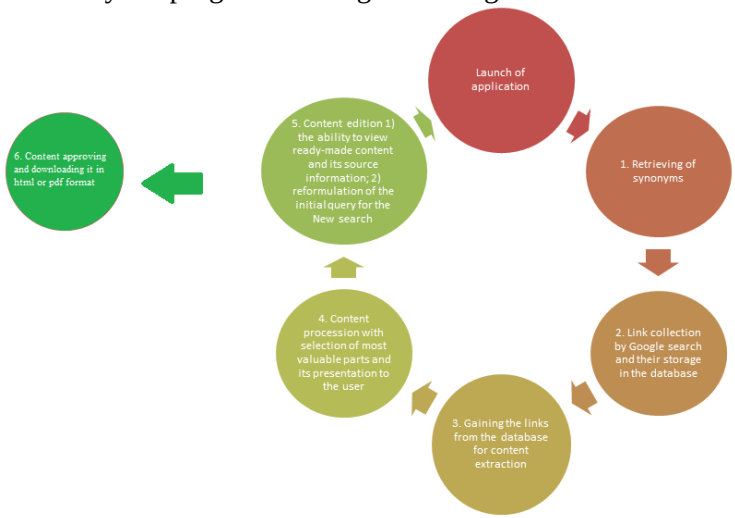


Figure 1. Scheme of the program model for the dynamic creation of training courses

According to our approach, we have 6 steps. In the first step, some web crawlers create networks of synonyms. In the second step, our application uses the original request and/or their selected synonyms for advanced search using Google search. Next, in step three, we gain from Google links for the requests and process them (crawl, select necessary fragments), storing all the information in the database. According to steps 4 and 5, well-merged content should be generated and further exported in HTML or PDF formats in step 6.

4. Resource processing

Resource processing relates to step 3 where the application retrieves links from the database and fetches information from them. This procedure includes text, image, and video extraction along with a selection of meaningful information (due to the search request). A detailed view of this procedure can be demonstrated in Fig. 2.

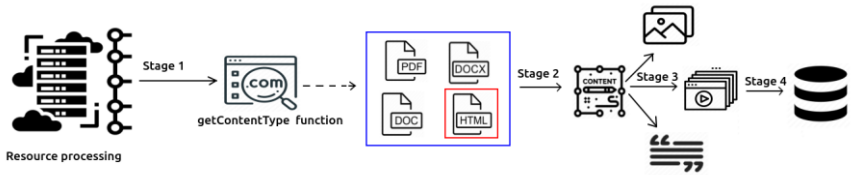


Figure 2. Content selection

The system itself comprises 4 stages for each resource. It begins with the identification of the resource kind. **getContentType** function gets headers of the *url* address and returns its MIME type. Then the necessary modality of handling resources executes. Here a simple web page should be separated from other kinds of sources representing uploaded documents on the Internet. Thus, the content of the web page can be obtained immediately in contrast with resources such as *doc*, *docx*, and *pdf*, which must be first downloaded and then processed.

Regarding web page resources, the retrieved content at stage 2 is used for searching image and video addresses by their tags *img* and *video at stage 3*. Text extraction (meaningful parts) is undertaken by **findInText** function. All gained information then is stored in the database.

Documents such as *doc*, *docx,* and *pdf* have another sequence of actions. After the download, they get processed. *Doc* format contains binary data and the application cannot obtain all the content except for the text. Hence, **findInText** function is applied at the next stage to find the searched information and save it. Both *docx* and *pdf* formats are more flexible in this respect as they allow us to fetch images and text. The procedure with text does not differ from *doc* format. However, throughout the image manipulation, the application creates some folder structures on the server and stores them there. Next, each image undergoes the procedure of optical character recognition (OCR). This is important as often information is scanned and saved into *pdf* or *docx* formats (books, magazines). Thus, OCR provides some text to find with **findInText** function.

At the end of resource processing, results are presented to the user.

5. Conclusion

The huge amount of information on the Internet is the mark of the 21st century. There is similar information, but it is outdated or fake. This necessitates the involvement of information technologies and approaches in order to select the most valuable pieces of it. It is especially important in the education area, where qualitative and up-to-date content plays an important role in the specialist formation.

The research proposes to build a focused web crawler for intelligent data extraction from web sources. We have adopted this idea and proposed an application model for the dynamic creation of training courses. Currently, we have focused on content processing in different file types and have elaborated on a common algorithm for its processing. Our next step is to implement Data mining and Text mining approaches to give the generated content more logic.

Acknowledgments. This article was written as part of the research project 20.80009.5007.22 “Intelligent information systems for solving ill-structured problems, processing knowledge and big data”.

References

[1] M. Yamin. *Information technologies of 21st century and their impact on the society*. In: “International Journal of Information Technology”,

vol. 11, no. 16-8 (2019), pp. 759-766, DOI: 10.1007/s41870-019-00355-1.

[2] U. Sajjanhar. *Focused Web Crawling for E-Learning Content*. Master's thesis. Indian Institute of Technology Kharagpur, Master of Technology In Computer Science and Engineering, 2008.

[3] H.W. Hijazi and J.A. Itmazi. *Smart Crawler Based e-Learning*. Hammamet-Tunisia, vol. 11, 2013, pp. 209 – 2016.

[4] Y. Biletskiy, M. Wojcenovic, and H. Baghi. *Focused Crawling for Downloading Learning Objects – An Architectural Perspective*. Interdisciplinary Journal of E-Learning and Learning Objects, vol. 5, no. 1, 2009, pp. 169 – 180.

[5] D. Kremer. *Brief History of Web Scraping*, 2021, https://webscraper.io/blog/brief-history-of-web-scraping.

[6] Bo. Zhao, *Web Scraping. Springer International Publishing*, vol. 5, (2017), 632 p, DOI: 10.1007/978-3-319-32001-4_483-1.

[7] S. Ruihua. *Joint Optimization of Wrapper Generation and Template Detection*. KDD07: The 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Minin, August, (2007), pp. 894 – 902, DOI: 10.1007/978-1-4419-7735-9_4.

[8] *Introduction to Web Scraping*, 2019, https://www.geeksforgeeks.org/introduction-to-web-scraping/.

Alexandr Parahonco[1], Mircea Petic[2]

[1]Affiliation/Institution: junior researcher / Vladimir Andrunachievici Institute of Mathematics and Computer Science, MD-3100, Moldova

software engineer / Department of Information Technologies, Alecu Russo State University of Balti, MD-3100, Moldova
E-mail: alexandr.parahonco@usarb.md

[2]Affiliation/Institution: leading researcher / Vladimir Andrunachievici Institute of Mathematics and Computer Science, MD-3100, Moldova

PhD, associate professor / Alecu Russo State University of Balti, MD-3100, Moldova
E-mail: mircea.petic@math.md

Formalization of decision knowledge and reasoning for casualty prioritizing

Iulian Secrieru, Constantin Gaindric, Elena Guţuleac,
Olga Popcova, Tudor Bumbu

Abstract

Primary triage, based on vital signs, is the first and the most important stage of mass casualty situations management. There is still no single approach, both among experts in emergency medical care and among developers of decision support systems. This is explained by the existence of various medical protocols (mandatory for execution at the national level), focusing on the main types of disasters that prevail in the region.

In this article we describe the formalization process, aiming to create an efficient inference for casualty prioritizing, based on vital signs.

Keywords: medical informatics, pre-hospital triage, mass casualty situations, knowledge acquisition and formalization.

1. Introduction

Knowledge formalization and the corresponding reasoning schemas selection is considered the key phase in the development of any inference module.

The domain of mass casualty situations and disasters management, on the one hand, is well structured, on the other hand, there is no general and universal approach in the world. The same applies to the main stage of the process of management – casualties primary triage, based on vital signs (see Fig. 1 - Obj. 1). Different countries design their triage model for emergencies according to their native medical protocols, resources, and forces.

The main drawback of many of the proposed approaches is that at the knowledge acquisition stage, given that national protocols/standards are 'over-structured', and the developers are forced to choose 'rigid' schemas of its representation. As a result, the inference does not correspond to the daily work and habits of the first-aid person, who is the end-user.

The discrepancy between the inference module of the knowledge-based system and the form of the doctor's diagnostic reasoning may become the cause of different mistakes or it may lead to rejection of the user to utilize it in medical practice.

Decision knowledge and reasoning formalization techniques, aiming to create an efficient inference for casualty prioritizing, based on vital signs, are described in this article.

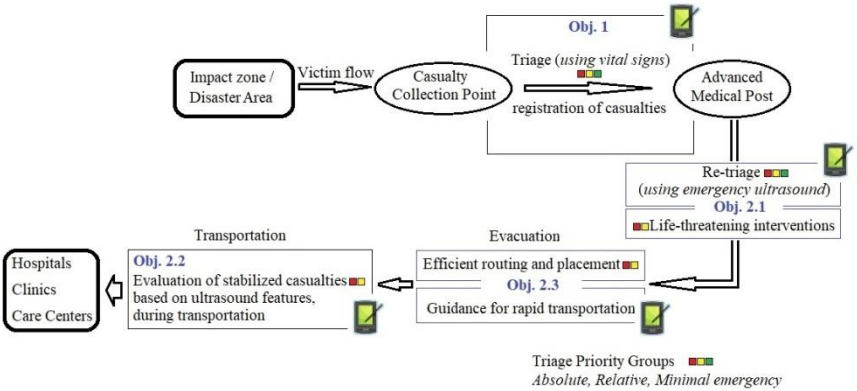


Figure 1. Victim flow management. Example scenario

2. Knowledge representation schemas

The main goal of the knowledge representation schema is to represent professional knowledge in a manner to facilitate drawing conclusions (inferencing) [1].

We distinguish two approaches: single and hybrid knowledge representation schemas.

Hybrid schemas represent integrations of two or more single knowledge representation schemas.

The most popular single knowledge representation schemas are the following:

- Decision trees and their descendants (frames or schemas);
- Semantic nets, Conceptual graphs, ontologies;
- Symbolic rules;
- Fuzzy rules (fuzzy logic);
- Case-based representations;
- Neural networks;
- Belief networks (or probabilistic nets).

Semantic nets, decision trees, and ontologies represent knowledge in the form of a graph (or a hierarchy) [2]. Nodes in the graph represent the concepts and edges represent the relations between the concepts. All of these knowledge representation schemas are very natural and well suitable for representing structural and relational knowledge.

Symbolic rules (symbolic reasoning) are one of the most popular knowledge representation schemas [3], representing general domain knowledge in the form of IF-THEN rules:

if <conditions> then <conclusion>,

where the term <conditions> represents the conditions of a rule, whereas the term <conclusion> represents its conclusion. The inference engine uses the knowledge in the rule base as well as facts about the problem at hand to draw conclusions. The efficiency of the inference process depends on the length of the inference chains.

3. Casualty prioritizing in mass casualty situations

By studying the existing types of medical triage, we can distinguish two main logical approaches: algorithmic [4] and numerical [5].

When the algorithmic model is used, the casualty should be assigned to a certain category (casualties with severe, moderate, or minor injuries), taking into account every vital sign (ability to move, breathe, level of consciousness, etc.). If the considered parameter is within the normal range, then the next parameter is studied according to its priority within the triage system.

When the numerical approach is used, the doctor who is performing medical triage should simultaneously assess all parameters of the model. As a result, the final assessment of the casualty state is made, being based

on the overall assessment of all parameters of the model. In accordance with the final assessment, the casualty is assigned to one of the categories of this medical triage.

4. Formalization of decision knowledge and reasoning. Knowledge acquisition

In collaboration with a team of medical experts in clinical emergency medicine, the minimum set of parameters needed for casualty registration was identified, so that the record, accompanying the casualty, contains all the information that will enable doctors from specialized medical centers to intervene operatively in the treatment.

These parameters cover all stages of the initial assessment of casualty and the organization/structuring of primary medical data.

The medical record for casualty registration consists of personal data, time interval, type of injury (resulting from visual inspection of casualty), and values of basic attributes (parameters) that describe vital signs.

The following 9 parameters were selected: visual inspection, Glasgow Coma Scale, airways, pulse, systolic and diastolic blood pressure, respiratory rate, oxygen saturation, and individual mobility. As the knowledge representation schema, there was selected the tabular form (most often used in the field of emergency medicine) [6].

There was developed an acquisition web-module – Medical Data Management Platform (https://g5700.math.md/admin/).

This platform allows one to record and store data about the casualty state, prioritizing them in 4 emergency categories (RED (I), RED (II), YELLOW, GREEN), based on vital signs.

5. Knowledge base kernel creation

The inference module is based on 4 decision rules, identified in collaboration with medical experts.

As the formalization schema of the decision rules, there was used symbolic rules, once again – being one of the most popular methods in the field of medical information systems.

All 4 rules were formalized and integrated into the inference module, representing now the knowledge base kernel.

Find below how a rule is represented in the inference module.

Pseudo-code (Java):

```
// conditions for category Red I
if ((glassgow_coma >=3 && glassgow_coma <= 8) &&
      (airways==1) &&
      (pulse<40 || pulse > 120) &&
      (systolic<80) &&
      (respiratory_rate<13 || respiratory_rate > 35) &&
      (oxygen <= 85) &&
      (mobilitiy==1)) {
   alert("Category: Red I");
   triage_category = 1;
   return triage_category;
}
```

These rules represent the scoring system for the triage of casualties. To validate this result, it was decided to create a synthetic data set and pre-test the decision algorithm.

6. Pre-testing and conclusion

The pre-testing process showed: i) a large number of cases with an 'undetermined' emergency category, if a totally random selection is applied.

The process of creating the data set showed: ii) the need to use a systemic approach to generate the synthetic data set; iii) the need to create an additional field in the medical record for a casualty on the web platform – for the comments of medical experts in the synthetic data validation process.

As a result, a synthetic data set of 56 cases (medical records) was created, following the approach below:

A) Values for all 9 parameters were selected from the same column, representing some emergency category;

B) 7-8 values from one column and 1-2 values from the neighboring column(s).

For 32% of these 56 cases, the emergency category was determined unequivocally, for 68% – the opinion of medical experts is required. In this sense, there was proposed and implemented the creation of an

additional field in the medical record for a casualty on the web platform – for the comments of medical experts in the synthetic data validation process.

The obtained result allows the authors to state that the proposed approach is a viable one for creating an efficient inference for casualty prioritizing, based on vital signs. For the selection of all used methods and technologies, there were formulated the grounds and rationale.

7. Future work

Both the study of the problem domain and the obtained result showed the existence of a one-sided orientation in the selection of the way to describe the reasoning for casualty prioritizing – algorithmic (decision trees, decision rules, etc.) or numerical (tabular form, scoring system, etc.).

At first glance, this selection is determined by the type of data source and the data itself. If developers have access to expert data and experts, which can formulate their professional knowledge in the form of rules, then the algorithmic approach is chosen. Otherwise, if developers have access only to the set of precedents (real cases) – the numerical approach is chosen.

Another reason is the time available for decision-making. If time is extremely limited, then the algorithmic method is chosen because it allows one to generate the conclusion without specifying the values for all parameters. If time allows for determining all the values for the system of selected parameters, then the numerical method is chosen, and even more, an attempt to create a scoring system is made.

A deeper analysis of the obtained result allowed the authors to make the following hypothesis: consciously or unconsciously the developers of medical information systems make their choice based on the restrictions imposed by the end-user habits – first-aid person and/or by the subdomain, in which the future information system will be used (disaster type like an earthquake, chemical or nuclear accident, explosion, flood, etc.; restrictions by age or any anthropometric data).

To verify the formulated hypothesis, it would be interesting to carry out both approaches – algorithmic and numerical (including a scoring system) for the same restrictions of the problem subdomain and the same set of real cases, described with the same set of 9 parameters.

Acknowledgments. The Moldovan State Program project 20.80009.5007.22 "Intelligent information systems for solving ill-structured problems, processing knowledge and big data" has supported part of the research for this paper.

References

[1] Iu. Secrieru. *Structured knowledge management techniques for the development of interactive and adaptive decision support system*. In: Computer Science Journal of Moldova, nr. 1 (49), vol. 17, 2009, Kishinev, pp. 58-73, ISSN 1561-4042.

[2] M. Negnevitsky. *Artificial intelligence: A guide to intelligent systems*, Reading, MA: Addison Wesley (2002).

[3] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach* (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, (2003), ISBN 0-13-790395-2.

[4] Gh. Ciobanu, M. Pîsla, F. Gornea et al. *Ghid naţional privind tirajul medical în incidente soldate cu victime multiple şi dezastre*. Centrul Nat. Șt.-Practic Medicină de Urgenţă, Centrul Republican Medicină Calamităţilor. – Chişinău, 2010, 36 p, ISBN 978-9975-80-337-3.

[5] S.P. Gormican. *CRAMS scale: field triage of trauma victims.* Ann Emerg Med. 1982 Mar; 11(3), pp. 132-135.

[6] C. Gaindric, S. Şandru, S. Puiu, O. Popcova, Iu. Secrieru, and E. Guţuleac. *Advanced pre-hospital triage based on vital signs in mass casualty situations.* Proceedings of the Workshop on Intelligent Information Systems WIIS2021, October 14-15, 2021, Chisinau, Republic of Moldova, pp. 130-134, ISBN 978-9975-68-438-5.

Iulian Secrieru[1], Constantin Gaindric[2], Elena Guţuleac[3], Olga Popcova[4], Tudor Bumbu[5]

[12345]Affiliation: Vladimir Andrunachievici Institute of Mathematics and Computer Science, Chisinau
E-mails: [1]iulian.secrieru@math.md, [2]constantin.gaindric@math.md [3]elena.gutuleac@math.md, [4]oleapopcova@yahoo.com, [5]tudor.bumbu@math.md

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Simple Stein medial quasigroups

Victor Shcherbacov, Valentina Demidova, Petr Radilov

Abstract

Using computer calculations, we research simple Stein medial quasigroups.

2000 Mathematics Subject Classification: 20N05

Keywords: computer science, quasigroup, Stein quasigroup, T-quasigroup, medial quasigroup.

1 Introduction

This paper is connected in main with the following **Belousov problem 1 [2, 6]**: Find a complete characterization of groups isotopic to quasigroups that satisfy one of the identities: $x(y \cdot yx) = y$ (of type T_2, [2, 10]), $xy \cdot yx = x$ (Schröder 2nd law, [9]), $xy \cdot yx = y$ (Stein 3rd law, [12], [9]).

These identities guarantee that a quasigroup is orthogonal to its parastrophe [2, 10].

Necessary definitions can be found in [1, 4, 8, 10, 11].

Definition 1. Binary groupoid $(Q, \circ)$ is called a left quasigroup if for any ordered pair $(a, b) \in Q^2$ there exists the unique solution $x \in Q$ to the equation $a \circ x = b$ [3, 4, 8, 1].

Definition 2. Binary groupoid $(Q, \circ)$ is called a right quasigroup if for any ordered pair $(a, b) \in Q^2$ there exists the unique solution $y \in Q$ to the equation $y \circ a = b$ [3, 4, 8, 1].

Definition 3. Binary groupoid $(Q, \circ)$ is called a quasigroup if it is a right and left quasigroup [3, 4, 8, 1, 5, 10].

Definition 4. Binary groupoid $(Q, \cdot)$ is called medial if this groupoid satisfies the following medial identity:

$$xy \cdot uv = xu \cdot yv \tag{1}$$

for all $x, y, u, v \in Q$ [1, 10].

We recall

Definition 5. Quasigroup $(Q, \cdot)$ is a T-quasigroup if and only if there exists an abelian group $(Q, +)$, its automorphisms φ and ψ, and a fixed element $a \in Q$, such that $x \cdot y = \varphi x + \psi y + a$ for all $x, y \in Q$ [7].

A T-quasigroup with the additional condition $\varphi\psi = \psi\varphi$ is medial [1, 10].

2 Simple medial quasigroups with 3-rd Stein identity

A quasigroup Q is simple if its only normal congruences are the diagonal congruence $\widehat{Q} = \{(q, q) \mid q \in Q\}$ and universal congruence $Q \times Q$. Any quasigroup of prime order is simple [10, p. 60].

A program was written in the language Pascal. Using this program, the following theorem is proved.

Theorem 1. For the following prime numbers, there exist simple medial quasigroups, in which the identity $xy \cdot yx = y$ is true (3-rd Stein identity). 5; 13; 17; 29; 37; 41; 53; 61; 73; 89; 97; 101; 109; 113; 137; 149; 157; 173; 181; 193; 197; 229; 233; 241; 257; 269; 277; 281; 293; 313; 317; 337; 349; 353; 373; 389; 397; 401; 409; 433; 449; 457; 461; 509; 521; 541; 557; 569; 577; 593; 601; 613; 617; 641; 653; 661; 673; 677; 701; 709; 733; 757; 761; 769; 773; 797; 809; 821; 829; 853; 857; 877; 881; 929; 937; 941; 953; 977; 997, 1009.

Proof. We have constructed quasigroups of these prime orders. □

In Table 1, for short we use ordered triplets of numbers. For example, triplet (1, 3, 5) means that there exist quasigroups with 3-rd Stein identity over the group of residues $Z_5 : (Q, \cdot)$, $x \cdot y = 1 \cdot x + 3 \cdot y \mod 5$.

Moreover, there exist the following three quasigroups over the group Z_5: $(Q, *)$, $x*y = 3*x+1*y \mod 5$; $(Q, \star)$, $x\star y = (1+1)\star x+(3+1)\star y \mod 5$, i.e., $x\star y = 2\star x+4\star y \mod 5$; $(Q, \circ)$, $x\circ y = (3+1)\circ x+(1+1)\circ y \mod 5$, i.e., $x \circ y = 4 \circ x + 2 \circ y \mod 5$.

Quasigroups $(Z_5, \star)$ and $(Z_5, \circ)$ are idempotent.

We construct the quasigroup $x \bullet y = 234x + 774y \mod (1009)$ and check that in this quasigroup, 3-rd Stein identity is fulfilled: $234(234x+774y)+774(234y+774x) = y \mod (1009)$, $54756x+18116y+181116y+599076x = y \mod (1009)$, $653832x + 362232y = y \mod (1009)$, $y = y \mod (1009)$.

We construct the quasigroup $x \circ y = 2657x + 7063y \mod (9721)$ and check that in this quasigroup, 3-rd Stein identity is fulfilled: $2657(2657x+7063y)+7063(2657y+7063x) = y \mod (9721)$, $7059649x+8766391y + 8766391y + 9885969x = y \mod (9721)$, $56945618x + 37532782y = y \mod (9721)$, $y = y \mod (9721)$.

We have constructed the following simple medial quasigroups with identity $xy \cdot yx = y$ (see Table 1).

3 Conclusion

In this paper, we research simple medial quasigroups with the identity $xy \cdot yx = y$ (3-rd Stein identity).

Acknowledgments. We were supported in the frame of projects 20. 80009.5007.22 "Intelligent information systems for solving weak-structured problems, processing knowledge and big data"and "Algebraic systems with additional structures, theory quasigroups and their application in coding theory".

Table 1. Simple quasigroups with identity $xy \cdot yx = y$

(1, 3, 5);	(2, 10, 13);	(6, 10, 17);	(8, 20, 29);	(15, 21, 37);	(4, 36, 41);
(11, 41, 53);	(5, 55, 61);	(13, 59, 73);	(27, 61, 89);	(37, 59, 97);	(45, 55, 101);
(7, 105, 113);	(18, 118, 137);	(52, 96, 149);	(64, 92, 157);	(46, 126, 173);	(9, 171, 181);
(40, 152, 193);	(91, 105, 197);	(53, 175, 229);	(44, 188, 233);	(88, 152, 241);	(120, 136, 257);
(93, 175, 269);	(26, 254, 281);	(77, 215, 293);	(12, 300, 313);	(101, 215, 317);	(94, 242, 337);
(106, 242, 349);	(155, 197, 353);	(134, 238, 373);	(57, 331, 389);	(31, 365, 397);	(190, 210, 401);
(71, 337, 409);	(89, 343, 433);	(33, 415, 449);	(54, 402, 457);	(206, 254, 461);	(150, 358, 509);
(117, 403, 521);	(244, 296, 541);	(219, 337, 557);	(241, 327, 569);	(276, 300, 577);	(38, 554, 593);
(62, 538, 601);	(17, 595, 613);	(211, 405, 617);	(243, 397, 641);	(74, 578, 653);	(277, 383, 661);
(307, 365, 673);	(325, 351, 677);	(67, 633, 701);	(306, 402, 709);	(176, 556, 733);	(43, 713, 757);
(19, 741, 761);	(353, 415, 769);	(158, 614, 773);	(107, 689, 797);	(245, 563, 809);	(147, 673, 821);
(291, 537, 829);	(166, 686, 853);	(103, 753, 857);	(75, 801, 877);	(193, 687, 881);	(302, 626, 929);
(370, 566, 937);	(48, 892, 941);	(255, 697, 953);	(362, 614, 977);	(80, 916, 997)	(234, 774, 1009).

References

[1] V.D. Belousov. *Foundations of the Theory of Quasigroups and Loops*. Nauka, Moscow, 1967. (in Russian).

[2] V.D. Belousov. *Parastrophic-orthogonal quasigroups.* Translated from the 1983 Russian original. Quasigroups Relat. Syst., 13, no. 1, 2005, pp. 25–72.

[3] G. Birkhoff. *Lattice Theory*, Nauka, Moscow, 1984, (in Russian).

[4] R.H. Bruck. *A Survey of Binary Systems*. Springer Verlag, New York, third printing, corrected edition, 1971.

[5] S. Burris and H.P. Sankappanavar. *A Course in Universal Algebra,* Springer-Verlag, 1981.

[6] E. Schröder. *Über eine eigenthümliche Bestimmung einer Function durch formale Anforderungen*. J. Reine Angew. Math., 90:189-220, 1881.

[7] T. Kepka and P. Němec. *T-quasigroups, II*, Acta Univ. Carolin. Math. Phys., 12, no.2, 1971, p. 31-49.

[8] H.O. Pflugfelder. *Quasigroups and Loops: Introduction*. Heldermann Verlag, Berlin, 1990.

[9] A. Sade. *Quasigroupes obéissant á certaines lois*. Rev. Fac. Sci. Univ. Istambul, 22, 1957, pp. 151–184.

[10] Victor Shcherbacov. *Elements of Quasigroup Theory and Applications*. CRC Press, Boca Raton, 2017.

[11] V. Shcherbacov. *Schroder T-quasigroups*, arXiv:2206.12844, 2022, 13 p.

[12] Sh. K. Stein. *Homogeneous quasigroups*, Pacific J. Math., 14, 1964, pp. 1091–1102.

Victor Shcherbacov[1], Valentina Demidova[2], Petr Radilov[3],

[1] Victor Shcherbacov
Researcher/Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: victor.scerbacov@math.md

[2] Valentina Demidova,
Researcher/Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: valentina.demidova@math.md

[3] Petr Radilov,
Ph.D. Student/Moldova State University
E–mail: illusionist.nemo@gmail.com

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

T-quasigroups with Stein 3-rd law

Victor Shcherbacov, Irina Shvedyuk, Nadeghda Malyutina

Abstract

We research T-quasigroups with Stein 3-rd law (identity).
2000 Mathematics Subject Classification: 20N05
Keywords: quasigroup, Stein quasigroup, T-quasigroup, medial quasigroup.

1 Introduction

This paper is connected in main with the following **Belousov problem 1 [2, 6]**: Find a complete characterization of groups isotopic to quasigroups that satisfy one of the identities: $x(y \cdot yx) = y$ (of type T_2, [2, 10]), $xy \cdot yx = x$ (Schröder 2nd law, [9]), $xy \cdot yx = y$ (Stein 3rd law, [12], [9]).

These identities guarantee that a quasigroup is orthogonal to its parastrophe [2, 10].

Necessary definitions can be found in [1, 4, 8, 10, 11].

Definition 1. Binary groupoid $(Q, \circ)$ is called a left quasigroup if for any ordered pair $(a, b) \in Q^2$ there exists the unique solution $x \in Q$ to the equation $a \circ x = b$ [3, 4, 8, 1].

Definition 2. Binary groupoid $(Q, \circ)$ is called a right quasigroup if for any ordered pair $(a, b) \in Q^2$ there exists the unique solution $y \in Q$ to the equation $y \circ a = b$ [3, 4, 8, 1].

Definition 3. Binary groupoid $(Q, \circ)$ is called a quasigroup if it is a right and left quasigroup [3, 4, 8, 1, 5, 10].

Definition 4. Binary groupoid $(Q, \cdot)$ is called medial if this groupoid satisfies the following medial identity:

$$xy \cdot uv = xu \cdot yv \tag{1}$$

for all $x, y, u, v \in Q$ [1, 10].

We recall

Definition 5. Quasigroup $(Q, \cdot)$ is a T-quasigroup if and only if there exists an abelian group $(Q, +)$, its automorphisms φ and ψ, and a fixed element $a \in Q$ such that $x \cdot y = \varphi x + \psi y + a$ for all $x, y \in Q$ [7].

A T-quasigroup with the additional condition $\varphi\psi = \psi\varphi$ is medial [1, 10].

2 T-quasigroups with Stein 3-rd identity

Theorem 1. *In T-quasigroup $(Q, \cdot)$ of the form $x \cdot y = \varphi x + \psi y$, Stein 3-rd identity is true if and only if $\varphi^2 + \psi^2 = 0$, $\varphi\psi y + \psi\varphi y = \varepsilon$.*

Proof. From identity

$$xy \cdot yx = y \tag{2}$$

we have

$$\varphi(\varphi x + \psi y) + \psi(\varphi y + \psi x) = y. \tag{3}$$

If in (3) $y = 0$, then

$$\varphi^2 + \psi^2 = 0, \tag{4}$$

$$\varphi^2 = I\psi^2. \tag{5}$$

If in (3) $x = 0$, then

$$\varphi\psi + \psi\varphi = \varepsilon. \tag{6}$$

Converse. If we substitute in identity (2) the expression $x \cdot y = \varphi x + \psi y$, then we obtain equality (3), which is true taking into consideration equalities (4), (6). Then we obtain that identity (2) is true in this case. □

Corollary 1. *In medial quasigroup $(Q,\cdot)$ of the form $x\cdot y=\varphi x+\psi y$, Stein 3-rd identity is true if and only if $\varphi^2+\psi^2=0$, $2\varphi\psi=\varepsilon$.*

Proof. This follows from mediality of the quasigroup $(Q,\cdot)$. □

By the symbol Z_n we denote the group of residues modulo n.

Corollary 2. *In medial quasigroup $(Z_n,\cdot)$ of the form $x\cdot y=\varphi x+\psi y$, Stein 3-rd identity is true if $\varphi+\psi+1=0$ or $\varphi+\psi-1=0$.*

Proof. From Corollary 1 it follows that $(\varphi+\psi)^2=1 \pmod n$, $\varphi+\psi=\pm1 \pmod n$. Therefore, $\varphi+\psi-1=0 \pmod n$ or $\varphi+\psi+1=0 \pmod n$. □

Example 1. If $\varphi=4$, $\psi=2$, then $\varphi^2+\psi^2=16+4=0 \pmod 5, n=5$. Further, $2\varphi\psi=2\cdot4\cdot2=16=\varepsilon=1 \pmod 5$, $x\cdot y=4x+2y \pmod 5$.

Check. $4(4x+2y)+2(4y+2x)=y \pmod 5$, $16x+8y+8y+4x=y \pmod 5$, $y=y \pmod 5$.

Example 2. Suppose we have the group Z_n of residues modulo n. If $\varphi=2$, $\psi=10$, then $\varphi^2+\psi^2=4+100=104=0 \pmod{13}, n=13$. Further, $2\varphi\psi=2\cdot2\cdot10=40=\varepsilon=1 \pmod{13}$, $x\cdot y=2x+10y \pmod{13}$.

Check. $2(2x+10y)+10(2y+10x)=y \pmod{13}$, $4x+20y+20y+100x=y \pmod{13}$, $y=y \pmod{13}$.

Example 3. Suppose we have the group Z_n of residues modulo n. If $\varphi=11$, $\psi=3$, then $\varphi^2+\psi^2=121+9=130=0 \pmod{13}, n=13$. Further, $2\varphi\psi=2\cdot3\cdot11=66=\varepsilon=1 \pmod{13}$, $x\cdot y=11x+3y \pmod{13}$.

Check. $11(11x+3y)+3(11y+3x)=y \pmod{13}$, $121x+33y+33y+9x=y \pmod{13}$, $y=y \pmod{13}$.

Example 4. Suppose we have the group Z_n of residues modulo n. If $\varphi=6$, $\psi=10$, then $\varphi^2+\psi^2=36+100=136=0 \pmod{17}, n=17$. Further, $2\varphi\psi=2\cdot6\cdot10=120=\varepsilon=1 \pmod{17}$, $x\cdot y=6x+10y \pmod{17}$.

Check. $6(6x+10y)+10(6y+10x)=y \pmod{17}$, $36x+60y+60y+100x=y \pmod{17}$, $y=y \pmod{17}$.

Example 5. Suppose we have the group Z_n of residues modulo n. If $\varphi = 7$, $\psi = 11$, then $\varphi^2 + \psi^2 = 49 + 121 = 170 = 0 \pmod{17}, n = 17$. Further, $2\varphi\psi = 2 \cdot 7 \cdot 11 = 154 = \varepsilon = 1 \pmod{17}$, $x \cdot y = 7x + 11y \pmod{17}$.

Check. $7(7x+11y)+11(7y+11x) = y \pmod{17}$, $49x+77y+77y+121x = y \pmod{17}$, $y = y \pmod{17}$.

3 Conclusion

In this paper, we research T-quasigroups with the identity $xy \cdot yx = y$ (3-rd Stein identity).

Acknowledgments. We were supported in the frame of projects 20. 80009.5007.22 "Intelligent information systems for solving weak-structured problems, processing knowledge and big data"and "Algebraic systems with additional structures, theory quasigroups and their application in coding theory".

References

[1] V.D. Belousov. *Foundations of the Theory of Quasigroups and Loops*. Nauka, Moscow, 1967. (in Russian).

[2] V.D. Belousov. *Parastrophic-orthogonal quasigroups*. Translated from the 1983 Russian original. Quasigroups Relat. Syst., 13, no. 1, 2005, pp. 25–72.

[3] G. Birkhoff. *Lattice Theory*, Nauka, Moscow, 1984, (in Russian).

[4] R.H. Bruck. *A Survey of Binary Systems*. Springer Verlag, New York, third printing, corrected edition, 1971.

[5] S. Burris and H.P. Sankappanavar. *A Course in Universal Algebra*. Springer-Verlag, 1981.

[6] E. Schröder. *Über eine eigenthümliche Bestimmung einer Function durch formale Anforderungen*. J. Reine Angew. Math., 90: pp. 189–220, 1881.

[7] T. Kepka and P. Němec. *T-quasigroups, II.* Acta Univ. Carolin. Math. Phys., 12, no.2, 1971, pp. 31–49.

[8] H.O. Pflugfelder. *Quasigroups and Loops: Introduction*. Heldermann Verlag, Berlin, 1990.

[9] A. Sade. *Quasigroupes obéissant á certaines lois*. Rev. Fac. Sci. Univ. Istambul, 22, 1957, pp. 151–184.

[10] Victor Shcherbacov. *Elements of Quasigroup Theory and Applications*. CRC Press, Boca Raton, 2017.

[11] V. Shcherbacov. *Schroder T-quasigroups*. arXiv:2206.12844, 2022, 13 p.

[12] Sh. K. Stein. *Homogeneous quasigroups*. Pacific J. Math. vol. 14, 1964, pp. 1091–1102.

Victor Shcherbacov[1], Irina Shvedyuk[2], Nadeghda Malyutina[3]

[1] Victor Shcherbacov,
Researcher/Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: victor.scerbacov@math.md

[2] Irina Shvedyuk,
Ph.D. Student/Moldova State University
E–mail: ira230396@mail.ru

[3] Nadeghda Malyutina,
Ph.D. Student/Moldova State University
E–mail: 231003.bab.nadezhda@mail.ru

Proceedings of Workshop on Intelligent Information Systems
WIIS2022, October 06-08, 2022, Chisinau, Republic of Moldova

Models of attacks against power grids: a short survey

Volodymyr G. Skobelev, Volodymyr V. Skobelev

Abstract

Modern power grids are critical infrastructures that include interacting networks of physical and computational components. In other words, they are critical smart grids. Therefore, they face serious safety and cyber-security problems. Failures in real power grids and numerous successful attacks on these grids, resulting in large economical losses, put forward the development of protection means against them as the main problems. The difficulty of solving these problems is due to the permanent emergence of new increasingly sophisticated attacks types against power grids. This results in the permanent development of protecting the power grid means from new types of attacks. The main aim of the given paper is a short survey of mathematical models for the most dangerous attacks against modern power grids and known protection schemes against these attacks.

Keywords: Power grids, safety, security, cyber attacks, detection, protection.

MSC 2020: 68M15, 93A15, 93B30, 93C83.

1 Introduction

Modern Smart Power Grids (SPGs) are critical infrastructures with the load controller (LC) for balancing the power supply due to the demand while maintaining nominal current frequency. Their main feature is the presence of numerous remotely controlled high-wattage IoT devices, e.g., water heaters, washing machines, air conditioning units, electric

ovens. So, any SPG can be considered as some Cyber-Physical System (CPS), for which safety and cyber security are the most pressing issues.

The main causes of SPGs safety vulnerability that lead to large blackouts are poor maintenance policy and/or mismanagement (see [1], for example). A single generator failure has led to the largest blackout in the US and Canada (August, 2003) [2]. Tripping tie-lines between Switzerland and Italy due to an overload have led to rather large blackout in Italy (September, 2003) [3]. Negligent maintenance of high voltage insulators on two transmission lines has led to blackout in Brazil (September, 2007) [4]. In the series of blackouts in Brazil (2010-2019) [5], conflicts of interests have played a significant role. A single high voltage line failure has led to the blackout in Arizona and Southern California (September 8, 2011) [6]. Failures in three transmission lines have led to the blackout in the South Australia (September, 2016) [7].

It should be noted that the main approach to SPG safety is based on the application of Reliability Theory methods and (probabilistic) mathematical models designed to assess the quality of maintenance.

The situation with SPGs cyber security is much more complicated due to the new types of these attacks emergence, the growth of their intensity, sophistication and scale. The variety of successful attacks types on SPGs is caused by various techniques and can be illustrated as follows (see [8], [9], for example).

1. Attacks via malware. The attack against the nuclear power plant in Ohio, US (January 2003) via infection of the third-party company by the Slammer computer worm [10]. The attack against the Brazilian power plant (February 2011) via infection it by the Conficker worm [11]. The attack via WannaCry ransomware against India's West Bengal State Electricity Distribution Company (May 2017) that caused bill-payment operations to be suspended for most of a day until backed-up data could be restored [12].

2. Attacks via denial of service. The attack that shut down for part of a day the website of the Brazilian energy company Petrobras - Latin America's (April 2020) [13]. The attack against the Ukrainian power load controller (December 2015) [8, 14] is considered as a serious

precedent for the safety of SPGs around the world. The attack against not named energy company that provides power in several western US states (March 2019) has resulted in interruptions of electrical system operations for more than 10 hours [15].

3. Attacks via false data injection. The attack against the Metcalf Power substation in San Jose, US (April 2013) by firing at high voltage transformers for 20 minutes [16].

4. Attacks by insider intruders. The attack by a technician against the Ignalina, Lithuania nuclear power station (1992) via introducing some virus into the control system of one reactor [8].

5. Attacks via zero day vulnerabilities. The attack against the Natanz, Iran nuclear center (2010), when the Stuxnet worm has damaged 984 centrifuges [17].

This list can be continued since the number of attacks against SPGs and the variety of attack vectors and their scenarios increase significantly each year.

A number of attempts have been made to classify the existing SPGs attacks (see [8], [18], for example). These classifications are based on the choice of the attack's target, the attack's objects, the level of the SPG structure abstraction, the time of the attack, etc.

Let's briefly consider the main models of attacks on SPGs and the proposed methods for detecting and protecting from these attacks.

2 Attacks against LC

In any SPG, the LC performs automated command control of SPG operation and it is also a communication channel between control applications and field devices by transmitting measurement and control signals [19].

The LC computer systems themselves are equipped with standard cryptographic means of protection against attacks. Therefore, their cyber security is completely determined by these means.

The main causes of vulnerabilities in any LC are heterogeneous and non-proprietary technologies used in it (see [20], for example).

To propect LC against cyber attacks, different Intrusion Detection Systems (IDSs) have been developed. To identify the cyber attack and assess its consequences, data from the IDS are compared with similar data for models of these attacks. Based on this comparison, some set of correlation indices (CIs) is calculated. This set of CIs is used for estimation the temporal and/or spatial characteristics of any given cyber attack. Thus, the possibility of any given cyber attack can be assessed.

Currently, several sets of CIs have been built. They are based on different principles, are designed to analyze different types of cyber attacks and use different levels of the analyzed network's abstraction. The various known sets of CIs and the problems for which they are intended are briefly described in [21].

It should be noted that any IDS can only detect regular attacks or some of their components. However, any IDS, regardless of the used set of CIs, is prone to false alarms, cannot identify some of the most dangerous cyber attacks, and does not provide the possibility to evaluate the results of cyber attacks on the analyzed SPG.

In [22], [23], an analysis of possible cyber attacks against LCs used in modern SPGs is presented and protection schemes against these attacks are considered.

3 Coordinated cyber attacks against SPGs

These attacks are often used to disrupt energy provision via imitation or realization of sudden spikes or drops in energy consumption resulting in SPGs equipment outages.

The danger of coordinated cyber attacks (CCAs) is that it is difficult to both detect them and protect SPGs from them. The spread of CCAs is due to simple and inexpensive access to botnets, i.e., networks of infected devices that can be remotely and synchronously controlled by an attacker.

A number of studies have been devoted to the CCAs simulation and the development of means for their detection and their consequences

leveling. Let's consider them briefly.

It has been shown in [24] via Matlab simulation that the balance between energy supply and consumption to maintain the nominal frequency of European SPG can be easily broken by a botnet attack. In doing so, the following three types of CCAs depending on the chosen object for the attack have been investigated:

1. Static load CCAs that target the SPG primary control via sudden increasing the power consumption of all bots to the maximum.

2. Dynamic load CCAs that target the behavior of the primary control via sudden increasing of the load to the maximum, waiting for the full primary control to be activated, then sudden decreasing of the load to the minimum, waiting for the primary control to be deactivated, and so on.

3. Inter-zone CCAs, the aim of which is to disconnect some communication line operating near the maximum, increasing the load on it via increasing the power consumption in the target area of this transfer.

The concept of manipulation on demand via IoT has been introduced in [1]. Based on this concept, the following three types of CCAs, depending on their goals, have been investigated:

1. CCAs that result in frequency instability via synchronous switching on or off many high-wattage IoT devices that lead to an imbalance between the supply and demand in the SPG.

2. CCAs that cause line failures and result in cascading outages via increasing the demand in some areas and decreasing it in others.

3. CCAs that increase operating costs via exceeding the projected day-ahead demand, since the SPG operator must purchase additional power from reserve generators, which typically have higher prices than the generators planned for the day.

The effectiveness of these attacks on real-world power grid models has been demonstrated in [1] via using the MATPOWER and the Power-World simulators.

It should be noted that the set of CIs proposed in [21] have been developed precisely taking into account the possibility of using them to detect and identify CCAs against SPGs. This is confirmed in [25],

where the corresponding IDS has been proposed.

Much attention has been paid to the development of stochastic models designed to analyze the management dynamics in order to ensure a balance between the supply of electricity and its demand in SPGs.

In [26], a solution to this problem has been proposed via using discrete-time Markov chains (DTMCs) and continuous-time Markov chains (CTMCs) simulated in the probabilistic model checker PRISM.

In [27], flooding DoS attacks against grids have been modeled and simulated via a 3-dimensional CTMC with block tridiagonal generator matrix that accounts the environment and admits random dropping for the grid's defense.

Noteworthy is the approach to CCAs modeling and simulation on the base of the probabilistic model checker PRISM, proposed in [28]. A distinctive feature of this approach is a very high abstraction level regarding the structure of the analyzed CPG, in terms of which scenarios for three types of CCAs against the SPG in Glasgow area, UK have been simulated.

Let's consider this approach in more detail.

It is assumed that the functioning of the SPG is provided by three types of power generators, namely nuclear, hydro, and gas ones.

The basic idea is that the behavior of each power generator PG, as well as the behavior of the LC and the botnet in relation to it, is modeled via the CTMC designed as follows.

The set of states for any power generator PG is

$$S_{PG} = \{a, g, r\},$$

where a (available), i.e., "ready to supply but not yet generating power into the SPG"; g (generating), i.e., "currently providing power to the SPG"; and r (restart), i.e., "detached from the SPG, not generating, nor supplying, but needed to be restarted".

The set of states for the power demand provided by this power generator is

$$S_D = \{l, m, h\},$$

where l is "low", m is "medium", and h is "high".

The set of states for the spike botnet B relatively to this power generator is

$$S_B = \{0, 1\},$$

where 0 is "all infected devices controlled by the botnet and influencing the power generator PG are switched off", and 1 is "all infected devices controlled by B and affecting the power generator PG are switched on".

Based on these state sets, to each power generator PG is assigned the marked CTMC

$$\mathcal{C}_{PG} = (\mathcal{S}, s_0, R_{PG}, L_{PG}),$$

where $\mathcal{S} = S_{PG} \times S_D \times S_B$ is the set of states, $s_0 = (a, m, 0)$ is the initial state, R_{PG} is the rate matrix, and

$$L_{PG} : S_{PG} \times S_{PG} \to 2^{\mathcal{P}}$$

is the state transitions marking for the power generator PG by atomic sentences from a given set of sentences $\mathcal{P}$ as follows:

$$L_{PG}(r \to a) = \begin{cases} \text{“}very\ slow\ restart\text{”}, & for\ a\ nuclear\ PG \\ \text{“}fast\ restart\text{”}, & for\ a\ hydro\ PG \\ \text{“}slow\ restart\text{”}, & for\ a\ gas\ PG \end{cases},$$

$$L_{PG}(a \to g) = \begin{cases} \text{“}slow\text{”}, & for\ a\ nuclear\ PG \\ \text{“}fast\text{”}, & for\ a\ gas\ PG \end{cases}.$$

All other state transitions marks in any marked CTMC $\mathcal{C}_{PG}$ are the empty sentence Λ.

In [28], when defining state transition marks, fuzzy concepts "very slow", "fast", and "slow" are used. This can still be understood for the case study used in [28]. However, such labels are inexplicable for a formal model applicable to simulate any SPG. Apparently, the way out of this situation is to use DTMC, in which each state transition with a non-empty label is replaced by a sequence of transitions of the appropriate length.

It should be noted that the approach proposed in [28], taking into account the above comment, can be generalized to the case of k ($k > 3$) power generators types as follows.

Let the analyzed SPG $\mathcal{Z}$ be presented by the partition

$$\pi_{\mathcal{Z}} = \{\{PG_1^{(1)}, \ldots, PG_{l_1}^{(1)}\}, \ldots, \{PG_1^{(k)}, \ldots, PG_{l_k}^{(k)}\}\},$$

where $PG_j^{(i)}$ $(i = 1, \ldots, k; j = 1, \ldots, l_i)$ is the j-th power generator of the type i.

Then the DTMCs associated with the SPG $\mathcal{Z}$ can be presented by the partition

$$\rho_{\mathcal{Z}} = \{\{\mathcal{C}_1^{(1)}, \ldots, \mathcal{C}_{l_1}^{(1)}\}, \ldots, \{\mathcal{C}_1^{(k)}, \ldots, \mathcal{C}_{l_k}^{(k)}\}\},$$

where $\mathcal{C}_j^{(i)}$ $(i = 1, \ldots, k; j = 1, \ldots, l_i)$ is the DTMC associated with the power generator $PG_j^{(i)}$.

The current state of the SPG $\mathcal{Z}$ is $\mathbf{s} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$, where

$$\mathbf{a} = (\alpha_{11}, \ldots, \alpha_{1l_1}, \ldots, \alpha_{k1}, \ldots, \alpha_{kl_k}),$$

$$\mathbf{b} = (\beta_{11}, \ldots, \beta_{1l_1}, \ldots, \beta_{k1}, \ldots, \beta_{kl_k}),$$

$$\mathbf{c} = (\gamma_{11}, \ldots, \gamma_{1l_1}, \ldots, \gamma_{k1}, \ldots, \gamma_{kl_k}),$$

and

$$\alpha_{ij} \in \{a, g, r\} \ (i = 1, \ldots, k; j = 1, \ldots, l_i),$$

$$\beta_{ij} \in \{l, m, h\} \ (i = 1, \ldots, k; j = 1, \ldots, l_i),$$

$$\gamma_{ij} \in \{0, 1\} \ (i = 1, \ldots, k; j = 1, \ldots, l_i).$$

For the transition of the SPG $\mathcal{Z}$ from the state $\mathbf{s} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$ into the state $\mathbf{s\prime} = (\mathbf{a}', \mathbf{b}', \mathbf{c}')$, we get:

1. The time needed for this states transition is

$$T_{\mathbf{s} \Rightarrow \mathbf{s}'} = \max\{t_{ij} | i = 1, \ldots, k; j = 1, \ldots, l_i\},$$

where t_{ij} is the time needed for the transition $\alpha_{ij} \Rightarrow \alpha'_{ij}$.

2. The cost of this states transition is

$$C_{\mathbf{s}\Rightarrow\mathbf{s}'} = \sum_{i=1}^{k}\sum_{j=1}^{l_i} c_{ij},$$

where c_{ij} is the cost of the transition $\alpha_{ij} \Rightarrow \alpha'_{ij}$ plus the cost of the transition $\beta_{ij} \Rightarrow \beta'_{ij}$.

3. The power supply variation in the SPG for this transition of states is

$$\Delta W_{\mathbf{s}\Rightarrow\mathbf{s}'} = \sum_{\{(i,j)|\alpha_{ij}=g\}} W_{ij}(\mathbf{s}) - \sum_{\{(i,j)|\alpha'_{ij}=g\}} W_{ij}(\mathbf{s}'),$$

where $W_{ij}(\mathbf{s})$ (correspondingly, $W_{ij}(\mathbf{s}')$) is the power supply provided by the power generator $PG_j^{(i)}$ in the state $\mathbf{s}$ (correspondingly, $\mathbf{s}'$).

4. The power demand variation in the SPG for this transition of states is

$$\Delta D_{\mathbf{s}\Rightarrow\mathbf{s}'} = \sum_{\{(i,j)|\alpha_{ij}=g\}} D_{ij}(\mathbf{s}) - \sum_{\{(i,j)|\alpha'_{ij}=g\}} D_{ij}(\mathbf{s}')),$$

where $D_{ij}(\mathbf{s})$ (correspondingly, $D_{ij}(\mathbf{s}')$) is the demand of the power suply produced by the power generator $PG_j^{(i)}$ in the state $\mathbf{s}$ (correspondingly, $\mathbf{s}'$).

5. The spike botnet B influence variation in the SPG for this state transition of states is

$$\Delta B_{\mathbf{s}\Rightarrow\mathbf{s}'} = \sum_{\{(i,j)|\alpha_{ij}=g\}} \gamma_{ij} - \sum_{\{(i,j)|\alpha'_{ij}=g\}} \gamma'_{ij}.$$

Thus, for any state $\mathbf{s}$ of the given SPG, the selection of the optimal state transition $\mathbf{s} \Rightarrow \mathbf{s}'$, such that

$$\sum_{i=1}^{k}\sum_{j=1}^{n_k} W_{ij}(\mathbf{s}') = \sum_{i=1}^{k}\sum_{j=1}^{n_k} D_{ij}(\mathbf{s}'),$$

can be carried out on the basis of solving the following multi-criteria problem

$$\begin{cases} T_{\mathbf{s}\Rightarrow\mathbf{s}'} \to \min \\ C_{\mathbf{s}\Rightarrow\mathbf{s}'} \to \min \\ |\Delta W_{\mathbf{s}\Rightarrow\mathbf{s}'}| \to \text{extr} \\ |\Delta D_{\mathbf{s}\Rightarrow\mathbf{s}'}| \to \text{extr} \\ \Delta B_{\mathbf{s}\Rightarrow\mathbf{s}'} \to \max \end{cases},$$

where the extr types depend on the sign of the difference

$$\sum_{\{(i,j)|\alpha_{ij}=g\}} W_{ij}(\mathbf{s}) = \sum_{\{(i,j)|\alpha_{ij}=g\}} D_{ij}(\mathbf{s}).$$

If, due to the size of the analyzed SPG, the search for a solution to this problem is not feasible, then its approximate solution can be found and its quality assessed.

4 Dynamic load altering attacks against SPGs

These attacks are aimed to change the SPG state estimation via injecting false data into power sensors and/or routers that relay the data from these sensors to LC. For such attacks to succeed, an attacker needs detailed information about the SPG configuration.

Dynamic load altering attacks (DLAAs) against SPGs has been presented firstly in [29]. This research is based on the following linear models.

The SPG state estimation is presented as the vector $\mathbf{z} = H\mathbf{x} + \mathbf{e}$, where H is the measurement matrix, $\mathbf{x}$ is the state variables measurements, and $\mathbf{e}$ is the measurement errors. When the meter error is normally distributed with the zero mean, the measurement residual is $\mathbf{z} - H\hat{\mathbf{x}}$, where $\hat{\mathbf{x}} = (H^T W H)^{-1} H W \mathbf{z}$, and W is the diagonal matrix whose diagonal elements are reciprocals of the meter errors variances. The presence of bad measurements is assumed if $\|\mathbf{z} - H\hat{\mathbf{x}}\| > \tau$, where τ is the given threshold. It is supposed that in the presence of DLAA,

the observed measurements can be presented as $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$, where $\mathbf{a}$ is the attack vector, and the estimates $\hat{\mathbf{x}}_{bad}$ of $\mathbf{x}$ for the malicious measurements $\mathbf{z}_a$ can be presented as $\hat{\mathbf{x}}_{bad} = \mathbf{x} + \mathbf{c}$, where $\mathbf{c}$ is a non-zero vector. Under these suppositions, if $\mathbf{a} = H\mathbf{c}$, then $\mathbf{z}_a$ can pass the detection as long as $\mathbf{z}$ can pass it (see Theorem 1 in [29]).

The following two types of DLAAs has been analyzed in [29]:

1. Random false data injection attacks, i.e., the attacker tries to find any attack vector that can lead to a wrong state variables estimation.

2. Targeted false data injection attacks, i.e., the attacker tries to find an attack vector to inject any errors into certain state variables.

Subsequently, the following three types of restrictions have been identified that affect the possibility of successful implementation of DLAAs:

1. Limited attacker resources for the number of compromising measurements of the SPG [30].

2. The extent to which an attacker has access to measurements in the SPG that can potentially be compromised [31].

3. The level of an attacker information completeness about the SPG measurement matrix [32].

It should be noted that the following three aspects for enhancing the known DLAAs detection methods have been investigated in [31]:

1. Attacker versus defender dynamics, i.e., possible interactive attacks and defense strategies are analyzed from secure phasor measurement unit location.

2. Distributed attack detection and state recovery, i.e., the ways for achieving the optimal centralized performance are analyzed for a distributed approach.

3. The quickest detection, i.e., a method for searching a trade-off between detection speed and detection performance is proposed.

Further DLAAs studies tend to be based on the results presented above. Among them, the following study should be pointed out.

In [33], dynamic DLAAs have been investigated based on the analysis of change amount in compromised load along the trajectory where this load changes, and a protection scheme against these attacks via

solving a non-convex pole-placement optimization problem has been proposed.

Reviews of the state-of-the-art studies of DLAAs against SPGs and protection schemes against these attacks are presented in [34]-[36].

5 Concluding remarks

This paper provides a brief overview of studies of the most dangerous coordinated attacks against modern SPGs and schemas designed to detect these attacks and mitigate their consequences. As noted above, new types of these attacks are emerging; their intensity, complexity, and scale are growing. Reviews [37] - [39] are devoted to the analysis of the current state of the SPGs cyber-security problem.

References

[1] S. Soltan, P. Mittal, and H. V. Poor. "BlackIoT: IoT botnet of high wattage devices can disrupt the power grid," In Proc. *The 27th USENIX Security Symposium*, Baltimore, MD, USA, 2018, pp. 15–32.

[2] Office of Electricity, "Blackout 2003: Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," April 2, 2004. Available: https://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf

[3] Union for the Coordination of the transmission of electricity (UCTE), "Final report of the investigation committee on the 28 September 2003 blackout in Italy," April 2004. Available: http://www.rae.gr/old/cases/C13/italy/UCTE_rept.pdf

[4] "Brazilian blackout traced to sooty insulators, not hackers," *Wired*, no. 11, November 9, 2009. Available: https://www.wired.com/2009/11/brazil-blackout/

[5] B. Alves. "Average number of electricity outages per consumer unit in Brazil from 2010 to 2019," *Statista*, July 2, 2021. Available: https://www.statista.com/statistics/987917/number-electricity-outages-brazil/

[6] Federal Energy Regulatory Commission and North American Electric Reliability Corporation, "Arizona – Southern California outages on September 8, 2011: causes and recommendations," Washington, D.C., April 27, 2012. Available: http://www.ferc.gov/legal/staff-reports/04-27-2012-ferc-nerc-report.pdf

[7] Australian Energy Market Operator (AEMO), "Black system South Australia 28 September 2016: final report," 23 March, 2017. Available: https://www.aemo.com.au/-/media/Files/Electricity/NEM/Market_Notices_and_Events/Power_System_Incident_Reports/2017/Integrated-Final-Report-SA-Black-System-28-September-2016.pdf

[8] G. Desarnaud. "Cyber attacks and energy infrastructures. Anticipated risks," *IFRI Center for Energy*, vol. 48, Issue 39, RN: 48078577, January, 2017.

[9] N. Kshetri and J. Voas. "Hacking Power Grids: A Current Problem," *Computer*, vol. 50, no. 12, pp. 91–95, December 2017, Available: https://ieeexplore.ieee.org/document/8220480.

[10] K. Poulsen. "Slammer worm crashed Ohio nuke plant network," *SecurityFocus*, Mountain View, California, August 19, 2003. Available: http://www.securityfocus.com/news/6767

[11] R. McMillan. "A power plant hack that anybody could use," *IDG News Service*, August 4, 2011. Available: https://www.computerworld.com/article/2509910/a-power-plant-hack-that-anybody-could-use.html?page=2.

[12] "WannaCry ransomware: Computers at West Bengal State electricity firm hit," *The Indian Express*, May 15, 2017. Available: https://indianexpress.com/article/technology/tech-news-technology/wannacry-ransomware-computers-at-west-bengal-state-electricity-firm-hit/.

[13] K. Eiden, E. Goraieb, K. Nobels, and D. Wallance. "The Latin American energy sector: How to address cybersecurity," *McKinsey&Company*, January 12, 2021. Available: https://www.mckinsey.com/business-functions/risk-and-resilience/our-insights/the-latin-american-energy-sector-how-to-address-cybersecurity.

[14] R. M. Lee, M. J. Assante, and T. Conway. "Analysis of the cyber attack on the ukrainian power grid," *SANS Industrial Control Systems*, March 18, 2016. Available: file:///C:/Users/admin/Downloads/E-ISAC_SANS_Ukraine_DUC_5.pdf.

[15] Z. Whittaker. "Denial of service condition' disrupted US energy company operations," *TechCrunch*, May 2, 2019. Available: https://techcrunch.com/2019/05/02/ddos-attack-california-energy/.

[16] B. Whitaker. "Vulnerable U.S. electric grid facing threats from Russia and domestic terrorists," *CBS News*, February 27, 2022. Available: https://www.cbsnews.com/news/america-electric-grid-60-minutes-2022-02-27/.

[17] B. Kesler. "The vulnerability of nuclear facilities to cyber attack," *Strategic Insights*, vol. 10, Issue 1, pp. 15–25, 2011.

[18] T. Krause, R. Ernst, B. Klaer, I. Hacker, and M. Henze. "Cybersecurity in Power Grids: Challenges and Opportunities," *Sensors*, vol. 21, Issue 6225, 2021. Available: https://doi.org/10.3390/s21186225.

[19] B. Zhu, A. Joseph, and S. Sastry. "A taxonomy of Cyber-attacks on SCADA systems," In Proc. *Internet of Things, 4th IEEE International Conference on Cyber, Physical and Social Computing*, Dalian, China, 2011, pp. 380–388.

[20] C. Liu, A. Stefanov, J. Hong, and P. Panciatici. "Intruders in the grid," *IEEE Power and Energy magazine*, vol. 10, no. 1, pp. 58–66, 2012.

[21] C. Moya and J. Wang. "Developing correlation indices to identify coordinated cyber-attacks on power grids." *IET Cyber-Physical Systems: Theory & Applications*, vol. 3, no. 4, pp. 178–186, 2018.

[22] Y. Zhang, L. Wang, Y. Xiang, and C.-W. Ten. "Power system reliability evaluation with SCADA cybersecurity considerations." *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1707–1721, 2015.

[23] Q. S. Qassim, N. Jamil, M. N. Mahdi, Z. C. Cob, F. A. Rahim, and L. M. Sidek. "A Short review: issues and threats pertaining the security of SCADA systems." In: N. Abdullah, S. Manickam, M. Anbar (eds). *Advances in cyber security. ACeS 2021. Communications in Computer and Information Science*, vol 1487. Springer, Singapore, pp. 230–247, 2022.

[24] A. Dabrowski, J. Ullrich, and E. R. Weippl. "Grid shock: Coordinated load-changing attacks on power grids: The non-smart power grid is vulnerable to cyber attacks as well." In Proc. *The 33rd Annual Computer Security Applications Conference*, December 4–8, 2017, San Juan, PR, USA, pp. 303–314.

[25] C. Moya, J. Hong, and J. Wang. "Application of correlation indices on intrusion detection systems: protecting the power grid against coordinated attacks." *arXiv preprint*, 2018. Available: https://doi.org/10.48550/arXiv.1806.03544.

[26] G. Norman, D. Parker, M. Kwiatkowska, S. Shukla, and R. Gupta. "Using probabilistic model checking for dynamic power management." *Formal aspects of computing*, vol. 17, no. 2, 2005, pp. 160–176.

[27] H. Baumann and W Sandmann. "Markovian modeling and security measure analysis for networks under flooding DoS attacks." In Proc. *The 20th Euromicro International Conference on Parallel, Distributed and Network-based Processing*, 15-17 February, 2012, Munich, Germany, pp. 298–302.

[28] L. Arnaboldi, R. M. Czekster, C. Morisset, and R. Metere. "Modelling load-changing attacks in cyber-physical systems." *Electronic*

Notes in Theoretical Computer Science, vol. 353, 2020, pp. 39–60. Available: https://doi.org/10.1016/j.entcs.2020.09.018.

[29] Y. Liu, P. Ning, and M. K. Reiter. "False data injection attacks against state estimation in electric power grids." In Proc. *16th ACM Conference on Computer and Communication Security*, November 9-13, 2009, Chicago, IL, USA, pp. 21–32. Available: https://dl.acm.org/doi/abs/10.1145/1653662.1653666.

[30] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye. "Detecting false data injection attacks on DC state estimation." In Proc. *1st Workshop on Secure Control Syst.*, April 12, 2010, Stockholm, Sweden, pp. 1–9.

[31] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer. "Coordinated data-injection attack and detection in the smart grid: a detailed look at enriching detection solutions." *IEEE Signal Processing Magazine*, vol. 29, no. 5, 2012, pp. 106–115.

[32] M. A. Rahman, and H. Mohsenian-Rad. "False data injection attacks with incomplete information against smart power grids." In Proc. *IEEE Global Communication Conference*, December 3-7, 2012, Anaheim, CA, pp. 3153–3158.

[33] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad. "Dynamic load altering attacks against power system stability: attack models and protection schemes." *IEEE Transactions on Smart Grid*, vol. 9, issue 4, 2018, pp. 2862–2872.

[34] J. Liu, and F. Labeau. "From wired to wireless: challenges of false data injection attacks against smart grid sensor networks." In Proc. *IEEE Canadian Conference on Electrical & Computer Engineering*, May 13-16, 2018, Quebec City, Canada, pp. 1–6.

[35] S. Aoufi, A. Derhab, and M. Guerroumi. "Survey of false data injection in smart power grid: Attacks, countermeasures and challenges." *Journal of Information Security and Applications*, vol. 54, no. 5, 2020, Article 102518.

[36] Y. Xu. "A review of cyber security risks of power systems: from static to dynamic false data attacks." *Protection and Control of*

Modern Power Systems, vol. 5, issue19, 2020, 12 p. Available: https://doi.org/10.1186/s41601-020-00164-w.

[37] M. Hossain, and C. Peng. "Cyber–physical security for on-going smart grid initiatives: a survey." *IET Cyber-Physical Systems: Theory & Application*, vol. 5, issue 3, 2020, pp. 233–244.

[38] S. Tufail, I. Parvez, S. Batool, and A. Sarwat. "A survey on cybersecurity challenges, detection, and mitigation techniques for the smart grid." *Energies*, vol. 5894, no. 14, 2021, 22 pp. Available: https://www.mdpi.com/1996-1073/14/18/5894.

[39] D. Faquir, N. Chouliaras, S. Vlachou, O. Kalopoulou, and L. Maglaras. "Cybersecurity in smart grids, challenges and solutions." *AIMS Electronics and Electrical Engineering*, vol. 5, issue 1, 2021, pp. 24–37.

Volodymyr G. Skobelev, Volodymyr V. Skobelev

Volodymyr G. Skobelev
V.M. Glushkov Institute of Cybernetics of NAS of Ukraine
40 Glushkova ave., Kyiv, Ukraine, 03187
Taras Shevchenko National Univercity of Kyiv
4d Glushkova ave., Kyiv, Ukraine, 03680
E–mail: skobelevvg@gmail.com

Volodymyr G. Skobelev
V.M. Glushkov Institute of Cybernetics of NAS of Ukraine
40 Glushkova ave., Kyiv, Ukraine, 03187
E–mail: force1304@mail.ru

AUTHORS INDEX

Table of contents