Proceedings MFOI-2018

Conference on Mathematical Foundations of Informatics

Institute of Mathematics and Computer Science
July 2-6, 2018, Chisinau, Moldova

CZU 51+004(082)

C 65

INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE
5, Academiei street, 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: Prof. S.Cojocaru, Prof. C.Gaindric, Ph.D. I.Drugus.

Authors are fully responsible for the content of their papers.

Descrierea CIP a Camerei Naţionale a Cărţii

Conference on Mathematical Foundations of Informatics, Jule 2-6, 2018, Chişinău, Moldova : Proceedings MFOI-2018 / Inst. of Mathematics and Computer Science ; ed.: S. Cojocaru [et al.]. – [Chişinău] : Institute of Mathematics and Computer Science, 2018 (Tipogr. "Valinex" SRL). – 276 p. : fig., tab.

Referinţe bibliogr. la sfârşitul art. – 60 ex.

ISBN 978-9975-4237-7-9.

51+004(082)

ISBN 978-9975-4237-7-9

An Anthropomorphic Avatar-based Approach for Virtual Tutoring Software

Olesea Caftanatov

Abstract

This paper presents the currently most used types and styles of avatars, reviews the attention effect of the anthropomorphic avatars on users and, finally we purpose an anthropomorphic avatar-based design for our GeoMe's virtual tutor.

Keywords: avatar, virtual tutor, visual identification, anthropomorphic avatar.

1. Introduction

In an increasingly digital world, where the popularity of web-based social networking is permanently growing, people communicate through technical devices, practically, same often as without using them. In order to visually represent themselves when communicating with each other online, the users often exploit a software component called "avatar". Generally, an avatar is an embodiment of a person or an idea. The term *"avatar"* originates in Hinduism, from the Sanskrit noun ***"avatāra"***, the meaning of which is *"down coming"*. It refers to a *"descent"* of the divine into realm of material existence, usually for the purpose of protecting or restoring dharma (cosmic order, righteousness) [1]. Avatar literally means *"descent, alight, to make one's appearance"* [2] and refers to the embodiment of the essence of a superhuman being or a deity in another form. This word also implies *"to overcome, to remove, to bring down, and to cross something"* [3].

In computing, an avatar is the graphical representation of the user or the user's alter ego or a character [4]. Avatar's form can vary from 2D model to 3D model or even to 4D model. In general, two-dimensional avatar represents a static or animated user picture used in communication between people, like in socialized web sites, blogs, forums or even live chats. Moreover 3D avatar model is used for computer games, since 1979. The early use was in role-playing computer game *"Avatar"* created in the University of Illinois Control Data Corporation PLATO System [5], which was influenced by many other games, but especially by **dnd** (A.k.a *"The Game of Dungeons"*) [6]. The first time this word was used as a term for the user's identity on screen in *"Ultima IV: Quest of the Avatar"* of 1985 [7]. Recently, the greater challenge became the creation of the 4D avatars (like hologram avatar). An example of this can be found in [8] – this is a footage from an exhibition game match in *"Yu-Gi-Oh!"* from World Championship 2017.

Nowadays, in addition to games, the use of avatars influences all fields of our life including business, marketing, science, healthcare, digital art, telecommunication and educational systems. In this work, we will focus our attention on the design of an avatar for educational software **GeoMe.** In our case, the avatar is the tutor's embodiment, which helps the users to learn geometry courses.

2. Types of Avatar Design

According to [9-12] reviews various types of avatar are used, and based on one of which experts can detect and make the psychological portrait of user. Avatars offer a broad range of behavioral and emotional expressions. Below there are described only five types of avatar design.

Anonymous avatar – basically, this type of avatar is used by new users (beginners) or by the category of people, who visit some web sites for the first and last time, or by people that do not care how their avatar looks like. In early days, different applications used a grey picture of a man or a woman to represent an anonymous avatar.

While in present, exist various ways to make more creative anonymous avatar. For instance, GitHub generates ***identicons*** for anyone without avatar on their platform [13]. Their identicons are unique geometric 5x5 pixel patterns generated based on a hash of a user's IP address (see Figure 1).

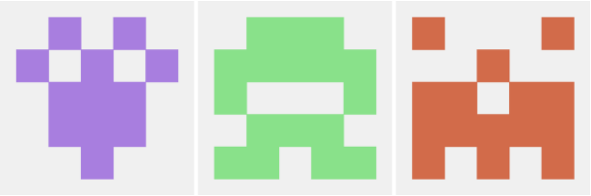


Figure 1. Identicons generated by GitHub

Another creative style of anonymous avatar is ***initials*** with a unique background color. Depending on what kind of information is available from the user profile, system will display the first or last initial, like Google does (see Figure 2), or combination of initials as Dropbox does.

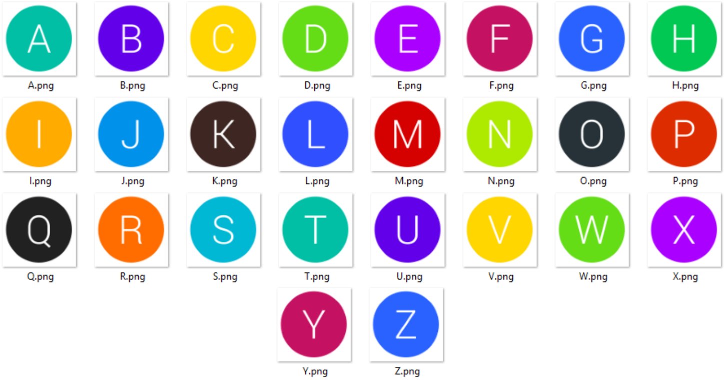


Figure 2. Initials generated by Google

Additionally, Google found a funny way to represent the anonymous users who visualize the shared document. Instead of only using initials icon with different background colors for each person, they associate each person with an animal [14] (see Figure 3). In addition, to some expected animals like dog, kitty or bear, they also display extinct or mythical animals [15], such as auroch, chupacabra, ifrit etc.

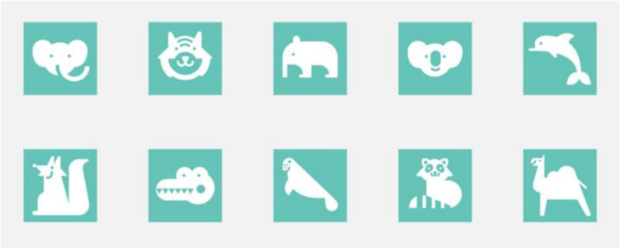


Figure 3. Anonymous animal type avatar generated by Google Drive

Animal avatar – regarding this type of avatar, people themselves use it as much as various systems that generate anonymous icons. According to [16] people like to use animal avatars because it somehow relates to their pets or because animals symbolize certain traits or attributes (e.g. strength, loyalty, grace, independence, cunning, etc.) or characteristics admired by the person. Also, in some case the chosen animal probably bears some psychological significance to a person, perhaps it represents some aspect of his identity. For example: a cat means weasel, domesticity, independence; a dog – loyalty, honesty; a snake – wisdom; a fox – cunningness; a bull – strength, perseverance [17].

Figure 4. Animal avatars

The opposed type of avatar to anonymous one is ***real face avatar***. Some users prefer to use real face avatar, it can be their self-photo or their loved one. Usually, when a user presents a photo with them, it may be a gesture of honesty, romance or a sign of friendship.

The peculiarities of people with this type of avatar are their complete confidence in their rightness of their way of life and thoughts (Figure 5).

Meanwhile, real face avatar has its patterns too, for instance: if the presented image displays a couple – it means demonstration of their own relations, their own usefulness; a photo from childhood – means melancholy, users miss the time from past; a selfie – usually, behind of a self-avatar can hide a real *narcissus*. Owners of self-avatars often change profile image, it is a way to demonstrate what they like, what they can be in various angles in different situations with a large range of emotions that they express.

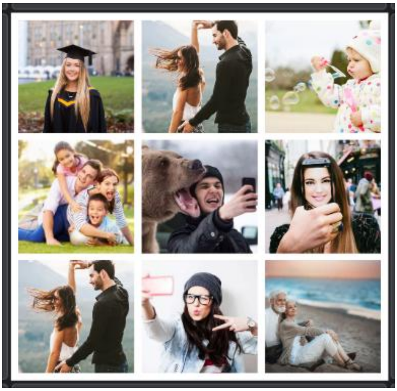


Figure 5. Real face avatars

Nature avatar – it can be natural landscapes (e.g. mountains, sea, waterfall, space, etc.), natural phenomena (e.g. lightning, explosion, fire, snowing, sandstorm, rain, etc.), floral, etc. Architectural design considers belonging to the same type of avatar (see Figure 6). For this type, the background is more important than the details. People that hide beside such avatar are inclined to non-standard thinking. Usually, they do not like to talk about themselves, but they are always ready to speculate about the world order (e.g. philosophy, politics, science, economics and other favorite topics).

Figure 6. Nature avatars

Character avatar – this type includes a wide range of fantasy styles, it can be an idol from movie that they greatly admire and love. Many users have desires (conscious or unconscious) to have superpowers like movie heroes (Spider-Man, Superman, Batman, Flash, Cat-Woman, etc.). Often behind such avatar a teenager hides. Anyway, who does not want to be strong and invulnerable? This type of idolization does not limit only to movies, but to cartoons, books, comics, games, manga, anime too (see Figure 7).

Figure 7. Character avatars

3. Avatar vs Gravatar

As we know, an avatar is an image that represents us online - a little icon that usually displays next to our nickname. It can be static image *(.jpeg, .bmp, .png, etc.)* or animated *(.gif)*. The acronym ***"GIF"*** stands for the ***G****raphics* ***I****nterchange* ***F****ormat* and is one of the most common avatar formats on World Wide Web. In order to not overuse the memory capacity of application, all systems have a size file limitation indifferent of what avatar format we use. Therefore, depending of websites or other application we need to resize our images according to avatars size limitation. It can be 24 x24 pixels or even 200x200 pixels.

Sometimes when we resize our favorite image to the needed requirements, our image deforms and this can really frustrate. We are compelled to go through this process many times, because we daily use not only one website or live chat. For those users who like to change their avatar very often, obvious it costs time, so one solution can be using Gravatar.

A ***Gravatar*** is a ***G****lobally* ***R****ecognized* ***Avatar*** [18]. Gravatar is a service for providing globally – unique avatars and was created by Tom Preston – Werner [19]. The main point is that user needs to create and upload profile just once, after that wherever he (she) participates in any Gravatar-enable sites, their gravatar image will follow them automatically. Therefore, this allows each commenter to have their identity throughout the World Wide Web.

Those users who want to keep their profile low should use avatar, but people who have their own business, developers, bloggers or anyone who wants to build a brand should start using a gravatar [20], because the gravatars help them to be recognized on large scale.

4. Avatar Design Principles

According to [21] Park Ji Yong and Nam Yong Hyun review, various types, style of avatar build up a well-articulated set of design principles for effective avatar design. Based on avatar design consideration (*mediator, personality, customization and interactivity*) the foundations for five avatar design principles were set.

The first principle is about the possibility to be ***identifiable***. All avatars should be identifiable like brand mark for companies, in such way it would be recognized as “online identifier” for natural person.

Another principle is about ***esthetical aspect of avatars***, because their main purpose is to attract other users’ attention and by communicating with each other to create a friendship bond, or romantic bond, or even a collaboration bond.

Additionally, avatars should be ***creative and unique*** because all people intend to be unique. There are not identical people in real world, even if there are twins that look the same, their personality for sure is different. That is why this is another design principle that we should consider.

In addition, one of the important parts of avatars is their possibility to be ***manipulable***. In terms of customization, avatar should be able to be applied with other objects and be combined with other elements such as background, text, stickers, animation, etc.

The last but not the least is the property to be ***expendable*** because as user grows up and changes, or as his mood changes, the avatar should be able to change too, in order to fit people’s trends.

5. Virtual Tutor’s Avatar Design

According to [22], personifying the virtual tutor has a positive impact on learners, especially if the virtual tutor’s embodiment includes emotional aspects, because it catches and focuses student’s attention. Therefore, we intend to develop virtual tutor for our educational application ***GeoMe***. In addition, C. Okonkwo considers that there are some gender-based and individual differences in the user perception of an emotional agent, which need to be taken into account when designing a virtual tutor. In our project we choose a parrot species to be an embodiment of our virtual tutor~~,~~ (see Figure 8).

The main argument is that parrots try to repeat the interesting sounds from their environment as students try to repeat assimilated knowledge after their teachers. In general, humanity may repeat almost everything what they see or learn from others. Similarly, parrots are like small children who are very curious to learn as many interesting things as possible from the environment in which they live. Also, parrots observe and then try to mimic what they hear, often the short sounds.

Moreover, it is a perfect solution for our mobile application, because in general we intend to create short messages almost for everything, being it an encouraging message or a definition of a geometrical shape. All ideas for avatar's design belong to me, but all arts are drawn by my friend ***Victoria Țvetkova***.

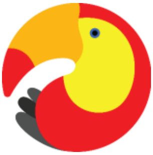


Figure 8. Our virtual tutor's avatar

In our project we intend to design 4 levels (*0, 1, 2, 3*) of avatar for each level difficulty. ***0 level*** is designed for all new users, because virtual teacher does not have enough information about their skills and their knowledge level. For this level, we purpose an interesting egg design appearance. In order to pop up a parrot from egg, the users need only to read the basic theory level recommended by eggy tutor. In Figure 9 there are presented two avatar's 2D images, one before basic theory activity and the second image is after this activity was taken. It means that virtual teacher has some information about their students, thus it means that the tutor now may purpose some assessment activity.

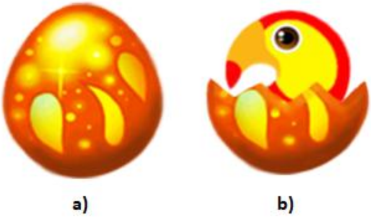


Figure 9. 2D image of virtual tutor with level 0

Regarding ***level 1-3*** it stands for three categories of difficulty: ***easy, moderate and challenging***. In all easy activities our avatar design will have a *chibi appearance*, as in figure 10 A. For level 2 (moderate) our virtual tutor's appearance evolve to a *teenager appearance* (see Figure 10

B), and for the challenging level our avatar's design looks like an *grown up parrot* (see Figure 10 C).

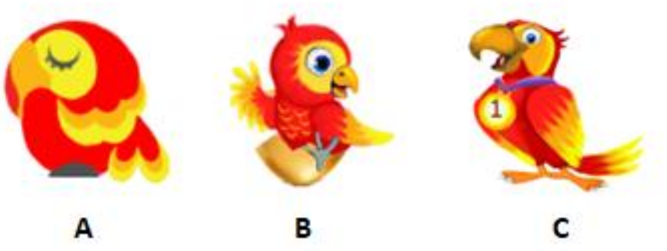


Figure 10. The evolution of our virtual tutor

The main point is that users are motivated to help their virtual teacher to access the greater level of knowledge, subsequently they learn according to that level. In other words, it can be described as being a:"knowledge challenger".

6. Attentional effect of anthropomorphic avatar

Anthropomorphism is the idea that an animal or an object has feelings, characteristics or behaviors like those of a human being. Exists some debate about the influence of anthropomorphic avatars, in some research like in [24], people reported the less anthropomorphic image to be more credible and likeable than the more anthropomorphic image. Alternatively, like in [25] study, participants reported feeling that they were more engaged in their communication when they did not use any anthropomorphic images representing their partners on the small screen of a cell phone. On the other hand, in some studies [26-29] people believe that avatars that are more anthropomorphic are more credible, engaging and likeable than less anthropomorphic images.

Additionally, there was a concern that the animated character might prove to be a distraction for the participants. In [23, 30, 31] studies found that the anthropomorphic animated character was marginally less distracting than the simple animated character.

Despite of all discusses, we intend to design an anthropomorphic avatar because it:

- motivates user on a long interaction with application;

- attract the user's focus of attention;
- conveys emotional signal;
- guides the user through application.

7. Designing Avatar's Emotions

Regarding emotions, we intend to design them to be more like human feelings (see Figure 11). In order to focus the user's attention we intend to add voice message to those avatars. Some of the voice messages will be to encourage users when they answer correctly /incorrectly to their assessment's activity something like ***"Good job!"*** for positive answers, and ***"Aww, Try again!"*** for negative answers.

Considering that the target users are the students from elementary school, thus we all know that they do not really like to read, so we will add voice sounds to the theoretical part too. For example, in [32] you can hear the angle definition in Romanian version. Initially, we wanted to add voice message that mimic parrot's speech but it is hard to understand. Hence, in order to maintain the easy comprehension we mimic child's voice. It also will catch the user's attention.

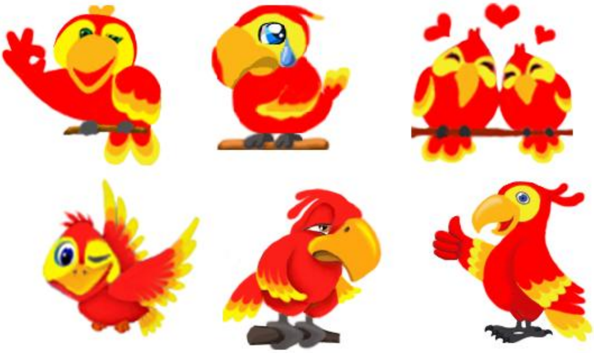


Figure 11. Some example of avatar's emotion

8. Final thoughts

Our main project goal consists in developing an education application with a personalized content and design. In order to obtain the personalization we want to add anthropomorphic avatar-based approach for our virtual tutor. Well, just deciding on an avatar style takes thoughtfulness, designing, research, analyzing, etc. and it is only a small step towards our largest goal of personalization. Nevertheless, even these small changes in avatar's appearance according to user's behavior make the work meaningful, funny and creative. It has the potential to attract and keep user's attention by creating delightful experience that affects users in the virtual and real world.

References

[1] *Avatar. New World Encyclopedia*. http://www.newworldencyclopedia.org/entry/Avatar

[2] S. Noel. *"Hindu Avatāra and Christian Incarnation: A Comparison"*. Philosophy East and West. Volume 52, January 1, 2002. p.98-125. http://muse.jhu.edu/article/26576/pdf

[3] W. Monier. *Avatāra* A Sanskrit English Dictionary. Oxford University Press. P.90. https://books.google.bg/books?id=_3NWAAAAcAAJ&pg=PA90&redir_esc=y#v=onepage&q&f=false

[4] *Avatar (computing)*. Wikipedia. https://en.wikipedia.org/wiki/Avatar_(computing)#Origins

[5] Video Games *A.k.a. Avatar*. University of Illinois. 1979. link: https://www.uvlist.net/game-174616-avatar

[6] Video Game *Dnd A.k.a. The Game of Dungeons*. Southern Illinois University. 1974 link:. https://www.uvlist.net/game-160118-dnd

[7] J. Mahner. *Ultima IV: Quest of the Avatar*. 1985. The digital antiquarian website link: https://www.filfre.net/2014/07/ultima-iv/

[8] Aditya. *Augmented Reality Duel - Yu-Gi-Oh! World Championship 2017*. Video link: https://www.youtube.com/watch?v=mgs6fgqYquw

[9] E. Koba. *Top -15 primetov: O ciom psihologam raskajet vash avatar*. June 8. 2015. Website link: https://hornews.com/top/top-15_primerov_o_chem_psihologam_rasskajet_vash_avatar/

[10] *Kak po avatarke sostaviti psihologiceskii portret celoveka*. Part 2. Nov. 16, 2011. Blog post link: https://www.baby.ru/blogs/post/42647419-7091480/

[11] *Psihologia vibora avatara. Test "O ciom raskajet ava – uznai o celoveke po avatare"*. Website link: https://psycabi.net/testy/520-psikhologiya-vybora-avatara-test-o-chem-rasskazhet-ava-uznaj-cheloveka-po-avataru

[12] *Avatari i harakter. Kak po avatare uznati haracter celoveka?*.Website link: http://www.otvet.quarkon.ru/avatar.php

[13] J. Long. *Identicons.* August 14, 2013. The GitHub Blog post link: https://blog.github.com/2013-08-14-identicons/

[14] A. Chitu. *Anonymous Animals in Google Drive.* April 19, 2013. Google System Blog post link: https://googlesystem.blogspot.com/2013/04/anonymous-animals-in-google-drive.html

[15] A. Robbins. *Anonymous Animals Icons in Google Drive.* February 2, 2016. Blog post link: http://www.hercampus.com/school/ucd/weirdest-google-doc-anonymous-animals-explained

[16] J. Suler. *The Psychology of Avatars and Graphical Space in Multimedia Chat Communities. A Study of the Palace*. Chat Communication, Michael Beiswenger, pp. 305-344. Ibidem. Stuttgart, Germany link: http://www-usr.rider.edu/~suler/psycyber/psyav.html

[17] GeorgiNNa. *Avatarki i ih hozeaeva.* Softmixer. Online Journal. Post link: http://www.softmixer.com/2011/06/blog-post_6076.html

[18] *What is Gravatar?* Gravatar website documentation link: http://en.gravatar.com/support/what-is-gravatar/

[19] *Gravatar*, Wikipedia website link: https://en.wikipedia.org/wiki/Gravatar

[20] Editoral Staff. *What is Gravatar and Why you should start using it right away. Beginner's guide for WordPress.* March 12, 2013. Website link: http://www.wpbeginner.com/beginners-guide/what-is-gravatar-and-why-you-should-start-using-it-right-away/

[21] P.J. Yong, N.Y. Hyun. *The understanding of avatar design: various types and styles of avatars and the design considerations.* The Korea Society of Illustration Research Vol.18. 2009. https://eprints.qut.edu.au/29594/1/29594_-_The_Understanding_of_Avatar_Design.pdf

[22] C. Okonkwo. *Affective Pedagogical Agents and User Persuasion.* University of Saskatchewan. Canada. https://pdfs.semanticscholar.org/ 0e2d/db1b6362f83b7a6a53b6bb34143f1353d0a3.pdf

[23] E. Andre, T. Rist, J. Müller. *WebPersona: A life-like presentation agent for the World-Wide Web.* German Research Center for Artificial Intelligence. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=E4F6E9B88AE37FD069A3828F7180A817?doi=10.1.1.124.1449&rep=rep1&type=pdf

[24] K.L. Nowak, F. Biocca. *The effect of the agency and anthropomorphism on users' sense of telepresence, compresence, and social presence in virtual environments.* MIT Press. Teleoperators and virtual environments. Project documentation link: http://mindlab.org/images/d/DOC836.pdf

[25] S.H. Kang, J.H. Watt, K. Isbister. *The effect of static anthropomorphic images on emotion perceptions in mobile phone communication.* Rensselaer Polytechnic Institute. Presence 2006. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.555.8491&rep=rep1&type=pdf

[26] K.L. Nowak, G. Zoric, N.Sträfling, N. Krämer. *The Psychology of Avatars: Real life effects of virtual communication.* Proceedings of the 11th Annual International Workshop on Presence Padova, 16-18 Octomber 2008. https://www.researchgate.net/profile/Sabine_Trepte/publication/228831565_The_Psychology_of_Avatars_Real_Life_Effects_of_Virtual_Communication/links/54ae982b0cf2b48e8ed44fa7/The-Psychology-of-Avatars-Real-Life-Effects-of-Virtual-Communication.pdf

[27] K.L.Nowak,C. Rauh. *The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily and attraction.* Journal of Computer Mediated Communication, Vol 11, Issue 1, November 1, 2005. Pp.153-178. https://academic.oup.com/jcmc/article/11/1/153/4616661

[28] A. Wexelblat. *Don't make that face: a report on anthropomorphizing an interface.* AAAI Technical Report SS-98-02. 1998. https://pdfs.semanticscholar.org/5014/5ab20d2c274def86eb60b43ed87c68657788.pdf

[29] T. Koda. *User Reactions to anthropomorphized interfaces.* JST CREST Digital City Project. Kyoto University, Japan. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.5606&rep=rep1&type=pdf

[30] C. Hongpaisanwiwat, M. Lewis. A*ttentional effect of animated character.* IOS Press, IFIP, 2003, pp. 423-430. http://www.idemployee.id.tue.nl/g.w.m.rauterberg/conferences/interact2003/INTERACT2003-p423.pdf

[31] J.H. Walker, L. Sproull, R. Subramani. *Using a human face in an interface.* Human Factors in Computing Systems, Boston, Massachusetts, USA. April 24-28, 1994. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.8319&rep=rep1&type=pdf

[32] O. Caftanatov. *Angle definition in Romanian version.* GeoMe Application. May 25, 2018. Voice Sound link: https://files.fm/f/9sk2pg5u

Olesea Caftanatov[1]

[1] Institute of Mathematics and Computer Science, Chisinau, Moldova
E-mail: olesea.caftanatov@math.md

On some groupoids of order three with Bol-Moufang type of identities

Vladimir Chernov, Alexander Moldovyan, Victor Shcherbacov

Abstract

We count number of groupoids of order 3 with some Bol-Moufang type identities.

Keywords: groupoid, Bol-Moufang type identity.

1 Introduction

A binary groupoid $(G, \cdot)$ is a non-empty set G together with a binary operation "$\cdot$". This definition is very general, therefore usually groupoids with some identities are studied. For example, groupoids with associativity identity (semi-groups) are researched.

Here we continue the study of groupoids with some Bol-Moufang type identities [6, 2, 9, 3].

Definition. Identities that involve three variables, two of which appear once on both sides of the equation and one of which appears twice on both sides are called Bol-Moufang type identities.

Various properties of Bol-Moufang type identities in quasigroups and loops are studied in [5, 7, 4, 1].

Groupoid $(Q, *)$ is called a quasigroup, if the following conditions are true [2]: $(\forall u, v \in Q)(\exists!\, x, y \in Q)(u * x = v \,\&\, y * u = v)$.

For groupoids the following natural problems are researched: how many groupoids with some identities of small order there exist? A list of numbers of semigroups of orders up to 8 is given in [8]; a list of numbers of quasigroups up to 11 is given in [6, 10].

2 Results

Original algorithm is elaborated and the corresponding program is written for generating groupoids of small (2 and 3) orders with some Bol-Moufang identities, which are well known in quasigroup theory.

To verify correctness of the written program, the number of semigroups of order 3 was counted. The obtained result coincides with the well known one, namely, there exist 113 semigroups of order 3.

We use list of Bol-Moufang type identities given in [4]. In Tables 1 and 2 we present number of groupoids of order 3 with the respective identity.

References

[1] Reza Akhtar, Ashley Arp, Michael Kaminski, Jasmine Van Exel, Davian Vernon, and Cory Washington. *The varieties of Bol-Moufang quasigroups defined by a single operation.* Quasigroups Related Systems, vol. 20(1), pp. 1–10, 2012.

[2] V.D. Belousov. *Foundations of the Theory of Quasigroups and Loops.* Nauka, Moscow, 1967. (in Russian).

[3] Vladimir Chernov, Nicolai Moldovyan, and Victor Shcherbacov. *On some groupoids of small order.* In The Fourth Conference of Mathematical Society of the Republic of Moldova dedicated to the centenary of Vladimir Andrunachievici (1917-1997), June 28 - July 2, 2017, Chisinau, Proceedings CMSM4', pp. 51–54, Chisinau, Moldova, 2017.

[4] B. Cote, B. Harvill, M. Huhn, and A. Kirchman. *Classification of loops of generalized Bol-Moufang type.* Quasigroups Related Systems, vol. 19(2), pp. 193–206, 2011.

[5] F. Fenyves. *Extra loops. II. On loops with identities of Bol-Moufang type.* Publ. Math. Debrecen, vol. 16, pp. 187–192, 1969.

[6] H.O. Pflugfelder. *Quasigroups and Loops: Introduction.* Heldermann Verlag, Berlin, 1990.

Table 1. Number of groupoids of order 3 with some identities.

Name	*Abbreviation*	*Identity*	*Number*
Semigroups	*SGR*	$x(yz) = (xy)z$	113
Extra	*EL*	$x(y(zx)) = ((xy)z)x$	239
Moufang	*ML*	$(xy)(zx) = (x(yz))x$	196
LeftBol	*LB*	$x(y(xz)) = (x(yx))z$	215
RightBol	*RB*	$y((xz)x) = ((yx)z)x$	215
$C - loops$	*CL*	$y(x(xz)) = ((yx)x)z$	133
$LC - loops$	*LC*	$(xx)(yz) = (x(xy))z$	220
$RC - loops$	*RC*	$y((zx)x) = (yz)(xx)$	220
MiddleNuclearSquare	*MN*	$y((xx)z) = (y(xx))z$	350
RightNuclearSquare	*RN*	$y(z(xx)) = (yz)(xx)$	932
LeftNuclearSquare	*LN*	$((xx)y)z = (xx)(yz)$	932
Comm.Moufang	*CM*	$(xy)(xz) = (xx)(zy)$	297
AbelianGroup	*AG*	$x(yz) = (yx)z$	91
$Comm.C - loop$	*CC*	$(y(xy))z = x(y(yz))$	169
Comm.Alternative	*CA*	$((xx)y)z = z(x(yx))$	110
Comm.Nuclearsquare	*CN*	$((xx)y)z = (xx)(zy)$	472
Comm.loops	*CP*	$((yx)x)z = z(x(yx))$	744
Cheban 1	*C1*	$x((xy)z) = (yx)(xz)$	219
Cheban 2	*C2*	$x((xy)z) = (y(zx))x$	153
Lonely I	*L1*	$(x(xy))z = y((zx)x)$	117
Cheban I Dual	*CD*	$(yx)(xz) = (y(zx))x$	219
Lonely II	*L2*	$(x(xy))z = y((xx)z)$	157
Lonely III	*L3*	$(y(xx))z = y((zx)x)$	157
Mate I	*M1*	$(x(xy))z = ((yz)x)x$	111
Mate II	*M2*	$(y(xx))z = ((yz)x)x$	196
Mate III	*M3*	$x(x(yz)) = y((zx)x)$	111
Mate IV	*M4*	$x(x(yz)) = y((xx)z)$	196
Triad I	*T1*	$(xx)(yz) = y(z(xx))$	162
Triad II	*T2*	$((xx)y)z = y(z(xx))$	180
Triad III	*T3*	$((xx)y)z = (yz)(xx)$	162
Triad IV	*T4*	$((xx)y)z = ((yz)x)x$	132
Triad V	*T5*	$x(x(yz)) = y(z(xx))$	132
Triad VI	*T6*	$(xx)(yz) = (yz)(xx)$	1419

Table 2. Number of groupoids of order 3 with some identities (continuation of Table 1).

$Name$	$Abbreviation$	$Identity$	$Number$
$TriadVII$	$T7$	$((xx)y)z = ((yx)x)z$	428
$TriadVIII$	$T8$	$(xx)(yz) = y((zx)x)$	120
$TriadIX$	$T9$	$(x(xy))z = y(z(xx))$	102
$Frute$	FR	$(x(xy))z = (y(zx))x$	129
$CrazyLoop$	CR	$(x(xy))z = (yx)(xz)$	136
$Krypton$	KL	$((xx)y)z = (x(yz))x$	268

[7] J. D. Phillips and Petr Vojtechovsky. *The varieties of loops of Bol-Moufang type.* Algebra Universalis, vol. 54(3), pp. 259–271, 2005.

[8] S. Satoh, K. Yama, and M. Tokizawa. *Semigroups of order 8.* Semigroup forum, vol. 49, pp. 7–29, 1994.

[9] Victor Shcherbacov. *Elements of Quasigroup Theory and Applications.* CRC Press, Boca Raton, 2017.

[10] Wikipedia. Semigroup with two elements, 2015. https://en.wikipedia.org/wiki/Semigroup_with_two_elements.

Vladimir Chernov[1], Alexander Moldovyan[2], Victor Shcherbacov[3]

[1]Shevchenko Transnistria State University
Email: volodya.black@gmail.com

[2]St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences
Email: nmold@mail.ru

[3]Institute of Mathematics and Computer Science, Moldova
Email: victor.scerbacov@math.md

Scattered and Digital Topologies in Image Processing

Mitrofan M.Cioban, Ivan A. Budanaev

Abstract

In this paper we present some geometrical and topological concepts to accommodate the needs of information theories. It is demonstrated that image processing methods and algorithms are well correlated with various aspects of the notion of $\mathcal{P}$-scattered space, where $\mathcal{P}$ is a given property. Theorem 2.8 illustrates how the $\mathcal{P}$-scatteredness concept can be explored through mathematical induction, which opens the possibilities to apply algorithmic procedures. It is established that the Khalimsky topology on the discrete line is unique as the minimal symmetric digital topology on $\mathbb{Z}$.

Keywords: digital space, discrete line, scattered space, digital image processing.

1 General notions and problems

Any topological space X is considered to be a Kolmogorov space, i.e. a T_0-space: for any two distinct points $x, y \in X$ there exists an open subset U of X such that $U \cap \{x, y\}$ is a singleton set [18]. Denote by $cl_X F$ the closure of the set F in the space X and by $|L|$ the cardinality of the set L. Let $\omega = \{0, 1, 2, ...\}$, $\mathbb{N} = \{1, 2, ...\}$ and $\omega(n) = \{0, 1, 2, ..., n\}$ for each $n \in \omega$.

It is well known that distinct algebraical and topological structures have been introduced to accommodate the needs of information theories. In the process of studying of the continuous objects by the computers methods they are approximated by finite objects or by digital images [1, 8, 9, 11, 14, 25, 30, 31, 32].

Digital image processing is a process which from a topological point of view may be described in the following way:

1. Fix an infinite space X (a continuous image of the original) and a property $\mathcal{P}$ of subspaces of the space X.

2. By some procedure we construct a number $n \in \omega$, a finite subset $H = \{h_i : i \in \omega(n)\} \subset \mathbb{Z}$ of levels and a finite family $\{G_i : i \in \omega(n)\}$ of open non-empty subsets of the space X with the properties:

- $G_i \cap G_k = \emptyset$ for all $0 \leq i < k \leq n$;

- for any $i \in \omega(n)$ and each $x \in G_i$ there exists an open subset $G(x)$ such that $x \in G(x) \subset G_i$ and $G(x)$ is a subset with the property $\mathcal{P}$ in X;

- the set $G = \{G_i : i \in \omega(n)\}$ is dense in X.

The set G is the $\mathcal{P}$-kernel and $X \setminus G$ is the $\mathcal{P}$-residue of the space X.

3. The intensity mapping $I_{\mathcal{P}} : X \to \omega = H$ is determined with the property: $I_{\mathcal{P}}(x) = maximal\{h_i : x \in cl_X G_i\}$ for each $x \in X$. We have $G_i \subset I_{\mathcal{P}}^{-1}(h_i)$ for each $i \in \omega(n)$.

4. A digital topology for which the mapping $I_{\mathcal{P}}$ is continuous, is determined on H.

5. By some procedure we construct a finite T_0-space K and for any $x \in X$ we determine a non-empty subset $D_{\mathcal{P}}(x)$ of K such that:

- for any $c \in K$ the set $X(c) = \{x \in X : c \in D_{\mathcal{P}}(x)\}$ is closed and is called a $\mathcal{P}$-cell of X;

- for any $c \in K$ there exist $i \in \omega(n)$ and an open non-empty subset $X'(c) \subset G_i$ such that $X(c) = cl_X X'(c)$.

The family $\{X(c) : c \in K\}$ is called a $\mathcal{P}$-complex and the mapping $D_{\mathcal{P}}$ represent an approximation of X by a finite space K. Methods of constructions of the objects K, $D_{\mathcal{P}}$ and $X(c)$ are called the methods of digitalization. This procedure is known in image processing literature as “thinning”, “skeletonization”, “digitalization” and “segmentation” process. In the concrete situations, the $\mathcal{P}$-cells are called pixels, voxels etc. The mapping $D_{\mathcal{P}}$ can be considered as a model of a digitizer.

Typical problems arising in this context are:

- Which topological (geometrical) properties does the finite spaces

H and K share with the space X?

- Is $D_{\mathcal{P}}$ or its inverse mapping continuous in some sense as a set-valued mapping?

- Classification of points and curves in the digital spaces. Study of digital invariants.

- Determine more "simple" topologically (homotopically) equivalent spaces of the space X, if the space K is complicate.

The inverse problem of discretization and digitalization is in some sense the problem of finding a continuous model for a given finite space K.

Methods of discreetness of spaces bring us to the notions of a $\mathcal{P}$-scattered space and of a $\mathcal{P}$-decomposable space.

2 Algorithms and scattered spaces

In many cases it is necessary to find a procedure or an algorithm that allows us to study from a certain point of view a given space or some object. As a rule, this procedure of study can be extended to a much larger class of spaces.

Let $\mathcal{P}$ be a property of spaces. We say that the subspace Y of X has the property $\mathcal{P}$ in X if there exists a subspace Z of X with property $\mathcal{P}$ such that $Y \subset Z$. A space X is a space with local property $\mathcal{P}$ if for any point $x \in X$ there exists an open subspace U with the property $\mathcal{P}$ in X such that $x \in U$.

In [12, 15, 35, 36] the following classes of spaces were introduced.

Definition 2.1. *A space X is called a $\mathcal{P}$-scattered space if for any non-empty closed subspace Y of X there exists a non-empty open subset U of Y such that the subspace U has the property $\mathcal{P}$ in X.*

Definition 2.2. *A space X is called a $\mathcal{P}$-decomposable space if there exist an ordinal number $\alpha_0 \geq 1$ and a family $\{X_\alpha : \alpha < \alpha_0\}$ of non-empty subspaces of X such that:*

1) $X = \{X_\alpha : \alpha < \alpha_0\}$ and $X_\alpha \cap X_\beta = \emptyset$ for any $0 \leq \alpha < \beta < \alpha_0$;

2) the set $\cup\{X_\alpha : \alpha < \beta\}$ is open in X for each $\beta \leq \alpha_0$;

3) X_α is a space with local property $\mathcal{P}$ in X.

We say that $\{X_\alpha : \alpha < \alpha_0\}$ *is a* $\mathcal{P}$*-decomposition of the space* X*.*

If X is a $\mathcal{P}$-decomposable space, then the index of $\mathcal{P}$-decomposition $id_{\mathcal{P}}(X)$ is the minimal ordinal number α_0 for which there exists a $\mathcal{P}$-decomposition $\{X_\alpha : \alpha < \alpha_0\}$ of the space X.

In [12, 15, 35, 36] were proved the following assertions:

A1. *Any* $\mathcal{P}$*-scattered space is* $\mathcal{P}$*-decomposable. In this case* $id_{\mathcal{P}}(X)$ *is the index of* $\mathcal{P}$*-scatteredness of the space* X*.*

A2. *If any closed subspace of a space with property* $\mathcal{P}$ *is a space with the property* $\mathcal{P}$*, then any* $\mathcal{P}$*-decomposable space is* $\mathcal{P}$*-scattered.*

A3. *Any non-empty closed subspace of a* $\mathcal{P}$*-scattered space is a* $\mathcal{P}$*-scattered space.*

A4. *If any non-empty subspace of a space with property* $\mathcal{P}$ *is a space with the property* $\mathcal{P}$*, then any subspace of a* $\mathcal{P}$*-scattered space is a* $\mathcal{P}$*-scattered space.*

Let X be a $\mathcal{P}$-scattered space. If Y is a closed subspace of X, then $\mathcal{P}\text{-}kernel_X(Y) = \cup\{U \subset Y : U \textit{ is open in } Y \textit{ and has property } \mathcal{P} \text{ in } X\}$. Let $X_0 = \mathcal{P}\text{-}kernel_X(X)$ and $X_\alpha = \mathcal{P}\text{-}kernel_X(X \setminus \cup\{X_\beta : \beta < \alpha\})$ for each ordinal number α. Then $id_{\mathcal{P}}(X) = minimal\{\alpha : X_\alpha = \emptyset\}$.

Example 2.3. Let $\mathcal{C}$ be the property of a space to be a connected space. Then any connected or locally connected space is a $\mathcal{C}$-decomposable space. A closed subspace of a connected space is not obligatory connected. The unit segment $[0, 1]$ in the Euclidean topology is connected and the Cantor subspace of the unit interval does not have any non-empty open connected subsets. Hence, $[0, 1]$ is a $\mathcal{C}$-decomposable space and is not a $\mathcal{C}$-scattered space.

Example 2.4. Let $\mathcal{S}$ be the property of a space to be a singleton space. The $\mathcal{S}$-scattered space is called a *scattered space.* A space X is scattered if for any non-empty subspace Y of X there exists a point $y \in Y$ such that the set $\{y\}$ is open in Y, i.e. y is an isolated point of Y. Denote by $id_s(X) = id_{\mathcal{S}}(X)$ the index of scateredness of the scattered space X. Any $\mathcal{S}$-decomposable space is scattered.

Example 2.5. Let k be the property of a space to be a compact space. A space X is *k-scattered* if for any non-empty closed subspace Y of X there exist a non-empty open subset U of Y and a compact subset F of Y such that $U \subset F$ [3, 34]. Any closed subspace of a k-scattered space is k-scattered. A subspace of a k-scattered space is not obligatory k-scattered. Indeed, the subspace of rationals from the unit interval $[0,1]$ is not k-scattered and the unit interval is k-scattered. Any scattered space is k-scattered. The unit interval is k-scattered and not scattered. Any k-decomposable space is k-scattered.

Example 2.6. Let SP be the property: intersection of a countable family of open subsets is open. A space X is *SP-scattered* [22] if for any non-empty subspace Y of X there exists a non-empty open subset U of Y such that U is a space with the property SP. Any subspace of a SP-scattered space is SP-scattered. Any scattered space is SP-scattered. Any SP-decomposable space is SP-scattered. A space with the property SP is called a P-space. A point $a \in X$ of a space X is a P-point if any countable family of neighbourhoods of the point a contains a neighbourhood of the point a in X. A space X is a P-space if any point $a \in X$ is a P-point. There exists a hereditarily paracompact not scattered P-space.

Example 2.7. Let FP be the property of a space to be a finite space. A space X is *FP-scattered* [22] if for any non-empty subspace Y of X there exists a non-empty open subset U of Y such that U is a finite set. Any subspace of a FP-scattered space is FP-scattered. A space is scattered if and only if it is FP-scattered.

We mention the following universal theorem.

Theorem 2.8. *Let $\mathcal{P}$, Γ and $\mathcal{Q}$ be the properties of spaces with the following conditions:*

- any space with property $\mathcal{P}$ has the properties $\mathcal{Q}$ and Γ;

- a closed subspace of the space with the property Γ is a space with the property Γ;

- if $Y = Z \cup S$ is a space with the property Γ, where S is a closed subspace with property $\mathcal{P}$ and Z is a subspace with local property $\mathcal{Q}$ in Y, then the space Y has the property $\mathcal{Q}$;

- if S and Z are open subspaces of the space Y with the property Γ, F is a subspace of Y with the property $\mathcal{P}$ and $x \in S \setminus Z \subset F$, then there exist an open subset U of Y and a subspace Φ with the property

Γ *such that* $x \in U$, $U \subset \Phi \subset Z \cup (F \setminus Z)$ *and* $\Phi \setminus Z$ *has the property* $\mathcal{P}$.

Then any $\mathcal{P}$*-decomposable space* X *with the property* Γ *has the property* $\mathcal{Q}$.

Proof. Fix a $\mathcal{P}$-decomposition $\{X_\alpha : \alpha < \alpha_0\}$ of the space X. It is sufficient to prove that X is a space with local property $\mathcal{P}$. For any point $x \in X$ we will construct an open subset Ux with the property $\mathcal{Q}$ in X such that $x \in Ux$.

If $x \in X_0$, then there exists an open subset Ux with the property $\mathcal{P}$ in X such that $x \in Ux$. Since any subspace with the property $\mathcal{P}$ in X has the property $\mathcal{Q}$ in X, then Ux has the property $\mathcal{Q}$ in X such that $x \in Ux$.

Assume that $0 < \alpha < \alpha_0$ and for any point $x \in \cup\{X_\beta : \beta < \alpha\}$ the open set Ux is constructed. Fix a point $a \in X_\alpha$. Then there exist an open subset S of X and a subset F of X with the property $\mathcal{P}$ such that $x \in S \subset \cup\{X_\beta : \beta \leq \alpha\}$ and $S \cap X_\alpha \subset F$. The set $Z = \cup\{X_\beta : \beta < \alpha\}$ is open in X, Z is a subspace with local property $\mathcal{Q}$ in X and $a \in V \setminus Z$. Hence, there exist an open subset Ua of X and a subspace Φ with the property Γ such that $a \in Ua$, $U \subset \Phi \subset Z \cup (F \setminus Z)$ and $\Phi \setminus Z$ has the property $\mathcal{P}$. Since Z and $U \cap Z$ are subspaces with local property $\mathcal{Q}$ in X, the subspace Φ has the property $\mathcal{Q}$. Hence Ua has the property $\mathcal{Q}$ in X. The proof is complete.

Theorem 2.8 opens the possibility of studying $\mathcal{P}$-decomposable spaces using induction and algorithms.

In [12] Theorem 2.8 was proved for regular spaces and for normal T_1-spaces there was introduced the invariant $dim_{\mathcal{P}}X = supremum\{dimF :$ *F has the property $\mathcal{P}$ and it is a closed subset of X*$\}$. For a paracompact $\mathcal{P}$-decomposable space we have $dimX = dim_{\mathcal{P}}X$ ([12], p. 19). This fact follows from Theorem 2.8. It is true, since any regular countable space is zero-dimensional.

Corollary 2.9. *If X is an SP-scattered paracompact space, then* $dimX = 0$.

In [33] it was shown that every paracompact scattered space is zero-dimensional. The authors of [22] mention: "we do not know if this conclusion holds for paracompact SP-scattered spaces". By virtue of above corollary, the response is affirmative.

Remark 2.10. *Let $\mathcal{P}$, $\mathcal{Q}$ and Γ be as in Theorem 2.8 and $\mathcal{Q}$ means that there exists a procedure that allows us to study from a concrete sense the spaces with the property $\mathcal{P}$. Then this procedure can be extended to the procedure to study the $\mathcal{P}$-decomposable spaces with the property Γ. Indeed, assume that the properties $\mathcal{P}$, $\mathcal{Q}$ and Γ satisfy the following conditions:*

- any space with property $\mathcal{P}$ has the property Γ;

- for any space Y with the property Γ and locally with the property $\mathcal{P}$ there exists an algorithm $\mathcal{Q}_1$ to study the space Y;

- if Z is an open non-empty subspace of the space with the property Γ, then for each point $z \in Z$ there exists an algorithm $\mathcal{Q}_2$ to construct an open subset U such that $z \in U \subset Z$ and U is a space with the property Γ;

- if $Y = Z \cup S$ is a space with property Γ, Z is open and locally with the property that exists an algorithm to study locally the space Z and $Y \setminus Z$ has the property $\mathcal{P}$, then there exists an algorithm $\mathcal{Q}_3$ to study the space Y;

- a closed subspace of the space with the property Γ is a space with the property Γ.

Assume that X is a space with the property Γ and the $\mathcal{P}$-decomposition $\{X_\alpha : \alpha < \alpha_0\}$. Then:

1) for any $\alpha < \alpha_0$ and any point $a \in X_\alpha$ we apply the algorithm $\mathcal{Q}_2$ of construction of an open subset Ua such that $a \in Ua \subset \cup\{X_\beta : \beta < \alpha\}$ and Ua is a space with the property Γ;

2) we apply the algorithm $\mathcal{Q}_3$ to study Ua;

3) we apply the algorithm $\mathcal{Q}_1$ to study X.

3 Alexandroff spaces

For a topological space X and the points $a, b \in X$ we put $O(a) = \cap\{U \subset X : a \in U, U \textit{ is open in } X\}$ and $a \preceq b$ if and only if $b \in cl_X\{a\}$. Then $\preceq$ is an ordering on X and it is called the Alexandroff order or the Alexandroff-Birkhoff order generated by the topology of the space

X [2, 10]. A binary relation $\preceq$ on a space X is an order if it is reflexive, antisymmetric and transitive, i.e. for all a, b, $c \in X$, we have that:

- $a \preceq a$ (reflexivity);

- if $a \preceq b$ and $b \preceq a$, then $a = b$ (antisymmetry);

- if $a \preceq b$ and $b \preceq c$, then $a \preceq c$ (transitivity).

For a space X and point $x \in X$ we put $O(x) = \cap\{U \subset X : x \in U, U$ *is open in* $X\}$ and $A(x) = O(x) \cup clX\{x\}$. If $y \in A(x)$, then $x \in A(y)$ and the points x, y are called adjacent points in the space X.

A topological space X is called a pseudo-discrete space or an Alexandroff space if the intersection of any family of open sets is open. By definition, the space X is a pseudo-discrete space if and only if the sets $O(x)$, $x \in X$, are open in X [2, 6].

Any ordering $\preceq$ on a set generates the topology $\mathcal{T}(\preceq)$ with the base $O(x, \preceq) = \{y \in X : y \preceq x\} : x \in X\}$. The topological space $(X, \mathcal{T}(\preceq))$ is an Alexandroff space [2].

Quasi-metric [7, 28] on a set X we call a function $d : X \times X \longrightarrow R$ with the properties:

(M1): $d(x, y) \geq 0$ for all $x, y \in X$;

(M2): $d(x, y) + d(y, x) = 0$ if and only if $x = y$;

(M3): $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

If $d(x, y) = d(y, x)$ for all $x, y \in X$, then the quasi-metric d is called a metric.

A function d with the properties (M1) and (M2) is called a distance on a set X. A function d with the property (M1) is called a pseudo-distance on a set X. A function d with the properties (M1) and (M3) is called a pseudo-quasi-metric on a set X.

Let d be a pseudo-distance on X and $B(x, d, r) = \{y \in X : d(x, y) < r\}$ be the ball with the center x and radius $r > 0$. The set $U \subset X$ is called d-open if for any $x \in U$ there exists $r > 0$ such that $B(x, d, r) \subset U$. The family $T(d)$ of all d-open subsets is the topology on X generated by d. A pseudo-distance space is a sequential space, i.e. a set $B \subset X$ is closed if and only if together with any sequence it contains all its limits [18].

If d is a quasi-metric, then $T(d)$ is a T_0-topology. For any distance that statement is not true.

The pseudo-distance is an integer or a discrete pseudo-distance, if $d(x,y) \in \{0,1,2,...\}$ for any $x,y \in X$ [28, 11]. If d is a discrete quasi-metric on X, then $O(a) = B(a,d,1)$ for any point $a \in X$ and the space $(X, T(d))$ is an Alexandroff space.

If $\preceq$ is an ordering on a set X, then we define two quasi-metrics d_l and d_r on X, where:

- $d_l(x,x) = d_r(x,x) = 0$ and $d_l(x,y) = d_r(y,x)$ for any $x,y \in X$;

- for $x \preceq y$ and $x \neq y$ we put $d_l(x,y) = 1$, $d_l(y,x) = 0$, $d_r(x,y) = 0$, $d_r(y,x) = 1$;

- if $x \not\preceq y$ and $y \not\preceq x$, then $d_l(x,y) = d_r(x,y) = 1$.

In this case $d_s(x,y) = d_r(x,y) + d_l(x,y)$ is a metric. In general, a sum of quasi-metrics is also a quasi-metric, and may not be a metric.

For any points $a,b \in X$ we put $(-,a] = \{y \in X : y \preceq a\}$, $[a,+) = \{y \in X : a \preceq y\}$ and $[a,b] = \{y \in X : y \preceq b\} \cap \{y \in X : a \preceq y\}$. As in [16, 17] we say that V is an f-set if V is open and there exists a point $o_V \in V$ such that $V = [o_V, +)$. Any f-set is an ω-set. The set $L = [a,+)$ is called an ω-set and $o_L = a$.

For any point $a \in X$ we have $B(x, d_l, r) = (-,a]$ and $B(x, d_l, r) = [a,+)$ for any $r \in (0,1]$. Obviously, $T(\preceq) = T(d_r)$ is the topology induced by the ordering $\preceq$.

If X is an Alexandroff space, then any set $V = O(x)$ is an f-set with $o_V = x$.

From the above, it follows:

Theorem 3.1. *For a topological space X the following assertions are equivalent:*

1. X is an Alexandroff space.

2. Any ω-set is an f-set of X.

3. The topology of X is induced by some ordering.

4. The topology of X is generated by some integer pseudo-quasi-metric.

4 Locally finiteness and digital spaces

Let A be an Alexandroff space. On A consider the natural ordering: $a \preceq b$ if and only if $b \in cl_X\{a\}$. We put $a\delta b$ if $a \preceq b$ or $b \preceq a$, i.e. the points a, b are comparable. The space A is a topological digital space if and only if the space A is chain-connected, i.e. for any two points $x, y \in A$ there exist a number $n = i(x, y) \in \mathbb{N}$ and a finite sequence $x_1, x_2, ..., x_n \in A$ such that $x_1 = x$, $x_n = y$ and $x_i\delta x_{i+1}$ for any $i < n$ (see [6, 23, 24, 13, 14, 4, 5, 26, 37]).

A topological space X is called:

- locally finite if each point $x \in X$ has a finite open set containing x.

- strongly locally finite if each point $x \in X$ has a finite open set and a finite closed set containing x.

A local finite space is an Alexandroff space and a scattered space. For any point $x \in A$ we put $f\text{-}dim(x, A) = |O(x)|$ and $f\text{-}dimX = sup\{f\text{-}dim(x, A) : x \in A\}$.

A local finite space is an Alexandroff space and a scattered space. For any point $x \in A$ we put $f\text{-}dim(x, A) = |O(x)|$ and $f\text{-}dimX = sup\{f\text{-}dim(x, A) : x \in A\}$.

We say that a space A is an f-bounded Alexandroff space if there is given a natural number $n \in \mathbb{N}$ such that for any point $x \in A$ there exists an open subset of Wx sequence such that $x \in Wx$ and $|Wx| \leq n$, i.e. $f\text{-}dimA \leq n$.

A connected Alexandroff space is called a topological digital space.

Let $n \in \mathbb{N}$. We say that a space A is a topological n-digital space if for any two points $x, y \in A$ there exists a finite sequence $x_1, x_2, ..., x_n \in A$ such that $x_1 = x$, $x_n = y$ and $x_i\delta x_{i+1}$ for any $i < n$. A singleton space is considered topological 1-digital space. A topological space X is called a bounded digital space if A is a digital space with $f\text{-}dimX < \infty$.

A point $x \in X$ is called a maximal or a closed point of X if the set $\{x\}$ is closed in X. If $\preceq$ is the ordering generated by the topology of the space X, then the maximal points coincide with the maximal points relative to the ordering $\preceq$. If the set $\{x\}$ is open in X, then x

is an initial or an open point of X. Denote by $Max(X)$ the set of all maximal points. If X is a weakly locally finite space, then the initial points coincide with the minimal points relative to the ordering $\preceq$. If $x \in X$ is either open or closed, it is called pure, otherwise it is called mixed [27]. In [16] a maximal point is called a vertex point.

Let $f : X \longrightarrow X$ be a homeomorphism and $a \in X$. Then $f\text{-}dim(f(a), X) = f\text{-}dim(x, X)$ and x is a maximal (initial) point if and only if $f(a)$ is a maximal (initial) point.

We say that a space X is s-homogeneous if for any two points $a, b \in X$ with $f\text{-}dim(a, X) = f\text{-}dim(b, X)$ there exists a homeomorphism $f : X \longrightarrow X$ for which $f(a) = b$. It is obvious that a non-discrete locally discrete space is not homogeneous.

Example 4.1. Let $X = \{1, 2, 3, ..., n, ...\}$ be a space with the topology $\mathcal{T} = \{\emptyset, X\} \cup \{\{1, 2, ..., n\} : n \in \mathbb{N}\}$. By construction, the point 1 is the unique initial point of X and the set of maximal points is empty. The space X is digital and locally finite, $f\text{-}dim(n, X) = n$ for any point $n \in \mathbb{N}$.

Proposition 4.2. *Let γ be a family of open subsets of a space X, $n \in \mathbb{N}$ and $f\text{-}dimX \leq n$. If $Max(X) \subset \cup\gamma$, then γ is a cover of X.*

Proof. First, we prove the following assertions.

Claim 1. If Y is a non-empty subspace of X, then Y is a locally finite space and $f\text{-}dimY \leq f\text{-}dimX$.

This assertion is obvious.

Claim 2. If Y is a non-empty closed subspace of X, then $Y \cap Max(X) \neq \emptyset$.

Let Y be a non-empty closed subspace of the space X. For any point $a \in X$ we put $O(a) = \cap\{U \subset X : a \in U, U \textit{ is open in } X\}$. The set $O(a)$ is open in X. If $a \in O(y)$ and $a \neq y$, then $O(a) \subset O(y)$. Assume that $m = maximum\{|Y \cap O(y)| : y \in Y\}$. Obviously $m \leq n$. Fix $a \in Y$ for which $|O(a) \cap Y| = m$. If $y \in Y \setminus \{a\}$, then $a \notin O(y)$. Hence $\{a\} = Y \setminus \cup\{O(y) : y \in Y \setminus \{a\}$ is a closed subset of X and $a \in Max(X)$.

Claim 3. $X = \cup\gamma$.

The set $Y = X \setminus \cup\gamma$ is closed and $Y \cap Max(X) = \emptyset$. By virtue of

Claim 2, we have $Y = \emptyset$. Hence $X = \cup\gamma$. The proof is complete.

Corollary 4.3. *For a locally finite space X the following assertions are equivalent:*

1. X is a compact space.

2. X is a finite space.

3. f-$dimX < \infty$ and the set $Max(X)$ is finite.

The Claim 2 in the proof of Proposition 4.2 is true for any strongly locally finite space (see [16], Theorem 8).

Let $\mathbb{I} = [0, 1]$ be the unit interval with the usual Euclidean topology generated by the metric $d(x, y) = |x - y|$ for all $x, y \in \mathbb{I}$.

A space X is arc-connected if for any ordered pair of points $a, b \in X$ there exists a continuous function $f : \mathbb{I} \to X$ such that $f(0) = a$ and $f(1) = b$. In this case we say that $f(\mathbb{I})$ is an arc with the endpoints a and b, the point a is the initial point and b is the terminal point of the arc.

Theorem 4.4. *Let X be a T_0-space, $a, b \in X$, $n \in \mathbb{N}$ and if for any two points $x, y \in A$ there exists a finite sequence $x_0, x_1, x_2, ..., x_n \in X$ such that $x_0 = a$, $x_n = b$ and $x_i\delta x_{i+1}$ for any $i < n$. Then there exist the set $\{t_i \in \mathbb{I} : 1 \in \omega(n)\}$ and a continuous mapping $g : \mathbb{I} \to X$ such that $g(\mathbb{I}) = \{r_i \in \mathbb{I} : 1 \in \omega(n)\}$, $0 = t_0 < t_1 < t_2 < ... < t_n = 1$ and $g(t_i) = x_i$ for any $i \in \omega(n)$.*

Proof. We apply the mathematical induction for n.

Let $a = b$ and $n = 1$. In this case $t_0 = 0$, $t_1 = 1$ and $g : \mathbb{I} \to X$ is a constant function with $g(\mathbb{I}) = \{a\} = \{b\}$.

Let $n = 1$ and $a \neq b$. We have two possible cases.

Case 1. $a \preceq b$.

We put $g([0, 2^{-1})) = \{a\}$ and $g([2^{-1}, 1]) = \{b\}$. Since $\{a\}$ is an open set of the subspace $\{a, b\}$ and $\{b\}$ is closed in $\{a, b\}$, the mapping g is continuous.

Case 2. $b \preceq a$.

We put $g([0, 2^{-1}]) = \{a\}$ and $g((2^{-1}, 1]) = \{b\}$. Since $\{a\}$ is a closed set of the subspace $\{a, b\}$ and $\{b\}$ is open in $\{a, b\}$, the mapping g is continuous.

For $n = 1$ the theorem is true. Assume that $m \geq 2$ and the theorem is true for any $n < m$. We put $c = x_{n-1}$. Thus, there exist the set $\{r_i \in \mathbb{I} : 1 \in \omega(n-1)\}$ and a continuous mapping $\varphi : \mathbb{I} \to X$ such that $\varphi(\mathbb{I}) = \{x_i \in \mathbb{I} : 1 \in \omega(n-1)\}$, $0 = r_0 < r_1 < r_2 < ... < r_{n-1} = 1$ and $g(r_i) = x_i$ for any $i \in \omega(n-1)$. We put $t_i = 2^{-1}r_i$ for any $i \in \omega(n-1)$, and put $t_n = 1$.

We have two possible cases.

Case 3. $c \preceq b$.

We put $g(t) = \varphi(2t)$ for each $t \in [0, 2^{-1}]$, $g([2^{-1}, 1)) = \{c\}$ and $g(1)$ $= b$. Since there exists an open subset U of X such that $U \cap \{c, b\} = \{c\}$, the mapping g is continuous.

Case 4. $b \preceq a$.

We put $g(t) = \varphi(2t)$ for each $t \in [0, 2^{-1}]$ and $g((2^{-1}, 1]) = \{b\}$. Since there exists an open subset U of X such that $U \cap \{c, b\} = \{b\}$, the mapping g is continuous.

The proof is complete.

Corollary 4.5. *Any digital space is arc-connected.*

Corollary 4.6. *Any connected local finite space is a digital arc-connected space.*

5 Discrete line and scattered spaces

The image classification problem is to find a fragmentation of an image under study into certain regions such that each region represents a class of elementary partitions (that means a set of pixels or voxtels) with the same label. The regions are separated by boundaries (see [6, 4, 37]). The article [30] is an overview of recent research of generalized topological property in the field of digital image processing.

Digital image processing is by nature a discrete process. This discrete nature causes few problems at the geometric level. At a topological level, this is however different. The notion at the base of topology, the neighborhood, is radically different from continuous spaces to discrete spaces. Algorithms based on topological information are numerous (see [26]).

Assume that the domain X of the plane $\mathbb{R}^2$ represents the image of the original Φ and that image is represented by an observed data function $I : X \to \mathbb{R}$ of the level intensity. We have $I(X) = \{c_i : 1 \leq i \leq n\}$. The function I is constructed in the following way:

- we determine for the image X the levels $\{c_i : 1 \leq i \leq n\} \subset \omega$;

- find a family $\{O_i : 1 \leq i \leq n\}$ of open subsets of X, where the $O = \cup\{O_i : 1 \leq i \leq n\}$ is dense in X, $O_i \cap O_j = \emptyset$ for $1 \leq i < j \leq n$ and O_i is the set of points of the intensity c_i;

- for any $i \in \{1, 2, ..., n\}$ and any $x \in O_i$ we put $I(x) = c_i$;

- if $x \in X \setminus \cup\{O_i : 1 \leq i \leq n\}$, then $I(x) = sup\{i : x \in cl_X O_i\}$;

- by the method of digitalization we construct a finite subset K of X which represents the original image.

In [4] it is considered that $I(x) = c_n$ for any $x \in X \setminus \cup\{O_i : 1 \leq i \leq n\}$. The process of constructing the open sets $\{O_i : 1 \leq i \leq n\}$ is called a fenestration of the topological space X (see [25]).

On $\mathbb{Z} = \{0, 1, -1, 2, -2, ..., n, -n, ...\}$ one can consider one of the following topologies:

- the left topology $\mathcal{T}_l = \{Z_{(-\infty,n)} = \{m \in \mathbb{Z} : m \leq n\} : n \in \mathbb{Z}\} \cup \{\emptyset, \mathbb{Z}\}$;

- the right topology $\mathcal{T}_l = \{Z_{(n,+\infty)} = \{m \in \mathbb{Z} : m \geq n\} : n \in \mathbb{Z}\} \cup \{\emptyset, \mathbb{Z}\}$;

- the topology of Khalimsky $\mathcal{T}_{Kh}$ with the open base $\mathcal{B}_{Kh} = \{\{2n-1\} : n \in \mathbb{Z}\} \cup \{\{2n-1, 2n, 2n+1\} : n \in \mathbb{Z}\}$ [23, 24].

We mention that the function I of the domain X in the Euclidean topology in the space $(\mathbb{Z}, \mathcal{T}_l)$ is continuous.

The space $(\mathbb{Z}, \mathcal{T}_{Kh})$ is called the Khalimsky line, $(\mathbb{Z}^2, \mathcal{T}^2_{Kh})$ is called the Khalimsky plane, $(\mathbb{Z}^3, \mathcal{T}^3_{Kh})$ is called the Khalimsky space.

The Khalimsky's line, plane and space are s-homogeneous scattered locally finite non-compact spaces.

Remark 5.1. *1. Let D be a topological space and $g : D \to \mathbb{Z}$ be a function. For each $n \in \mathbb{Z}$ we put $O(g, n) = \cup\{U \subset X : g(U) = \{n\} : U$ is open in $X\}$. A continuous function f of D in $(\mathbb{Z}, \mathcal{T}_l)$ is an intensity level function on D provided $f(X) = \{0, 1, 2, ..., n\}$ for some $n \in \mathbb{N}$ and $O(f, i) \subset f^{-1}(i) \subset cl_D O(f, i)$ for any $i \in \{0, 1, 2, ..., n\}$.*

2. Any intensity function $f : D \to \mathbb{Z}$ determines on D the property $\mathcal{P}(f)$: a subset U of the subspace Y of the space D has the property $\mathcal{P}(f)$ if the set U is open in Y and $f(U)$ is an open singleton subset of $f(Y)$ as the subspace of the space $(\mathbb{Z}, \mathcal{T}_l)$. Relatively to this property D is a $\mathcal{P}(f)$-scattered space.

Any level intensity function on a space D generates some similarity on D. A similarity measure on a space D is a function of two variables $s : D \times D \longrightarrow R$, where $s(x, y) > 0$ and $s(x, x) - s(x, y) \geq 0$ for any $x, y \in D$ [8, 20, 21, 19, 29].

The space $\mathbb{Z} = \{0, 1, -1, 2, -2, ..., n, -n, ...\}$ is called the discrete line. The digital topologies on $\mathbb{Z}$ are important for the process of digitalization.

We say that the topology $\mathcal{T}$ on $\mathbb{Z}$ is symmetric if $(\mathbb{Z}, \mathcal{T})$ is a scattered Alexandroff space, the set $\{0\}$ is not open in $(\mathbb{Z}, \mathcal{T})$ and for any $n \in \mathbb{Z}$ the mapping $S_n : \mathbb{Z} \to \mathbb{Z}$, where $S_n(x) = 2n - x$ for each $x \in \mathbb{Z}$, is a homeomorphism. If $\mathcal{T}$ is a symmetric topology on $\mathbb{Z}$, then the translations $T_{2n} : \mathbb{Z} \to \mathbb{Z}$, where $T_{2n}(x) = 2n + x$ for all $n, x \in \mathbb{Z}$, are homeomorphisms of the space $(\mathbb{Z}, \mathcal{T})$.

Theorem 5.2. *For a topology $\mathcal{T}$ on $\mathbb{Z}$ the following assertions are equivalent:*

1. The topology $\mathcal{T}$ is symmetric.

2. There exists a non-empty subset $L \subset \{2n-1 : n \in \mathbb{N}\}$ such that:

- $U_0 = \{0\} \cup L \cup \{-n : n \in L\}$ is the minimal open neighbourhood of the point 0 in the space $(\mathbb{Z}, \mathcal{T})$;

- the family $\mathcal{B}(L) = \{T_{2n}(U_0) : n \in \mathbb{Z}\} \cup \{\{2n-1\} : n \in \mathbb{Z}\}$ is an open base of the topology $\mathcal{T}$ on $\mathbb{Z}$.

Proof. Assume that $L \subset \{2n - 1 : n \in \mathbb{N}\}$ is a non-empty subset, $U_0 = \{0\} \cup L \cup \{-n : n \in L\}$ and $\mathcal{B}(L) = \{T_{2n}(U_0) : n \in \mathbb{Z}\} \cup \{\{2n-1\} : n \in \mathbb{Z}\}$. Obviously, $\mathcal{B}(L)$ is an open base of the concrete symmetric topology $\mathcal{T}(L)$ on $\mathbb{Z}$. This fact proves the implication 2 → 1.

Fix a symmetric topology $\mathcal{T}$ on $\mathbb{Z}$. Let V_0 be the minimal neighbourhood of the point 0 and $M = V_0 \cap \mathbb{N}$.

Claim 1. $M \subset \{2n - 1 : n \in \mathbb{Z}\}$.

Assume that $k \geq 1$ and $2k \in M$. Then S_k is a homeomorphism and V_{2k} is a minimal open neighbourhood of the point $2k$ in the space $(\mathbb{Z}, \mathcal{T})$. By construction, we have $k \in V_0 \cap V_{2k}$ and $(\mathbb{Z}, \mathcal{T})$ is not a T_0-space, a contradiction. The Claim 1 is proved.

Claim 2. *$V_0 = \{0\} \cup M \cup \{-n : n \in L\}$ is the minimal open neighbourhood of the point 0 in the space $(\mathbb{Z}, \mathcal{T})$.*

This fact follows from construction and Claim 1.

Claim 3. *The set $\{2n - 1\}$ is open in $(\mathbb{Z}, \mathcal{T})$ for each $n \in \mathbb{Z}$.*

Since $(\mathbb{Z}, \mathcal{T})$ is a scattered space the set $\{a\}$ is open in $(\mathbb{Z}, \mathcal{T})$ for some $a \in \mathbb{Z}$. The points $2n$ are not isolated in the space $(\mathbb{Z}, \mathcal{T})$. Hence $a = 2k - 1$ for some $k \in \mathbb{Z}$. Since $S_{n-k}(2k - 1) = 2n + 1$ and S_{n-k} is a homeomorphism, the set $\{2n + 1\}$ is open in $(\mathbb{Z}, \mathcal{T})$ for each $n \in \mathbb{Z}$. Claim is proved.

From the Claims 2 and 3 it follows that the family $\mathcal{B}(M) = \{T_{2n}(V_0) : n \in \mathbb{Z}\} \cup \{\{2n - 1\} : n \in \mathbb{Z}\}$ is an open base of the topology $\mathcal{T}$ on $\mathbb{Z}$ and $\mathcal{T} = \mathcal{T}(M)$. This fact proves the implication 1 → 2. The proof is complete.

Remark 5.3. *1. The set of symmetric topologies on $\mathbb{Z}$ is oriented by the relation of inclusion. We have $\mathcal{T}(L) \subset \mathcal{T}(M)$ if and only if $L \subset M$. Hence, the topology $\mathcal{T}(L)$ is a minimal symmetric topology if and only if L is a singleton set.*

2. Let $m \in \{0, 1, 2, ...\}$ and $L_m = \{2m + 1\}$. Then the set $H_m = \cup\{n(2m + 1) : n \in \mathbb{Z}\}$ is an open and closed subset of the space $(\mathbb{Z}, \mathcal{T}(L_m))$. We have $\mathbb{Z} = H_m$ if and only if $m = 0$. Hence, the minimal symmetric topology $\mathcal{T}(L_m)$ is a digital topology if and only if $m = 0$.

3. The topology of Khalimsky $\mathcal{T}_{Kh}$ with the open base $\mathcal{B}_{Kh} = \{\{2n-1\} : n \in \mathbb{Z}\} \cup \{\{2n-1, 2n, 2n+1\} : n \in \mathbb{Z}\}$ is of the form $\mathcal{T}(L)$ for $L = \{1\} = L_0$. Therefore the topology of Khalimsky is the unique minimal digital symmetric topology on the discrete line $\mathbb{Z}$.

Acknowledgments. The second author's research was supported by the World Federation of Scientists.

References

[1] S. Abramsky S. and A. Jung. *Domain Theory*, in: S. Abramsky, D. Gabbay and T.S.E. Maibaum (eds.) *Handbook of Logic in Computer Science*, Oxford: Oxford University Press, 1994, pp. 1–168.

[2] P. Alexandroff. *Diskrete Räume*, Matem. Sbornik (N.S.), vol. 2 (1937), pp. 501–518.

[3] P.S. Alexandroff and I.V. Proskuryakov. *On reducible sets*, Izvestia Akad. Nauk SSSR, Matematika, vol. 5 (1941), no. 3, pp. 217–224.

[4] M. Al-Hajri, K. Belaid, L.J. Belaid. *Scattered spaces, compactifications and an application to image classification problem*, Tatra Mountains Math. Publ., vol. 66 (2016), pp. 1–12.

[5] G. Aubert and P. Kornprobst. *Mathematical problems in image processing*, Series: Appl. Math. Sci. 147, Springer-Verlag, New York, 2002.

[6] F.G.Arenas. *Alexandroff spaces*, Acta Math. Univ. Comenianae, vol. 68 (1999), no. 1, pp. 17–25.

[7] A.V. Arhangelskii. *Mappings and spaces*, Russian Math. Surveys, vol. 21 (1966), no. 4, pp. 115–162.

[8] J.L. Bentley B.W. Weide, and A.C. Yao. *Optimal expected-time algorithms for closest point problems*, ACM Trans. Math. Software, vol. 6 (1980), pp. 563–580.

[9] J. Blanck. *Domain representations of topological spaces*, Theoretical Computer Science, vol. 247 (2000), pp. 229–255.

[10] G. Birkhoff. *Rings of sets,* Duke Math J., vol. 3 (1937), no. 3, pp. 443–454.

[11] M.M. Choban and I.A.Budanaev. *About Applications of Distances on Monoids of Strings*, Computer Science Journal of Moldova, vol. 24 (2016), no. 3, pp. 335–356.

[12] M.M. Choban and N.C. Dodon. *Theory of P-dispersed spaces*, Kishinev, Ştiinţa, 1979 (in Russian).

[13] B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order*, Cambridge: Cambridge University Press, 1990.

[14] D.E. Denning. *A lattice model of secure information flow*, Communications of the ACM., vol. 19 (5) (1976), pp. 236–243.

[15] N.C. Dodon, M.M. Choban. *Mappings of P-scattered spaces*, C.R. Acad. Bulgare, vol. 28 (1975), nr. 4, pp. 441–443 (in Russian).

[16] Ulrich Eckhardt and Longin Latecki. *Digital Topology*, Hamburger Beiträage zur Angewandten Mathematik Reihe A, Preprint 89, October 1994, 31 p.

[17] Yu.L. Ershov. *Definability and Computability*, Springer, 1996.

[18] R. Engelking. *General Topology*, Warszawa: PWN, 1977.

[19] J.B. Hart and C.Tsinakis. *A concrete realization of the Hoare powerdomain*, Soft Comput., vol. 11 (2007), pp. 1059–1063.

[20] J.E. Hellerstein, E. Koutsoupias, D.P. Miranker, C.H. Papadimitriou, and V. Samoladas. *On a model of indexability and its bounds for range queries*, Journal of the ACM, vol. 49 (2002), no. 1, pp. 35–55.

[21] J.E. Hellerstein, E. Koutsoupias, D.P. Miranker and C.H. Papadimitriou. *On the analysis of indexing schemes*, in: Proceedings of the Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Tucson, Arizona, 12-15 May 1997, Arizona, 1997, pp. 249–256.

[22] M. Henriksen, R. Raphael, R.G. Woods. *SP-scattered spaces; a new generalization of scattered spaces*, Comment. Math. Univ. Carolin., vol. 48 (2007), pp. 487–505.

[23] E. Khalimsky. *Topological structures in computer science*, Journal of Applied Math. and Simulation, vol. 1 (1987), no. 1, pp. 25–40.

[24] E. Khalimsky, R. Kopperman and P.R. Meyer. *Computer graphics and connected topologies on finite ordered sets*, Topol. Appl., vol. 36 (1990), no. 1, pp. 1–17.

[25] E.H. Kronheimer. *The topology of digital images*, Topology and its Applications, vol. 46 (1992), pp. 279–303.

[26] J. Lamy. *Integrating digital topology in image-processing libraries*, Computer Methods and Programs in Biomedicine, vol. 85 (2007), no. 1, pp. 51–58.

[27] Erik Melin. *Locally finite spaces and the join operator*, Proceedings of the 8th International Symposium on Mathematical Morphology, Rio de Janeiro, Brazil, Oct. 10-13, 2007, MCT/INPE, vol. 1, 2007, pp. 63–74.

[28] Stoyan Y. Nedev. *o-metrizable spaces*, Trudy Moskov. Mat.Ob-va, vol. 24 (1971), pp. 201–236 (English translation: Trans. Moscow Math. Soc., vol. 24 (1974), pp. 213–247).

[29] Vladimir Pestov. *On the geometry of similarity search: dimensionality curse and concentration of measure*, Information Processing Letters, vol. 73 (2000), pp. 47–51.

[30] Nibedita Roy, Ajit Das. *Topological Structure in Digital Image Processing : A Survey*, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 2 (2017), no. 6, pp. 481–485.

[31] R.S. Sandhu. *Lattice-based access control models*, Computer, vol. 26 (11) (1993), pp. 9–19.

[32] D. Spreen. *On Some Decision Problems in Programming*, Information and Computation, vol. 122 (1995), no. 1, pp. 120–139.

[33] R. Telgarsky. *Total paracompactness and paracompact dispersed spaces* , Bull. Acad. Polon. Sci. Ser. Math. Astronom. Phys., vol. 16 (1968), pp. 567–572.

[34] R. Telgarsky. *C-scattered and paracompact spaces*, Fund. Math., vol. 73 (1971), pp. 59–74.

[35] R. Telgarsky. *On topological properties defined by games*, Preprint no. 46, Institute Math. Polish Acad. Sci., 1972.

[36] R. Telgarsky. *On topological properties defined by games*, Topics in Topology, Budapest, 1974, pp. 617–624.

[37] J. Weickert. *Efficient image segmentation using partial differential equations and morphology*, Pattern Recognition, vol. 34 (2001), pp. 1813–1824.

Mitrofan M. Cioban[1], Ivan A. Budanaev[2]

[1]Tiraspol State University
Email: `mmchoban@gmail.com`

[2] Institute of Mathematics and Computer Science, Republic of Moldova
Email: `ivan.budanaev@gmail.com`

On technology of free access to the characteristic cultural heritage in the form of old printed texts

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

Abstract

This paper is dedicated to challenges, approaches, and problems in free access to the cultural heritage. Creation of linguistic resources from old Romanian printed text is considered. The discussed platform consists of electronic linguistic resources, the access tools, texts processing tools, and service and management software.

Keywords: cultural heritage, old Romanian printed texts, digitization, lingustic software platform, big data.

1 Introduction

Presented research concerns modern challenge of enabling online and cross-border access to digital resources of European heritage. The challenge arose from the fact that only small part of Europe's cultural heritage is digitized and even digitized resources have low visibility especially across national boundaries. Digital resources of European heritage are potentially significant for cultural and creative economy sectors, which are today essential directions for achieving economical advantages.

Free access means providing cultural heritage digitization-and-mining services on the basis of web-platforms. The challenge encourages applying new technologies of computer science generally and, in particular, big data management, for developing supporting methods and techniques. Another specific feature of this challenge is the importance of understanding the rich diversity of the cultural heritage and, therefore, considering spatial-temporal dimension in developing of both access tools and resources.

This paper regards old Romanian books in the Cyrillic script. The Romanian language can be classified as “under-resourced”. Until recently, most of the research work in Natural Language Processing (NLP) has been focused on a few well-described languages with many researchers. Under-resourced languages nevertheless pose important scientific challenges. NLP for under-resourced languages tends to be carried out in isolated or sparse research groups, and the resulting products are often in different formats and standards. Discovering, accessing, and making those resources interoperable can become a challenge in itself. When dealing with under-resourced languages, issues of interoperability of data and metadata become of crucial importance for combining and re-using the few resources and tools that might be available.

The platform that provides easy access to textual cultural heritage is being developed basing on advanced computer science methods and techniques obtained from our large portfolio of computational linguistic related developments, as well as from our experience in HPC and Big Data processing. It will answer spatial-temporal dimension of the challenge granting access to a unique content. The cultural heritage of old Romanian books is relatively small but reflects specific features of Romanian. It keeps not only the special pronunciation and orthography of different epochs and regions but also history of Romanian book printing.

The platform combines digitized textual resources with free access and software for linguistic analysis. Resources are created in the life-circle of scanning, recognizing, and application of linguistic software like web-crawler, lemmatization, normalization, diachronic analysis, etc. Support for corpora will be also included.

2 Producing linguistic resources from old Romanian books

Electronic linguistic resources are, in a broad sense, any computer tools to support the linguist's work. They include both computer programs and linguistic data. In many cases, programs and data are inextricably linked, for example, in corpora and databases.

Linguistic data are divided into primary and secondary. In linguistics, the primary data are spoken or written expressions not interpreted linguistically. Secondary data appear when primary data are processed by appropriate programs. In this paper we consider the resources of secondary linguistic data obtained from old Romanian printed texts.

The starting point of our work is the scanned text, i.e., the text presented in the form of page images. There are a lot of big electronic libraries of texts in this form, e.g., Bucharest Digital Library[1]. This form of presentation does not allow any processing except reading it by an expert. Our task is to get the symbolic representation of the scanned text, annotate it, and then perform further processing.

Technology for recognition of the historic and linguistic Romanian heritage printed in the Cyrillic script in the 17th–20th centuries is supported by a pack of the following tools and utilities [1]:

- Alphabets for ABBYY FineReader (AFR).
- Dictionaries (word lists) for AFR.
- Recognition patterns as trained under AFR.
- Utility of transliteration from Cyrillic script to modern Latin and vice versa.
- Selection utility to start AFR with the alphabet, dictionary, and templates corresponding to a specific epoch and location.
- Virtual keyboard.

Specific Romanian Cyrillic script (RC) of the mid 18th century till 1830 is characterized by two substantial differences from that of the older time, with the same RC of up to 47 letters. First, the usual Arabic number system is used. Second, upper accents had become

[1] `http://www.digibuc.ro/`

rare and may be ignored. Therefore, the recognition doesn't imply sophisticated training.

AFR recognizes RC of the corresponding period. The recognition of texts of the 18th century resulted in WER (Word Error Rate) of 3–4.5%.

The previous period covers the 17th century and the 1st half of the 18th century when the Romanian typographies had strictly adhered the previous manual writing practices. This means that the numbers were encoded by letters with special ascending strokes, and accents over the line were substantial. Some words were traditionally printed with abbreviations and were also marked over them. Skipped letters were frequently set over the precedent letter, also with a special marker.

The recognition of such printing implies very subtle and thorough training. For example, each pair of a letter and another letter over it should be trained as a ligature.

Numbers (one or several letters with a marker) should also be trained as ligatures. This increases the number of recognition patterns, but decreased WER down to 6%.

Post-processing of digitized text is a complex task. To solve it, we are developing software that supports expert's efforts in improvement and analysis of the recognized texts.

The highest priority task of post-processing is to minimize errors in the recognized text.

As this purpose is met, we can use the text for:

- philological research, e.g., completion of corpora, or scientific re-printing;
- presentation to wide public, in electronic or re-printed form;
- making the text subject of full-text search;
- extraction of data for better subsequent of repeated recognition;
- statistics, e.g., error or accuracy rates.

The recognized text may show several types of errors. If we denote original word (fragment, text) as A, recognized word as B, and corrected word as C, the ideal case is $A = B = C$. Inequalities means errors. Possible cases are:

- $A = B$, $A \neq C$: false correction;
- $A \neq B$, $B = C$: unseen error;
- $A = C$, $B \neq C$: corrected error;
- $A \neq B$, $B \neq C$, $A \neq C$: unclear situation for expert solution.

Error correction may be manual, semi-automatic, or even automatic.

Manual correction is performed by an expert (philologist, native language speaker, etc.) The process is expensive and error-prone. A spell checker with a historical word list may be useful. A helpful additional feature is global correction: each time the expert corrects the text, the similar places are shown as the list, some may be unchecked but other will be corrected at once.

Semi-automatic or automatic correction tools use different error models. For example, the template-driven tool may provide the replacement of "cn" at the end of words to "en" that is suitable for many European languages, etc.

For philological research, original look and feel of the text should be kept but a lot of additional information is necessary.

The current practice is to produce multi-layer text presentation, usually as an XML file. One of layers keep original text (images). Other layer may add morphological and syntactic information for each word, etc.

With old texts, one of layers may contain the normalized text. Ambiguous spellings of the same word appearing in old texts as local or temporal peculiarities are replaced by a unified standard presentation.

For old Romanian, the corresponding transliteration of the word in the modern script and orthography may present normalization of this word in almost all cases.

Such multi-layer XML presentation, especially being correspondent to some of existing standards (XML-TEI,[2] etc.), may be used later, e.g., for completion of corpora.

The normalized text may be a suitable representation for mass users, or the full-text search.

[2] `http://www.tei-c.org/index.xml`

In our view, software tools for experts are as follows:

- Multi-window editor, containing source images, recognized text, and corrected text. Alignment should be implemented that makes it possible to navigate using any window.
- Access semi-automatic tools, like selection of suspected fragments, or global correction.
- Access automatic tools like statistics, tokenizer, or POS-tagger, one by one, or all at once.
- Access tools like lexicon and corpora management, normalization, transliteration, or XML output.

3 Example of resource evaluation: 18th century book on geography

Quality of the secondary linguistic resources is of great importance and should be thoroughly checked. Let's take as an example the 18th century Romanian text on geography [3]. The book was processed by our technology [1]. The resulting text was compared with the text provided by M. Onofrei (A.-I. Cuza University, Iasi).

Quality of recognition and post-processing can be measured in accuracy or error rates. These rates can be calculated for words, or for characters.

Character accuracy seems to be more adequate measure because word accuracy depends on error distribution in words. The question is exhaustively discussed in [2, Sec. 5]. In their example, 600 characters with 30 errors (95% character accuracy) and word length 6 may give from 95% down to 70% word accuracy, the latter corresponds to the distribution of one erroneous character per word. These calculations are especially applicable because the mean word length in old Romanian texts is between 5 and 6. Moreover, in our experience, errors were mostly distributed as one per word.

The aforementioned text contains 8020 words. Comparing with modern orthography, 756 words (9.4%) differ. OCR errors: 115 (less than 2%). Transliteration errors: 54 (0.7%). Difference in writing in one word or separately: 301 (3.6%); it's historical development of orthography. Different approaches to transliteration: 20 (0.2%). Remaining 266 words (3.4%) are subject to philological expertize. For example, the word "up to" was printed both as "până" (modern norm) and "pănă". It may be local or historical variation, or even simple misprint. All these results are clearly shown in the diagram (Fig. 1).

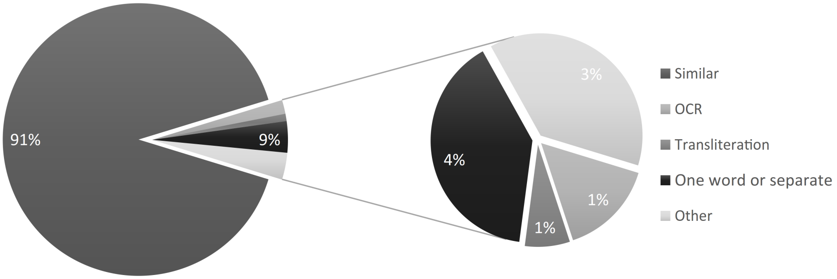


Figure 1. Word error statistics for a Romanian text on geography, the 18th century.

4 Software for lingustic processing and its application

Linguistic processing of texts uses diverse software. They include annotation, segmentation, tokenizing, POS-tagging, syntax analysis, semantic analyzing, concordance, etc. Let us consider diachronic analysis as an example, because it is based on historical nature of resources we create.

Diachronic ("across time") linguistics (aka historical linguistics) concerns the process of language development over time. Diachronic analysis learns in which way language changes spread across spatio-temporal dimension. Its tools is one of the most popular today software of NLP. Their application supposes digitization of huge collections of historic texts and creation the corresponding corpora. Diachronic analysis tools were described in [5].

The presented paper focuses on resource specific features, which have to be implemented to maintain diachronic analysis. Actually any historical text based corpus can be made diachronic by slicing into time slices by natural time component. Since the resources of the proposed platform will be created together with tools, we intend to add time component in metadata, nevertheless keeping in mind the possibility of automatic corpus completion from external resources.

Three manuals in elementary mathematics were selected to experiment with diachronic analysis. These manuals were of the 18th [6], 19th [7], and 21th century [8]. The first one was printed in parallel in German and Romanian (old Cyrillic), the second is in Romanian (Cyrillic), and the third is in Romanian (Latin).

Until present, the following stages were finalized:

- Approx. 30 pages from each book were selected.
- Selected pages from books [6,7] were recognized by fineReader 12 and transliterated to the Latin script.
- Transliterated pages were edited to correct spelling.
- All "stopwords" were deleted from all three texts with a dedicated Java program.

Finally the LDA (latent Dirichlet allocation) model will be apllied to compare mathematical terms of the three epochs.

5 On interface of the proposed platform

The presented research provides free access to characteristic textual cultural heritage by web-based platform. The platform will combine textual corpus with collection of methods and algorithms, which implement accessible services. The general platform architecture consists of: a presentation component, a user services component, components for accessing and/or processing resources, and resources themselves.

A prototype of web interface of presentation component, concerning only digitization, is described in work [4]. Web interface of components for accessing resources will be developed basing on previously developed one.

The main task in development of interface is, for today, management of user services. Web interface of user services component has to give access to whole functions set supported by platform. It is to be intuitive enough, and supplied by help being mainly used by researchers-humanitarians.

We intend to employ high performance techniques and equipment, and the corresponding approaches and algorithms, to guarantee efficient and improved access. Therefore the platform has to provide gate to high performance processing resources. The gate will connect presentation component with the component that implements the HPC scheduling and performing.

6 Conclusion

The proposed platform solves problems at creation of secondary linguistic resources with the accompanying processing software. The platform will smoothly support data protection, copyright, and archiving. With its user-friendly interface, these and other features make the developed system highly useful for professional and non-professional users. It will answer both challenges related to European heritage: free access to historical resources, and exposing characteristic features of each culture in multicultural European context.

References

[1] S. Cojocaru, A. Colesnicov, L. Malahov. *Digitization of Old Romanian Texts Printed in the Cyrillic Script.* Second International Conference on Digital Access to Textual Cultural Heritage DATeCH-2017, Goettingen, June 1-2, 2017, pp. 143–148.

[2] U. Springmann, A. Lüdeling. *OCR of historical printings with an application to building diachronic corpora: A case study using the RIDGES herbal corpus*, arXiv:1608.02153 [cs.CL]

[3] Amfilohie Hotiniul. *De obște gheografie*, Iași, 1795.

[4] T. Bumbu, S. Cojocaru, A. Colesnicov, L. Malahov, S. Ungur, *User Interface to Access Old Romanian Documents.* CMSM42017, June 25-July 2, 2017, Chisinau, Republic of Moldova.

[5] L. Burtseva, V. Demidova. *Building of P system based tools for diachronic text analysis*, Proceedings of the 4th Conference of Mathematical Society of Moldova CMSM4'2017, June 25-July 2, 2017, Chisinau, Republic of Moldova, pp.483-487.

[6] J.I. von Felbiger. *Anleitung zur Rechenkunst / Ducere de mână către aritmetică,* Viena, 1785.

[7] Gh. Asachi. *Elemente de matematică. Partea 1 : Aritmetică*, Iasi, 1836.

[8] V.O. Ignătescu. *Breviar teoretic pentru pregătirea de examenul de matematică, clasa VIII-a*, Buzău, 2010.

S. Cojocaru[1,2],L. Malahov[1,3], A. Colesnicov[1,4], T. Bumbu[1,5], L. Burtseva[1,6], V. Demidova[1,7]

[1]Institute of Mathematics and Computer Science
5 Academiei str., MD-2028, Chisinau
Republic of Moldova
[2]Email: `svetlana.cojocaru@math.md`
[3]Email: `lmalahov@gmail.com`
[4]Email: `acolesnicov@gmx.com`
[5]Email: `tudorbumbu10@gmail.com`
[6]Email: `luburtseva@gmail.com`
[7]Email: `valentina.demidova@math.md`

Analysis and preparation of data from Stroke.md database when creating a stroke prediction model

Svetlana Cojocaru, Constantin Gaindric,
Galina Magariu, Tatiana Verlan

Abstract

Data preparation when creating a prediction model in medicine, is the most time-consuming and labour-intensive stage. To the most degree the model performance depends on data quality. This work describes the process of data analysis and preparation, and features selection at stroke prediction model creation.

Keywords: data preparation, data analysis, features selection, stroke prediction model

1 Introduction

Stroke is one of the leading causes of mortality, morbidity and serious long-term loss of ability to work. So, stroke prediction is an important task for deciding the way of patient's living mode and treatment.

Statistics on stroke, provided by WHO data for May 2014, is very threatening. Thus, Moldova ranks 32^{nd} in the world and 3^{rd} in Europe in mortality due to stroke. On the other hand, stroke is the second cause of death in Moldova. And here Moldova "keeps pace with the whole world": stroke is the second cause of death in the world.

Surely, that world medical community is concerned by searching risk factors for prevention this terrible disease, leading to deaths or disability of population. The scientific world is also concerned with this problem for several dozens years already, creating different mathematical stroke prediction models. There are a lot of mathematical models, which are in some way connected with stroke or predicting stroke. Each of them has some nuance, narrow focus or specific character.

2 A brief review of the literature on models associated with stroke

In [1-13] some existing models somehow related to stroke are described.

I.V.Sidyakina et al. describe in [1] the predictive model of lethality and functional recovery evaluation after severe and extremely severe stroke. The model forecasts assessment according to Bartel scale in 1, 3, 6 and 12 months after stroke. Purpose of the model is to foresee the likely scenario of the disease progression basing the set of initial underlying features. If the scenario becomes better as the result of treatment, then the therapy is considered to be effective, and vice versa.

Eric E. Smith et al. in [2] present a stroke risk model, which provides clinicians with a practical bedside tool for death risk stratification for In-Hospital Ischemic Stroke Mortality. Data for this study were provided by 1036 hospitals that contributed 274 988 ischemic stroke patients between October 2001 and December 2007. Determining mortality risk for every patient at admission provides valuable prognostic information to patients, their family members and medical personnel by identifying those who are at high risk for poor outcomes.

The group of researchers from China introduced a Two-Level Model [3] for the Analysis of Syndrome of Acute Ischemic Stroke. In total, 166 acute ischemic stroke patients within 72 hours after their ictus were enrolled from several hospitals located in Northern China from August 2007 to December 2009. They were aged at 35 – 75, without some specific predefined diseases. A questionnaire of symptoms with 102 records was formulated on wind-phlegm collateral obstruction syndrome.

Three developed models for predicting functional outcome after stroke are described in [4]. The authors aimed to assess if the performance of stroke outcome models with simple clinical variables could be improved by the addition of more complex clinical variables and information from the first computed tomography (CT) scan. 538 acute ischaemic and haemorrhagic stroke patients were enrolled in the study between 2001 and 2002. Independent survival was assessed at 6 months.

Authors of [5] proposed stroke mortality prediction score early after hospitalization of patients with acute ischemic stroke. This was a retrospective study which included 12 262 community-based patients with an acute ischemic stroke at hospitals in Ontario, between 2003 and 2008. Their stroke index constitutes an objective tool to stratify mortality risk of individual patient at 30 days and 1 year.

In [6] the authors describe their models predicting survival and functional recovery within 3 months after acute stroke. Patients are described with regard to age and NIHSS score assessed within 6 hours after stroke. The models do not consider imaging or laboratory investigations, which were impossible to obtain in large original cohorts within an early time frame and a standardized evaluation protocol. Models are focused on readily accessible variables that do not require a sophisticated technique and rigorous time frame. They were developed with data from the stroke data bank of the German Stroke Collaborators. Data set of the Virtual International Stroke Trials Archive (VISTA) was used for models testing. VISTA has data for >15 000 patients, but only 5843 of them met the specified criteria used for test dataset.

The study [7] concerned predicting the recovery of dextrous function in the paretic hand at 6 months after acute stroke. 57 patients within 5 days after stroke were recruited, recovery of dextrous hand function was assessed weekly (the first 4 weeks), then monthly till 6 months after onset. The main purpose was to determine clinical characteristics for predicting. The 7 candidates became: side and site of brain infarct, stroke severity, cognition, spatial neglect, two-point discrimination, muscle tone and muscle strength of the paretic upper extremity.

Surely, there are researches as well as the mathematical models which predict the stroke itself. But such models are much less.

Aditya Khosla et al. in their study [8] compare the Cox proportional hazards model with a machine learning approach for stroke prediction on the Cardiovascular Health Study (CHS) dataset. They discuss important for medical datasets problems of data imputation, feature selection, and prediction, present a margin-based censored regression algorithm. The authors also managed to find potential risk factors that have not ever been found by traditional approaches. In their experiments they considered the 5-year stroke prediction problem on CHS. At first they excluded from the initial data set the persons with pre-baseline stroke, then removed the features with more than 60% missing entries. The resulted data set included 796 features for 4988 persons with 299 occurrences of stroke. Then they used L1 regularized logistic regression and reduced the number of features to about 200. At last using feature selection procedure for this reduced set they achieved the final set with 19 features (model M3 in Fig. 3).

Authors of [9] make a comprehensive investigation of the risk factors for strokes in the context of the CHS, applying Bayesian model averaging to the selection of variables in Cox proportional hazard models.

The objective of the study [10] was to construct a stroke prediction model for an elderly U.S. population (see model M2 in Fig. 3), and to assess the accuracy of other previously published prediction models in this population. The subjects were participants in the CHS: 2,495 men and 3,393 women of age 65 and older at baseline, and followed for 6.3 years. Among 5,711 participants free of baseline stroke, 399 strokes occurred. In this study the authors compare their model with previously published Framingham Study model and the Israeli Ischemic Heart Disease Project model. So, they had to find equivalents between the dataset used by them and the datasets of these two models.

There is a research based on the Framingham Heart Study (FHS) cohort (model M1 in Fig. 3, [12]). Stroke probabilities were computed using the Cox proportional hazards model for each sex. The examined cohort (2,372 men and 3,362 women of age 55-84 years, free of stroke at baseline) was followed up during 10 years. 472 stroke events have occurred during this time in that cohort. Based on the stroke risk factors (model M1 in Fig. 3), the values for which can be easily obtained during usual medical examination, a stroke risk can be evaluated.

There is a recent study [13] (model M5 in Fig. 3), notable for the number of examined subjects (3182325 men and 2532986 women) followed up during 10 years (starting from 2002-2003 until 2013). It used the Korean national health examination data. As the result, the model was developed for predicting the prevalence of stroke within 10 years, the level of stroke risk was classified into 5 categories. Based on the model, the personalized warnings and the lifestyle correction messages were formulated and uploaded to the personal health record service "*My Health Bank*".

The mentioned above models predicting stroke itself [8-10, 12] are based on such well-known datasets as CHS and FHS. These datasets are worth of special consideration (Section 3).

From the other hand, there are models based on small datasets, e.g., the authors of [11] (model M4, Fig. 3) present a system using artificial neural networks for prediction of Cerebrovascular Accident (CVA) risk, recognizing four its levels. The training dataset is: 108 patients, 11 risk factors. The model simulates human brain activity and, by entering necessary values of the new patient, it predicts his level of CVA risk.

3 Short introduction to CHS and FHS

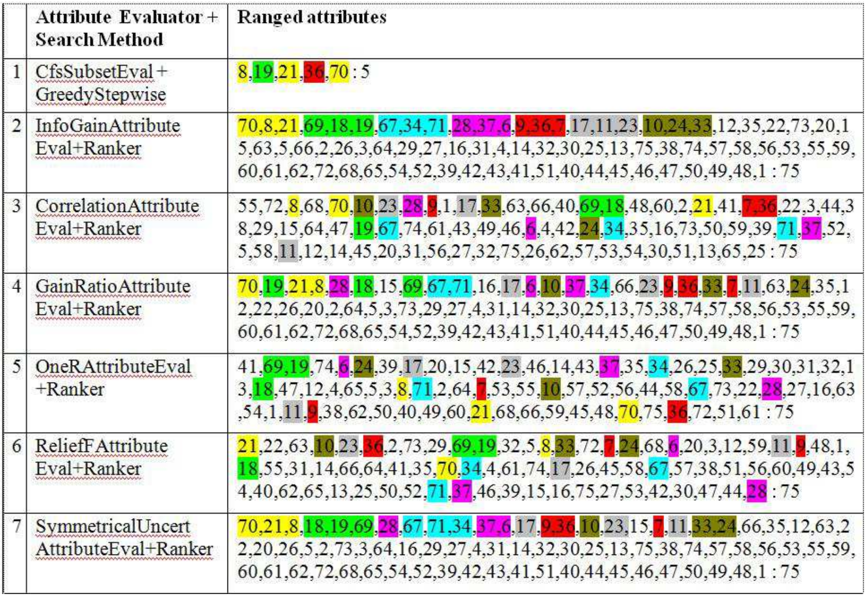
The CHS (https://chs-nhlbi.org/) was initiated by the National Heart, Lung and Blood Institute (NHLBI) in 1987 to identify risk factors for development and progression of cardiovascular disease (CVD) related to the onset of coronary heart disease and stroke in adults of 65 years or older: 5888 participants from 4 U.S. communities. Between 1989–1999 all the participants underwent extensive annual clinical examinations. Every half a year between clinic visits, and if clinic visits ended, participants were called by phone to define their health status and hospitalizations. The main results are coronary heart disease (CHD), angina, heart failure (HF), stroke, transient ischemic attack (TIA), claudication, and mortality. Exams from the very beginning included a home interview and a clinic examination that evaluated traditional risk factors for CVD and measures of subclinical disease (as carotid ultrasound, echocardiography, electrocardiography, and pulmonary function). Later on such new components as Cranial MRI scans, retinal photography, and tests of endothelial function were added. Since 1999, participants continue to be contacted every 6 months by phone, to ascertain health status. Adjudication of cause of death continues, but adjudication of other events ended in June 2015. Data from the CHS has led to more than 1,500 published research papers. As the cohort has aged, CHS has become an important study of the aging process.

The FHS (https://www.framinghamheartstudy.org/fhs-about/) is also a project of the NHLBI, since 1971 – in collaboration with Boston University. It is a long-term, ongoing cardiovascular cohort study on inhabitants of the town Framingham. It began in 1948, the original cohort: 5209 adults of age 30 – 62 at the time of first examination and having no history of heart attack or stroke. The study had been intended to last 20 years. But, by the end of this period, the study continued. Now it is on its third generation of participants. There are several cohorts, reflecting in general different generations. One of the cohorts founded in 1994, considers race and heritage in heart factors. The FHS participants, their children and grandchildren, voluntarily agreed to undergo a detailed medical history, physical examination, and medical tests every two years, creating a wealth of data about physical and mental health, especially about cardiovascular disease. In the past half century, FHS has produced almost 1,200 articles in leading medical journals. Concept of CVD risk factors has become an integral part of the modern medical curriculum and has led to the development of effective treatment and preventive strategies in clinical practice.

4 Parameters analysis and selection

The database Stroke.md [14] (DB) was created for neurologic department of Emergency Medicine Institute (Moldova) and intended to contain patients of two types: population in stroke risk group (3 risk levels), patients with stroke. At first the DB contained information for only 32 patients with stroke. Using these data (and elaborating the approach of their analysis) we tried to create the model for patients' classification on groups of stroke risk level [15]. Afterwards the information of two types was populated into the DB: patients who underwent routine examinations in some rural regions (96); patients who had arrived in the hospital with the diagnosis "stroke" (137). In total, information about 233 patients was included into DB. These were subjects for stroke prediction model creation. The dataset for this purpose was obtained from this DB, extracting the information on 99 parameters into the easy-to-use file (with parameters' order numbers). Analysis of data regarding their fullness and representation correctness was carried out.

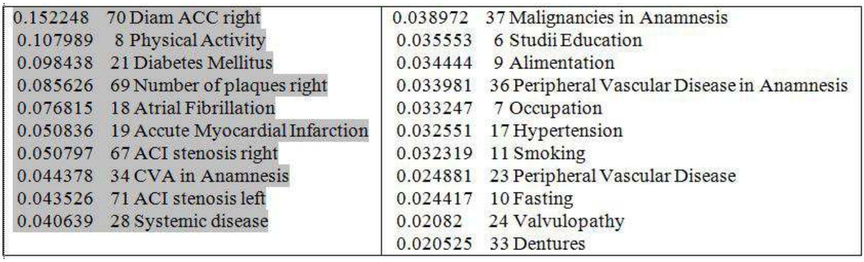
Data analysis was carried out applying two different approaches. **First approach:** Exclusion of the parameters which cannot serve as a basis of predictive model since their values for the majority of patients were missing. Afterwards, rows (patients), in which values for one or more parameters were missing, have been excluded too. After several iterations, we have got 69 rows with 76 parameters – this relation "number of patients" to "number of parameters" is incomparable.



	Attribute Evaluator + Search Method	Ranged attributes
1	CfsSubsetEval + GreedyStepwise	8,19,21,36,70 : 5
2	InfoGainAttribute Eval+Ranker	70,8,21,69,18,19,67,34,71,28,37,6,9,36,7,17,11,23,10,24,33,12,35,22,73,20,15,63,5,66,2,26,3,64,29,27,16,31,4,14,32,30,25,13,75,38,74,57,58,56,53,55,59,60,61,62,72,68,65,54,52,39,42,43,41,51,40,44,45,46,47,50,49,48,1 : 75
3	CorrelationAttribute Eval+Ranker	55,72,8,68,70,10,23,28,9,1,17,33,63,66,40,69,18,48,60,2,21,41,7,36,22,3,44,38,29,15,64,47,19,67,74,61,43,49,46,6,4,42,24,34,35,16,73,50,59,39,71,37,52,5,58,11,12,14,45,20,31,56,27,32,75,26,62,57,53,54,30,51,13,65,25 : 75
4	GainRatioAttribute Eval+Ranker	70,19,21,8,28,18,15,69,67,71,16,17,6,10,37,34,66,23,9,36,33,7,11,63,24,35,12,22,26,20,2,64,5,3,73,29,27,4,31,14,32,30,25,13,75,38,74,57,58,56,53,55,59,60,61,62,72,68,65,54,52,39,42,43,41,51,40,44,45,46,47,50,49,48,1 : 75
5	OneRAttributeEval +Ranker	41,69,19,74,6,24,39,17,20,15,42,23,46,14,43,37,35,34,26,25,33,29,30,31,32,13,18,47,12,4,65,5,3,8,71,2,64,7,53,55,10,57,52,56,44,58,67,73,22,28,27,16,63,54,1,11,9,38,62,50,40,49,60,21,68,66,59,45,48,70,75,36,72,51,61 : 75
6	ReliefFAttribute Eval+Ranker	21,22,63,10,23,36,2,73,29,69,19,32,5,8,33,72,7,24,68,6,20,3,12,59,11,9,48,1,18,55,31,14,66,64,41,35,70,34,4,61,74,17,26,45,58,67,57,38,51,56,60,49,43,54,40,62,65,13,25,50,52,71,37,46,39,15,16,75,27,53,42,30,47,44,28 : 75
7	SymmetricalUncert AttributeEval+Ranker	70,21,8,18,19,69,28,67,71,34,37,6,17,9,36,10,23,15,7,11,33,24,66,35,12,63,22,20,26,5,2,73,3,64,16,29,27,4,31,14,32,30,25,13,75,38,74,57,58,56,53,55,59,60,61,62,72,68,65,54,52,39,42,43,41,51,40,44,45,46,47,50,49,48,1 : 75

Figure 1. Ordered parameters

To improve this relation we have made selection of the most significant parameters using 7 ranking methods proposed by WEKA (https://www.cs.waikato.ac.nz/ml/weka/) [16]. Numbers in the lists of ranged attributes reflect parameters order in the dataset by degree of their influence over the target parameter (Fig.1). We have selected 21 most significant parameters for all the applied methods, and then we have left only 10 of them, besides the target parameter *Ischemic CVA* (Fig.2). For the selected parameters we constructed the dataset with the greater number of patients (215) for 10 params. This relation is more acceptable for model construction.



0.152248	70	Diam ACC right	0.038972	37	Malignancies in Anamnesis
0.107989	8	Physical Activity	0.035553	6	Studii Education
0.098438	21	Diabetes Mellitus	0.034444	9	Alimentation
0.085626	69	Number of plaques right	0.033981	36	Peripheral Vascular Disease in Anamnesis
0.076815	18	Atrial Fibrillation	0.033247	7	Occupation
0.050836	19	Accute Myocardial Infarction	0.032551	17	Hypertension
0.050797	67	ACI stenosis right	0.032319	11	Smoking
0.044378	34	CVA in Anamnesis	0.024881	23	Peripheral Vascular Disease
0.043526	71	ACI stenosis left	0.024417	10	Fasting
0.040639	28	Systemic disease	0.02082	24	Valvulopathy
			0.020525	33	Dentures

Figure 2. The selected ranked 21 parameters: ranking coefficient, order number, name

Second approach: The targeted selection of parameters for model creation – different sets of 10 params were selected based on various considerations: either the features considered as common risk factors, or the features used in published prognostic models. Afterwards, the rows which contained missing values for one or more parameters were excluded from the dataset. In particular, when selecting parameters frequently used in different stroke prognostic models, we have obtained the dataset with 168 patients and 10 params (besides the target parameter *Ischemic CVA*): *age, gender, hypertension (HTN), atrial fibrillation (AF), acute myocardial infarction (AMI), diabetes mellitus (DM), peripheral vascular disease (PVD), systolic blood pressure (SBP), creatinine, left ventricular hypertrophy (LVH).*

We tried to draw analogies between params from our dataset and those used in other stroke prognostic models (as in Fig.3). This turned out to be not so simple task, because:

- we had not all the params used in the existing models;
- parameters names are not standardized: in different models the same parameter can have different names, e.g., *hypertension* and *high blood pressure*, *stenocardia* and *angina*;
- one parameter can contain several indicators, as e.g., for parameter *Dyslipidemia* in model M4 (Fig.3) we took 3 params as conditional equivalent in our dataset: *triglyceride, cholesterol, beta-lipoprotein* since Dyslipidemias may be manifested by elevation of the total cholesterol, the "bad" low-density lipoprotein (LDL) cholesterol and the triglyceride concentrations, and a decrease in the "good" high-density lipoprotein (HDL) cholesterol concentration in the blood (https://www.medicinenet.com/script/main/art.asp?articlekey=33979).
- for several parameters in one model we can suggest only one parameter as conditional equivalent. For example, in model M4, two params (*Postprandial Glycaemia* and *A Jeun Glycaemia*) have only one indicator (*Glucose*) as the analogue in our dataset.
- such parameter as *Previous Condition* stores one of three possible diseases that affected the patient in the past and that increase the risk: angina, MI, coronary revascularization. We have information only about one of these deseases – MI.
- Also, such parameter as *Previous Condition* may have another name in other model and, as a compound parameter, may contain different information. Thus, in model M4 Previous Condition is angina, MI, coronary revascularization; in model M1 it is prior cardiovascular disease (coronary heart disease, cardiac failure, or intermittent claudication); in model M2 they have *"history of heart disease" = angina + MI + ECG evidence of past silent MI + coronary bypass surgery (angioplasty) + PVD, defined as claudication in the legs + congestive heart failure.*
- *General health*: there are different approaches for its assessment.
- It is necessary to note, that in the models described in Section 2, there are sophisticated parameters, the values for which are

difficult to obtain from the one hand, and the simple ones, which are not time consuming and do not require complex and costly equipment, from another hand. The last, but not the least are such as *Min. ankle arm ratio* (model M3), *Maximal inflation level* (model M3), *time to walk 15 feet (in sec)* (models M2, M3).

In Fig. 3 there are params considered as significant in models M1–M5. Denotations reflect the relation of the respective parameter with params in our dataset. Information before "/"concerns the presence of this parameter in the model indicated in the column ("+" means that this parameter is present, figure in brackets means that in this model this parameter is compound and consists of the respective number of indicators). Information after "/" concerns the correspondence of this parameter with parameter(s) in our dataset. Sign "=" means that such parameter exists in our dataset, sign "-" means its absence.

Thus, for example, in model M2 for parameter *Previous Conditions* we have: "**+(=6)/ 3**", which means that in this model parameter *Previous Conditions* consists of 6 indicators: *"history of heart disease" = angina + MI + ECG evidence of past silent MI + coronary bypass surgery + PVD + congestive heart failure.* We have only 3 of them: *MI, evidence of past silent MI, PVD.*

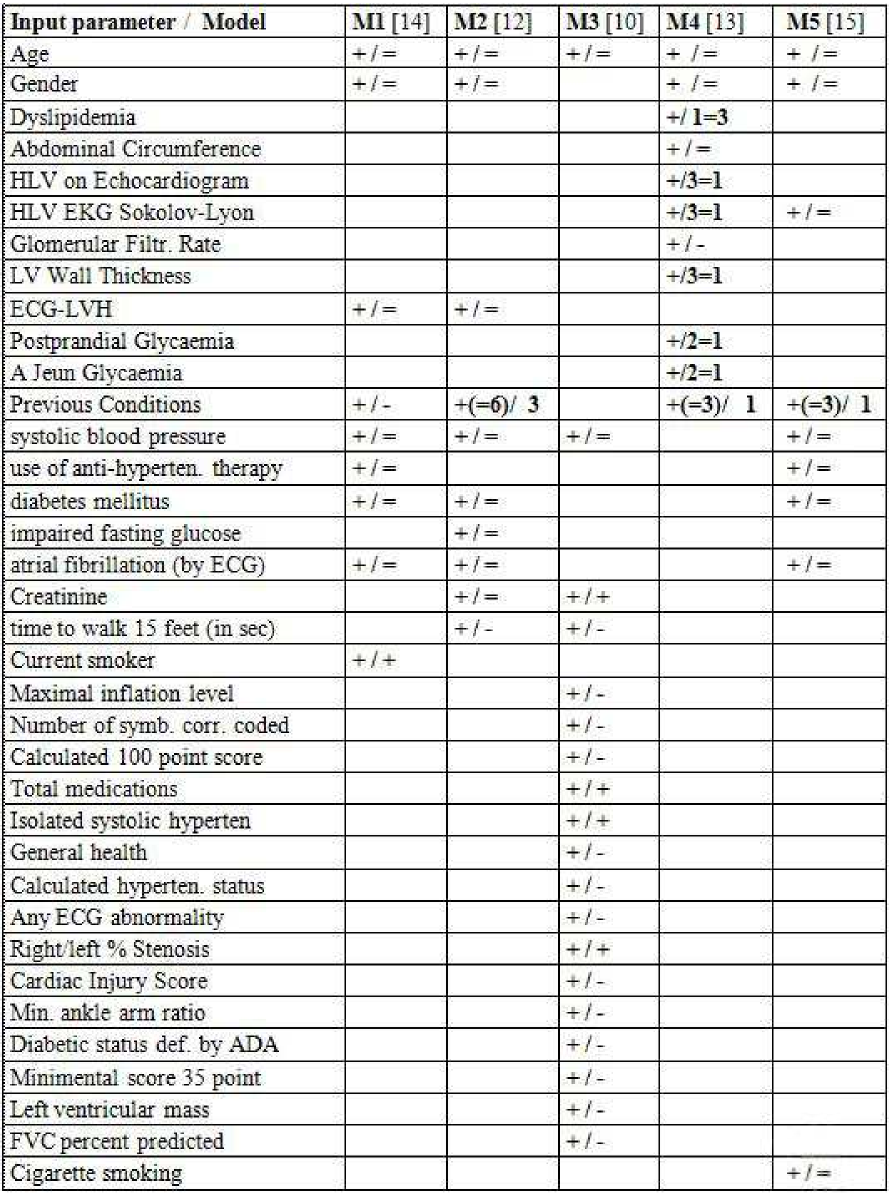
Formula after "/" means that:

- the parameter in our dataset is compound. For example, for *Dyslipidemia* in the model M4: "+/ 1=3" means that in the model M4 there is only one parameter *Dyslipidemia*, and our dataset contains 3 indicators: *triglyceride, cholesterol, beta-lipoprotein;*

- our dataset has only one parameter, to which several parameters of the respective model correspond. For example, in model M4 formula "+/2=1" for parameters *Postprandial Glycaemia* and *A Jeun Glycaemia* means, that these two parameters exist in the model M4, but our dataset has only one parameter *Glucose.*

5 Model construction for 10 parameters

Models were constructed and tested by 8 listed in Fig.4 methods, using 10 parameters obtained as the most significant both for the case when



Input parameter / Model	**M1** [14]	**M2** [12]	**M3** [10]	**M4** [13]	**M5** [15]
Age	+ / =	+ / =	+ / =	+ / =	+ / =
Gender	+ / =	+ / =		+ / =	+ / =
Dyslipidemia				**+/ 1=3**	
Abdominal Circumference				+ / =	
HLV on Echocardiogram				**+/3=1**	
HLV EKG Sokolov-Lyon				**+/3=1**	+ / =
Glomerular Filtr. Rate				+ / -	
LV Wall Thickness				**+/3=1**	
ECG-LVH	+ / =	+ / =			
Postprandial Glycaemia				**+/2=1**	
A Jeun Glycaemia				**+/2=1**	
Previous Conditions	+ / -	**+(=6)/ 3**		**+(=3)/ 1**	**+(=3)/ 1**
systolic blood pressure	+ / =	+ / =	+ / =		+ / =
use of anti-hyperten. therapy	+ / =				+ / =
diabetes mellitus	+ / =	+ / =			+ / =
impaired fasting glucose		+ / =			
atrial fibrillation (by ECG)	+ / =	+ / =			+ / =
Creatinine		+ / =	+ / +		
time to walk 15 feet (in sec)		+ / -	+ / -		
Current smoker	+ / +				
Maximal inflation level			+ / -		
Number of symb. corr. coded			+ / -		
Calculated 100 point score			+ / -		
Total medications			+ / +		
Isolated systolic hyperten			+ / +		
General health			+ / -		
Calculated hyperten. status			+ / -		
Any ECG abnormality			+ / -		
Right/left % Stenosis			+ / +		
Cardiac Injury Score			+ / -		
Min. ankle arm ratio			+ / -		
Diabetic status def. by ADA			+ / -		
Minimental score 35 point			+ / -		
Left ventricular mass			+ / -		
FVC percent predicted			+ / -		
Cigarette smoking					+ / =

Figure 3. Parameters used in stroke prognostic models described in scientific literature

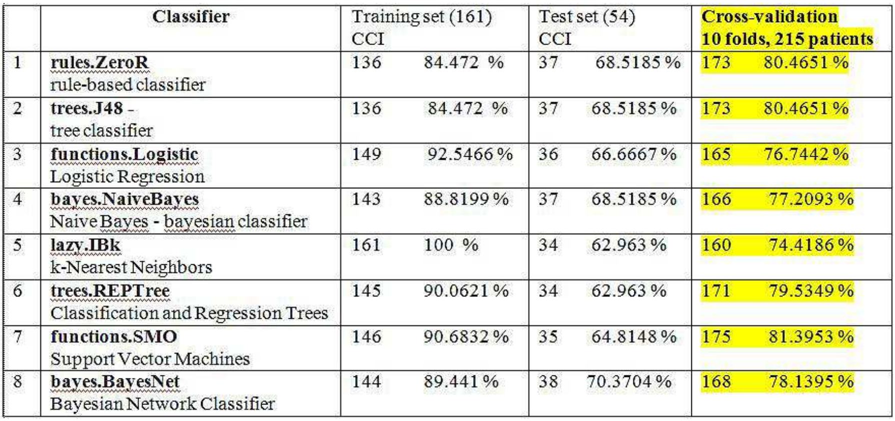
they were ranked, and for the case of targeted selection (according to the common conceptions about risk factors or selecting analogues of parameters usually used in the models).

Two approches were used:

1) initial file with information about 215 patients was split in 2 parts: training set (161 subjects) and testing set (54 subjects).

2) initial file with information about patients used as a whole with cross-validation (CV) in 5, 10 and 20 folds.

In the case of parameters ranking (Fig.1) the first 10 parameters were selected as the most significant (Fig.2). Results for models creation with these parameters and 215 patients, for splitting the dataset in training and testing sets (161 and 54 patients) and for CV (10 folds) are in Fig.4: method SMO gives the best result for CV (CCI = 81%).



	Classifier	Training set (161) CCI	Test set (54) CCI	Cross-validation 10 folds, 215 patients
1	**rules.ZeroR** rule-based classifier	136 84.472 %	37 68.5185 %	173 80.4651 %
2	**trees.J48** - tree classifier	136 84.472 %	37 68.5185 %	173 80.4651 %
3	**functions.Logistic** Logistic Regression	149 92.5466 %	36 66.6667 %	165 76.7442 %
4	**bayes.NaiveBayes** Naive Bayes - bayesian classifier	143 88.8199 %	37 68.5185 %	166 77.2093 %
5	**lazy.IBk** k-Nearest Neighbors	161 100 %	34 62.963 %	160 74.4186 %
6	**trees.REPTree** Classification and Regression Trees	145 90.0621 %	34 62.963 %	171 79.5349 %
7	**functions.SMO** Support Vector Machines	146 90.6832 %	35 64.8148 %	175 81.3953 %
8	**bayes.BayesNet** Bayesian Network Classifier	144 89.441 %	38 70.3704 %	168 78.1395 %

Figure 4. Correctly Classified Instances (CCI), 10 params, 215 patients

The procedure of models creation was reiterated for 10 parameters got by targeted selection (*age, gender, HTN, AF, AMI, DM, PVD, SBP, creatinine, LVH*) and 168 patients, for splitting the dataset in training and testing sets (110 and 58 patients), and for CV (10 folds). The CV approach gives better results, then for splitting in training and testing sets.

6 The work with data

Since we are limited in the number of patients, so at the next step of our work we decided not to refuse the missed data, so long as the system Weka permits to cope with this problem. So, we are able to work with the whole set of patients (233 subjects).

Due to such approach we could carry out additional analysis of data with a view to outliers – parameters values, which differ significantly from other values of the corresponding parameter. Thus, we could find out and correct mistakes in such parameters as *weight, height, body mass index, fibrinogen.* Such parameters as *ALAT, ASAT, glucose, triglyceride, cholesterol,* etc., which have outliers too, produce questions that need to be specified with physicians, and the introduced information to be verified.

Afterwards we repeated the procedure of parameters exclusion according to the following criteria (and keeping the number of patients): missed values >25%; unknown values >25%; dental interventions; *nationality, religion*, and women related information.

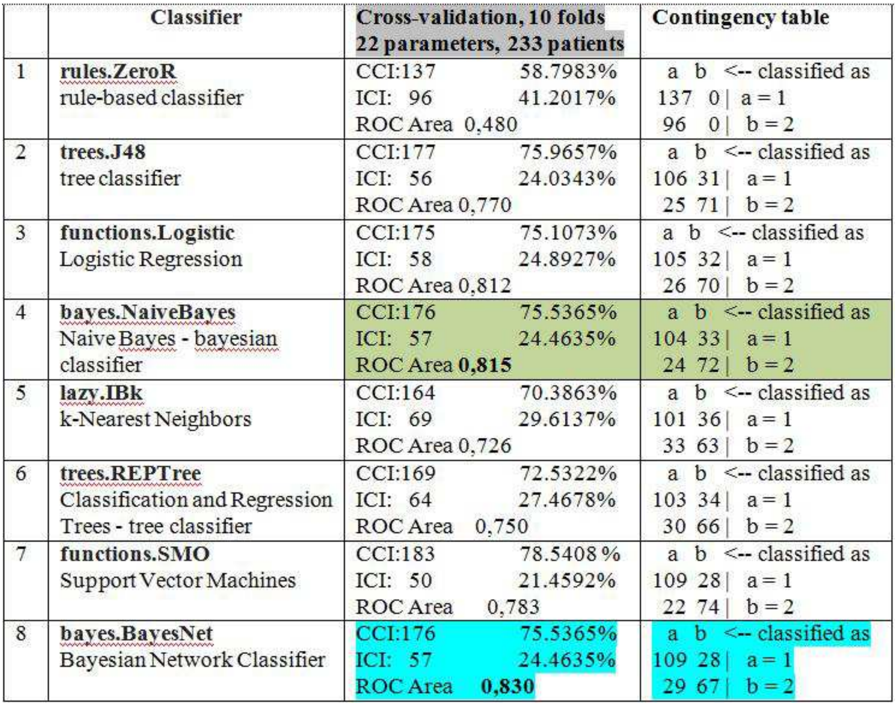
The second group in our dataset is patients interned into the hospital with diagnosis "stroke".

In our dataset patients with stroke mostly have arterial blood pressure and pulse rate within normal range or not pretty high – and this is a consequence of treatment that had already began. So, these parameters were also excluded from examination when model creation.

The process of parameters ranking and selection was repeated. According to the method GainRatioAttributeEval (the rest methods give more or less similar results) we have got 21 significant parameters: *Ischemic CVA, CIM right, Hemorrhagic CVA, CIM left, Age, Arrhythmia, Systemic disease, Diam ACC left, ACI stenosis right, Anticoagulant, Old MI, AF, Number of plaques left, TIA, Gender, HTN, Civil status, AMI, Smoking, LVH, ACI stenosis left.*

We have created models with these parameters, setting parameter "stroke on the date of clinical examination" as target parameter (Fig.5). The following estimations were analyzed: CCI, Incorrectly Classified

Instances (ICI), ROC Area and Contingency table.



	Classifier	Cross-validation, 10 folds 22 parameters, 233 patients	Contingency table
1	**rules.ZeroR** rule-based classifier	CCI:137 58.7983% ICI: 96 41.2017% ROC Area 0,480	a b <-- classified as 137 0 | a = 1 96 0 | b = 2
2	**trees.J48** tree classifier	CCI:177 75.9657% ICI: 56 24.0343% ROC Area 0,770	a b <-- classified as 106 31 | a = 1 25 71 | b = 2
3	**functions.Logistic** Logistic Regression	CCI:175 75.1073% ICI: 58 24.8927% ROC Area 0,812	a b <-- classified as 105 32 | a = 1 26 70 | b = 2
4	**bayes.NaiveBayes** Naive Bayes - bayesian classifier	CCI:176 75.5365% ICI: 57 24.4635% ROC Area **0,815**	a b <-- classified as 104 33 | a = 1 24 72 | b = 2
5	**lazy.IBk** k-Nearest Neighbors	CCI:164 70.3863% ICI: 69 29.6137% ROC Area 0,726	a b <-- classified as 101 36 | a = 1 33 63 | b = 2
6	**trees.REPTree** Classification and Regression Trees - tree classifier	CCI:169 72.5322% ICI: 64 27.4678% ROC Area 0,750	a b <-- classified as 103 34 | a = 1 30 66 | b = 2
7	**functions.SMO** Support Vector Machines	CCI:183 78.5408 % ICI: 50 21.4592% ROC Area 0,783	a b <-- classified as 109 28 | a = 1 22 74 | b = 2
8	**bayes.BayesNet** Bayesian Network Classifier	CCI:176 75.5365% ICI: 57 24.4635% ROC Area **0,830**	a b <-- classified as 109 28 | a = 1 29 67 | b = 2

Figure 5. Results of model creation with 22 parameters (including target parameter) and 233 patients

Methods NaiveBayes and BayesNet gave the best results (Fig.5): high percent of CCI (75%), the greatest ROC Area (0,81 and 0,83 for NaiveBayes and BayesNet respectively). The contingency table gives relatively high CCI in the main diagonal against ICI, i.e. for every class (class **a** – patients with stroke, class **b** – patients without stroke). According to this estimation method SMO gives even better result.

However, 6 parameters from our list of selected 21 are related to the Doppler examinations: *CIM right, CIM left, Diam ACC left, ACI stenosis right, Number of plaques left, ACI stenosis left.* Since this examination, as a rule, is difficult of access and the respective values can be missing at a number of patients, so when creating different prognostic models they are intentionally excluded from the list of examined parameters, or the comparative analysis is carried out: models with Doppler indicators as against models without them. We decided to run such analysis as well, and at the next step of models creation we excluded parameters with Doppler examinations. There remained 59 parameters, including targeted parameter, and 233 patients. Then the procedure of parameters ranking and models creation was reiterated. The selected ranked parameters are: *Ischemic CVA, Hemorrhagic CVA, Age, Arrhythmia, Systemic disease, Anticoagulant, Old MI, AF, TIA, Gender, HTN, Civil status, AMI, Smoking, LVH, Antidiabetic, Physical activity, Antiarrhythmic, Migraine, Occupation, Antihypertensive.* The best results were obtained with method BayesNet – CCI = 72%, ROC Area = 0,79; method SMO gave contingency table which shows the best correlation of CCI and ICI in classes **a** and **b**.

7 Conclusion

Our goal for today is to find significant parameters for our specific dataset. So, we tried various ways to find them (including ranking and targeted selection). One of the common approaches is to differentiate cohorts by gender. Thus, our next step is to divide our dataset by gender and find significant params for each group.

Acknowledgments. Part of this research was carried out within the framework of project 16.00418.80.07A "Mathematical modeling of risk factors and patients clustering for preventive stroke management".

References

[1] I.V. Sidyakina, et al. *Prognostic model of evaluation of lethality and functional recovery after severe and extremely severe stroke*, Medico-social examination and rehabilitation, no. 3 (2012), pp. 49–52. (in Russian)

[2] E.E. Smith, et al. *Risk Score for In-Hospital Ischemic Stroke Mortality Derived and Validated Within the Get With The Guidelines–Stroke Program*, Circulation, vol. 122, no. 15 (2010), pp. 1496–1504.

[3] W. Dai, et al. *A two-level model for the analysis of syndrome of acute ischemic stroke: from diagnostic model to molecular mechanism.* Evid Based Complement Alternat Med, (2013), pp. 1–15.

[4] J.M. Reid, et al. *Predicting functional outcome after stroke by modelling baseline clinical and CT variables*, Age Ageing, vol. 39 (2010), pp. 360–366.

[5] G. Saposnik, et al. *IScore: A Risk Score to Predict Death Early After Hospitalization for an Acute Ischemic Stroke.* Circulation, 123 (2011), pp. 739–749.

[6] I.R. König, et al. *Predicting Long-Term Outcome After Acute Ischemic Stroke A Simple Index Works in Patients From Controlled Clinical Trials.* Stroke, vol. 39, no. 6 (2008), pp. 1821–1826.

[7] S.S.Y. Au-Yeung, and C.W. Hui-chan. *Predicting recovery of dextrous hand function in acute stroke.* Disability and Rehabilitation,vol.31(2009),pp.394–401.

[8] A. Khosla, et al. *An Integrated Machine Learning Approach to Stroke Prediction.* Proc. of the 16th ACM SIGKDD intern. conf. on Knowledge discovery and data mining, pp. 183–192.

[9] C.T. Volinsky, et al. *Bayesian Model Averaging in Proportional Hazard Models: Assessing the Risk of a Stroke.* Applied Statistics, vol.46, no.4(1997), pp.433-448.

[10] Th. Lumley, et al. *A stroke prediction score in the elderly: validation and Web-based application.* Journal of Clinical Epidemiology, vol.55,(2002),pp.129–136.

[11] I. Tănăsoiu, and A. Albu. *A Connectionist Model for Cerebrovascular Accident Risk Prediction.* Proceedings of EHB 2017 - The 6th IEEE International Conference on E-Health and Bioengineering, 2017, pp. 45–48.

[12] P.A. Wolf, et al. *Probability of stroke: a risk profile from the Framingham Study,* Stroke, 1991, vol. 22, no. 3 (1991), pp. 312–318.

[13] J.W. Lee, et al. *The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record*, Computer Methods and Programs in Biomedicine, vol. 153 (2018), pp. 253–257.

[14] Elena Zamsa. *Medical Software User Interfaces, Stroke MD application design,* In: E-Health and Bioengineering Conference. Proceedings of a meeting held 19-21 November 2015, Iasi, Romania. ISBN 9781467375467, vol. II IEEE, Curran Associates, Inc, (2016), pp. 563–567.

[15] S. Cojocaru, et al. *Data preparation in the process of prognostic model STROKE.MD creation*, In Proc. MFOI-2017, (2017), pp. 59–64.

[16] E.Frank,et al. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, 4th ed. Morgan Kaufmann, (2016).

Svetlana Cojocaru[1], Constantin Gaindric[1], Galina Magariu[1], Tatiana Verlan[1]

[1] Institute of Mathematics and Computer Science, Chisinau, Republic of Moldova
Emails: `svetlana.cojocaru@math.md constantin.gaindric@math.md, galina.magariu@math.md, tatiana.verlan@math.md`

On interoperability of research information based on CERIF in the Republic of Moldova

Rodica Cujba, Andrei Rusu

Abstract

The paper is dedicated to the interoperability of research information based on the Common European Research Information Format (CERIF) and its realization in the Republic of Moldova by adapting it to local needs and realities.

Keywords: interoperability, research information, CERIF, CRIS.

1 Introduction

At present, the way the scientific research is organized and conducted, essential changes based on cooperation and new ways of knowledge dissemination using digital technologies and new collaborative tools are imposed. The new approach is driven by the exponential growth of information and the availability of digital technologies, driven by the globalization of the scientific community and by the growing demand from society to find solutions to today's challenges. Research, Development and Innovation (RDI) decision-makers and society as a whole need access to accurate, comprehensive, credible and visible information on scientific resources, activities and results. In these conditions, re-usability and interoperability of data became one of the main problems in organizing and monitoring of research activities, and research results dissemination.

The need for a standard of interoperability of data related to research process comes from the fact that at least a big part of research is done based on public funds and the society would like to know how efficiently these funds are spent. So there is a need for countries, research organisations and decision bodies, implied in exchanging information related to research, to have a common language of reference and the same understanding of notions.

The Common European Research Information Format (CERIF) comes to help to solve this problem.

2 CERIF

CERIF is developed by an international not-for-profit association, euroCRIS, with the goal to promote cooperation within and share knowledge among the research information community and interoperability of Current Research Information Systems (CRIS) through CERIF [1].

CERIF includes the concept (conceptual level), description (logical level) and formalization (Physical Level) about research entities and their relationships.

Basic entities of the CERIF are: project, person and organisation unit. All other entities that appear in CERIF are related to them, for example, result publication, result product, equipment, funding, event, country, etc (see Fig.1 [1]).

Some of the advantages of CERIF are: a CRIS can be implemented using a subset or superset of the full CERIF model; its architecture is neutral; it supports relational, object-oriented or information retrieval data model.

Today CERIF is used as a model for implementation of a standalone CRIS (but interoperation ready), as a model to define the wrapper around a legacy non-CERIF CRIS to allow homogeneous access to heterogeneous systems and as a definition of a data exchange format to create a common data warehouse from several CRISes.

In order to operate with the same terms EuroCRIS developed CERIF Ontology Specification and Semantic Vocabulary. CERIF Ontology Specification provides basic concepts and properties for describing research information as semantic data, and CERIF Semantic Vo-

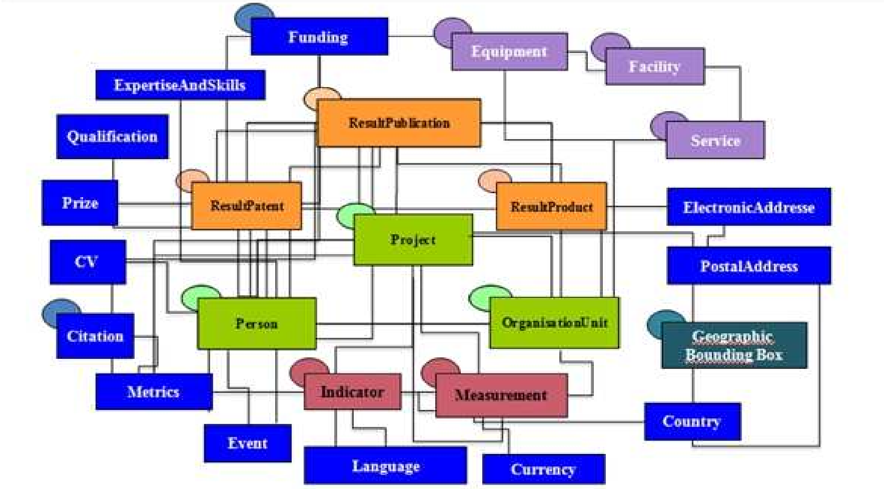


Figure 1. CERIF entities

cabulary provides general relationship and type terms for the research domain.

Today such giants of bibliographic data management systems as Clarivate Analytics (former Thomson Reuters), OpenAIRE, Elsevier developed tools based on CERIF to facilitate data exchange with their software systems.

3 MD-CERIF

There are several information systems in the Republic of Moldova with the goal to help researchers and the society to deal with processes of organizing, conducting, monitoring research activities and dissemination of research results. Some of them are: National Bibliometric Instrument (https://ibn.idsi.md/) [2], EXPERT on-line (https://expert.idsi.md/) [3], Research and Development Indicators of the Republic of Moldova (http://indicator.idsi.md/)[4]) – developed by Information Society Development Institute (ISDI), as well as institutional repositories based on DSpace platform [5] – developed and maintained by many Moldovan universities and some research organisations [6].

In order to achieve interoperability among these and other national and international systems, IDSI proposed a compatible standard to be used in the Republic of Moldova, called MD-CERIF.

MD-CERIF is based on a subset of entities of CERIF data model and the corresponding relations between them (Fig. 2).

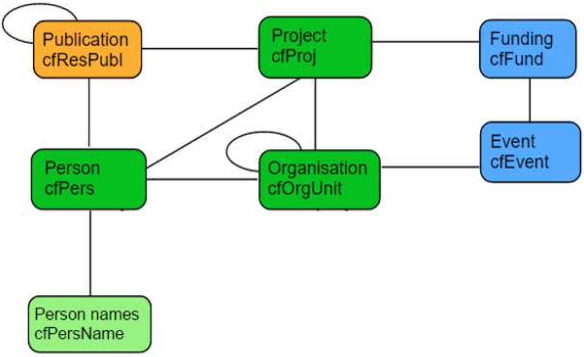


Figure 2. MD-CERIF data model based on a subset of CERIF

At the first stage, it was decided to limit to the following research entities and their relationships: Person, Project, Organisation Unit (base or first level entities), Publication (result entity), Funding, Event (second level entities).

MD-CERIF data model contains description of every entity: attribute, applied vocabulary, cardinality, format and related entity.

Semantic layer is very important for MD-CERIF data model. In order to use the same terms and nomenclatures, and to enhance their comprehension, IDSI adapted CERIF semantic vocabulary (v. 1.3) to the needs and realities of Moldovan RDI system. Thus, 28 new classification terms were included, and translations in Romanian language for all terms in the MD-CERIF semantic vocabulary were added.

A corresponding detailed documentation on MD-CERIF is presented on the web page of the SCIFORM project conducted by IDSI during 2015-2018 [7].

4 Conclusion

In this paper a short description of CERIF standard is presented together with its adaptation (MD-CERIF standard) to the national needs and realities.

MD-CERIF with adapted semantic vocabulary are proposed to be used in the Republic of Moldova in order to contribute to the interoperability of information related to scientific research within academia in Moldova and abroad.

The data model described by MD-CERIF will be permanently updated and improved. The Project team will continue to pursue the CERIF standard as well as some good practices examples such as OpenAIRE to harmonize MD-CERIF.

The proposed standard MD-CERIF and its associated documents (semantic vocabularies, etc) may contribute to the fortification of the e-Infrastructure of RDI sphere in the Republic of Moldova in accordance with the needs specified in [8]

Acknowledgments. The present research for this paper was conducted with the support of SCIFORM research project (15.817.06.13A) and E-IDSM research project (18.50.07.10A/PS).

References

[1] EuroCRIS, *CERIF (Common European Research Information Formatmodificat)*, http://www.eurocris.org/cerif/main-features-cerif.

[2] Information Society Development Institute, *National Bibliometric Instrument*, https://ibn.idsi.md/.

[3] Information Society Development Institute, *EXPERT On-line*, https://expert.idsi.md/.

[4] Information Society Development Institute, *RD Indicators*, http://indicator.idsi.md/.

[5] DuraSpace Community, http://www.dspace.org/.

[6] Nelly Ţurcan, Rodica Cujba. *Open Access Policy to research outputs in the Republic of Moldova. State of the art and perspectives.* In: Conference proceedings "CEE eDem and eGov Days 2017", May 4-5, 2017. Budapest, 2017, pp. 283-293. ISBN: 978-3-903035-14-0.

[7] Information Society Development Institute, *SCIFORM project*, https://idsi.md/sciform.

[8] Igor Cojocaru, Alfreda Roşca, Andrei Rusu, Mihail Guzun. *Public Research and Innovation Infrastructure of the Republic of Moldova: Challenges and Oportunities.* In: Conference proceedings "CEE eDem and eGov Days 2018", May 3-4, 2018. Budapest, 2018, pp. 421-430, DOI:10.24989/ocg.v331.35, ISBN 978-3-903035-14-0.

Rodica Cujba[1], Andrei Rusu[2,3]

[1]Information Society Development Institute
Email: `rodica.cujba@idsi.md`

[2]Information Society Development Institute
Email: `andrei.rusu@idsi.md`

[3]Ovidius University of Constanta
Email: `agrusu@univ-ovidius.ro`

Towards an Algebraic Set Theory for the Formalization of Data Structures

Ioachim Drugus

Abstract

A set theory intended to serve as a foundational framework for data structures used in computer science is formulated as an axiomatic theory in algebraic presentation. This theory is weaker than Bourbaki's set theory posited in the foundation of Bourbaki's "theory of structures" and is weaker than ZF set theory. It is presumably the weakest mainframe set theory.

Keywords: data structure, set theory, atomification, association, aggregation, well-founded set.

1 Introduction

In computer science, "data structures" and "algorithms" are treated as complementary subjects (see [1]). However, in mathematics, there is a "theory of algorithms" which plays a foundational role for the discipline of programming, but there is no widely accepted "theory of structures" to play a similar foundational role for the discipline of architecturing data structures. An attempt to formulate a "theory of structures" describing the structures used in mathematics, i.e. the "mathematical structures", was made by Bourbaki in the first book of well-known series [2], immediately after the presentation of what is currently known as "Bourbaki's set theory". However, this attempt was widely criticized, and an example of such criticisms can be found in the paper [3], which is often used as an informal account of the results of the difficult-to-read Bourbaki's paper [2].

The belief of the author of current paper is that the criticisms of Bourbaki's approach developed in [2] refer to the form of presentation, rather than to results. These are worth of thorough consideration, given that, apparently, no other successful attempts in direction of formalization of mathematical structures have been made before or thereafter. In Bourbaki's set theory, the axiom of foundation in any of its known forms is not deducible. Also, this theory uses Hilbert's "tau-operator", the effect of which is similar to that of an "axiom of global choice", a strong statement that cannot be even formulated in the language of ZF; it can be formulated only in a proper extension of ZF's language. These and other features of Bourbaki's set theory support the thesis that this is not one of the mainstream set theories, i.e. theories obtained by weakenning ZF set theory or by adding new axioms to it.

In this paper, a set theory, referenced as "aggregate theory", intended to serve as a framework for the formalization of the notion of structure, is introduced. The development of this set theory began in [4], continued in [5], and acheived in this paper a state of completeness. This is believed to be a minimal algebraic set theory. Whether or not this theory can serve as a pivot, around which other theories can be developed – theories which could be referenced as "mainstream algebraic theories", remains to be seen in the future.

The aggregate theory employs for algebraisation purposes universal algebras with finitary operations in contrast with the widely accepted "algebraic set theory" of [6], which employs for algebraisation ZF-algebras – algebraic structures with infinitary operations.

No proofs of the statements valid in aggregate theory or of the statements about this theory are presented in this paper because of the space limitations.

2 Informally about data structures

The idea to posit a set theory in the foundation of "structure theory" goes back to Bourbaki, as mentioned in Introduction. The approach used in this paper is to formulate on the role of foundation of the structure theory an *algebraic* set theory, rather than use for this purpose the Bourbaki's set theory or the ZF set theory, theories which are not algebraic. This is motivated by the following intuitions:

- "Structures" are "constructed things" (notice the common root of "structure" and "construct"), i.e. structures are objects built by repetitive application of operations – operations, which can be described as "structure composition operations".

- The structure composition operations can be taken to be primitive operations of an algebraic set theory posited in the basis of the structure theory.

- Structures can be specified as values of the terms in the signature of the algebraic theory foundational for the structure theory – values, obtained in result of interpreting the term's variables with values lying within the universe of discourse, in particular, atoms.

A theory of structures can be posited upon a set theory, in which the notion of ordered pair is defined through the notion of set, like when one uses Kuratowski's definition of ordered pair. However, such a foundation for the structure theory is not rich enough to explicate the notion of order, which is commonly perceived as of a nature different from the nature of sets. Moreover, the ordered pairs are perceived exactly as a special type of structures. Therefore, the algebraic set theory introduced below will have the ordered pair operator in its signature, and this is similar to Bourbaki's set theory. In such a "set theory", an ordered pair is an object of a type different from the type of sets, and the term "set theory" does not sound quite appropriate. Also, the term "set theory with atoms" calls for a generalization of the notions "atom" and "set".

Atoms and ordered pairs must be naturally included into the universe of discourse of a theory foundational for the structures, and a general term for both would be very useful. The term "aggregate" was chosen here for this purpose because, alongside sets, it sounds appropriate also for the atoms and ordered pairs. Also, the word “aggregate” sounds appropriate to serve as a shorter term for “data structure” and it will be used here also on this role. Notice, the word “aggregate” was used for the German term “menge” (“set”) in the first traslation of the article where Georg Cantor introduced the notion of set. The term “aggregate theory” (or “theory of aggregates”) will be preferred to the term “set theory” for the theory introduced in this paper, because this term does not exclude atoms from discourse and because the ordered pairs are explicitly specified in this theory’s signature.

An intuitive approach to data structures was introduced in [7], an approach according to which, any structure is built of “atomary structures” or “atoms” by repetitive application of two operations – the “*aggregation* operation”, used to form a set from its elements, and “*association* operation” used to form an ordered pair of two objects. This approach was further developed in [8], where a third operation called “*atomification* operation” was introduced, an operation which applied to an object results in same object, viewed “as a whole”. One can say, that this operation is changing the view upon an object rather than the object itself. This intuitive approach to aggregates, referenced as “A3 approach”, where the symbol “A3” stands for “Atomification-Aggregation-Association”, can be used in modeling as an alternative to the Entity-Relationship (or “ER”) approach.

To be used in practice, a well-tuned terminology for the A3 approach is needed, and next, the appropriateness of choice of term “aggregation” will be checked. Such a check is needed, because the word “aggregation” with the same root as the word “aggregate” sounds more appropriate for an operation of forming an aggregate than for an operation of forming a set. But the use of the term “aggregate” in two senses, for a set and for a more complex structure, is justified by the history of set theory. Thus, the term “aggregation” needs to be used also in two senses – there is an operation of forming a set called “aggregation (of a set)” within the context of the A3 approach, and there is an operation of forming an aggregate, which “requires” the name “aggregation" (even though the expression "aggregation of an aggregate" is obviously ill-formed).

The difference between the ER approach and the A3 approach is that the former one prescribes to focus on relationships, and the latter one prescribes to focus on operations, and consider three operations as fundamental for the creation of data structures (aggregates). Current paper is intended to algebraically explicate, even though partially, the mechanism of the A3 approach.

The knowledge of previous papers [4] and [5] is not presupposed and the main results presented in those papers will be reviewed here, even though with minor changes intended to improve the presentation and ease of use of their results (e.g. the sign "$'$" is frequently used to form variables' names, and this sign used for an operation in [5] is replaced here by the degree symbol "$^\circ$").

3 An algebraic aggregate theory

This is a theory of universal algebras, "aggregate algebras", viewed as universes of discource of set theory. The signature of the aggregate theory consists of four operations' symbols ("operators") listed below (with their standard set-interpretation, "treatment", indicated between the parentheses like here):

- A constant 0 (treated as the empty set);
- A unary operator "$^\circ$" (where "x°" is treated as the singleton "$\{x\}$"), of an operation called "successor operation";
- The operators of two binary operations, "+" (treated as union of two sets) and "–" (treated as set difference and also called "subtraction");
- A binary operator, which is a pair of parentheses enclosing the list of operands (so, that "(x, y)" is treated as an ordered pair).

The axioms of the aggregate theory are as follows:

$$a - a = 0, \tag{1}$$

$$a + (b + c) = (a + b) + c, \tag{2}$$

$$a + b = b + a, \tag{3}$$

$$a + a = a, \tag{4}$$

$$(a + b) - c = (a - c) + (b - c), \tag{5}$$

$$a - (b + c) = (a - b) - c, \tag{6}$$

$$a + (b - a) = a + b, \tag{7}$$

$$a + (a - b) = a, \tag{8}$$

$$(a - b) - c = (a - c) - (b - c), \tag{9}$$

$$a - (b - c) = (a - b) + (a - (a - c)), \tag{10}$$

$$a^\circ = b^\circ \to a = b. \tag{11}$$

$$(a,\ b) = (a',\ b') \to a = a'\ \&\ b = b' \tag{12}$$

The aspects of independence of some axioms from others, or of choosing a minimal number of axioms were not thoroughly considered here.

3.1 On the meaning of symbols used in aggregate theory

One can easily check the intuition behind the axioms of the aggregate theory, as well as their validity, by treating the symbols in the signature according to the set-theoretic interpretation (indicated above between parentheses). With this interpretation, the "aggregate theory" is a *minimal* set theory among the mainstream set theories, because all its axioms are validated in any of them. Really, in any of mainstream set theories, the binary operations of union and set difference are defined and satisfy the properties (1-10), the singleton operation satisfies the axiom (11), and the notion of ordered pair is defined in such a manner that it has the "characteristic property of ordered pair" which is exactly the axiom (12).

One of the reasons why symbols different from set-theoretic ones were preferred for the signature of aggregate theory, is that there exist interesting non-standard interpretations of this theory where the use of set theoretic symbols could create confusions. For example, notice, that for a denoting a “Quine’s atom” (a kind of singleton in Quine’s New Foundations), the equation $a^\circ = a$ holds, but in set theoretic notation this equation looks like this: $\{a\} = a$, a correlation which could raise confusions. Another reason for using generic algebraic symbols “+” and “–” rather than set theoretic symbols, is that this theory can also describe generalizations of Boolean algebras with operators, where logical interpretation also makes sense.

The symbols of aggregate theory are convenient enough to combine set theory and logic into one theory. The need for such a combination can be discovered through practical examples like this one: the set-theoretic symmetric difference operation is an exact counterpart of the logical Sheffer’s operation.

3.2 Basic statements and definitions of aggregate theory

Some basic statements and definitions are given below to elucidate the meaning of symbols. Most of these statements are easy to prove, but taken together, they make up an apparatus convenient for proofs of more difficult theorems. Some descriptions are given after formulas to help memorize them and link with similar formulas from other domains. Correlations of some statements with the axioms of other theories are suggested in these descriptions. Several acronyms and abbreviations are used, like these ones: “wrt” stands for “with respect to”, “rel.” for “relative”, “gen.” for “generalized”, “def.” for “definition”, “subtr.” for “subtraction”, and several other ones.).

$$a - a = b - b, \qquad \text{unicity of 0} \tag{13}$$

$$a + 0 = a \qquad \text{neutrality of 0 wrt add.} \tag{14}$$

$$a - 0 = a \qquad \text{neutrality of 0 wrt subtr.} \tag{15}$$

$$(a - b) + b = a + b \qquad \text{subtr. is rel. complement} \tag{16}$$

$$a - (b - a) = a \qquad \text{contraction} \tag{17}$$

$$(a-b)-c=(a-c)-b \qquad \text{exchange} \tag{18}$$

$$a \leq b \leftrightharpoons a-b=0 \qquad \text{def. of partial order} \tag{19}$$

$$a=(a-b)+(a-(a-b)) \qquad \text{gen. Huntington axiom} \tag{20}$$

$$a=a-(b-a) \qquad \text{gen. Robbins axiom} \tag{21}$$

$$a \bigtriangleup b \rightleftharpoons (a-b)+(b-a) \qquad \text{def. of symm. diff.} \tag{22}$$

$$a \cdot b \rightleftharpoons (a+b)-(a \bigtriangleup b) \qquad \text{def. of intersection} \tag{23}$$

$$a \cdot b=a-(a-b) \qquad \text{alternative def. of int.} \tag{24}$$

$$a \cdot b=b-(b-a) \qquad \text{alternative def. of int.} \tag{25}$$

$$a-(a-b)=b-(b-a) \qquad \text{quasi-commutativity} \tag{26}$$

$$a-(b \cdot c)=(a-b)+(a-c) \qquad \text{gen. De Morgan law} \tag{27}$$

$$a-(b+c)=(a-b) \cdot (a-c) \qquad \text{gen. De Morgan law} \tag{28}$$

4 On aggregate algebras

Aggregate theory is a theory of universal algebras referenced here as "aggregate algebras". These algebraic structures, the elements of which are called "aggregates", are in need of an algebraic treatment, which would uncover their algebraic properties. There are two subjects of research, which are interesting for some classes of universal algebras of the same type:

A. Identifying the reducts of such algebras, the algebras obtained by omitting some operations in order to bring into focus the properties of the remaining operations;

B. Describing the structure of such algebras in terms of direct or sub-direct products, subalgebras, homomorphic images etc.

In this section, these subjects are discussed in relation with the aggregate algebra. The reducts of the aggregate algebra identified here are the following ones:

1. The algebra obtained by omitting the operations of successor and of ordered pair which validates the axioms (1 - 10), an algebra which in [4] is called "extent algebra".

2. The algebra obtained by omitting the operation of ordered pair,

which validates the axioms (1 - 11), an algebra which in [5] is called "metrologic algebra".

3. The algebra with the operator of ordered pair and no other operators in its signature, an algebra which validates the axiom (12). Such an algebra does not seem to have obtained treatment in literature, and here it will, temporarily or not, be referenced as the "order algebra" – a term seemingly missing in literature.

5 On extent algebra

The following passage from [11] is highly relevant to the manner of algebraization proposed in this paper:

The history of Boolean algebra and much of its usefulness is motivated by the attempt to find an abstract characterization of the algebra of sets. The results are always, to a certain extent, unsatisfactory, since Boolean algebras contain a least and greatest element and are self-dual, whereas set theory as a whole admits no largest set and complements can be only defined with respect to a given set.

In the paper just mentioned, a useful generalization of Boolean algebra called "semi-Boolean algebra" was proposed in order to help overcome the difficulties described above. However, such algebras are not closed under joins, and this makes them insufficiently suitable for modeling the universe of set theory – a class closed under arbitrary unions of sets.

More suitable sound to be the generalized Boolean algebras (GBAs), which can be described as Boolean algebras with an optional "unit" or "top element". Because the unit may be missing, the elements of these algebras may not have complements, and the logic modeled by GBAs is a logic without negation. On the other hand, the GBAs turned out to be suitable for algebraization of most mainstream set theories. In this paper, a form of GBAs, algebras with a reduced signature called "extent algebras", are preferred to GBAs, because these (the extent algebras) allow for generalizations like set theories based on intuitionistic or other non-classical logics.

Since there are also other generalizations of Boolean algebra (e.g. there are pseudo-Boolean algebras), in a reference to this particular generalization, it's appropriate to mention the name of the author like this: "Stone's GBA" (one cannot use for this purpose the term "Stone algebra", because this term is used for algebras of a different type).

5.1 On extent algebras as a form of Stone's generalized Boolean algebra

The Boolean algebra is traditionally defined as a universal algebra with a signature consisting of three operations (union, intersection, complement). In [9], Stone treated Boolean algebra as an idempotent ring (a ring consisting only of idempotent elements) with a mandatory unit, and introduced the GBA as an idempotent ring with an optional unit. Stone described as "double-composition systems" the representations of Boolean algebra and its generalization as special rings. To distinguish between the two forms, the former is sometimes referenced as "Boolean lattices" and the latter (i.e. the "double-composition systems") as "Boolean rings".

Stone described the Boolean rings as "double-composition systems" according to the fact that a ring is an algebraic structure with two binary fundamental operations, addition and multiplication, unlike a group, which is an algebraic structure with one fundamental operation and which is described as a "single-composition system'.

In [4], the question whether there exists a "single-composition system" form of the Stone's GBA, was answered in the positive – such a form exists and this is a universal algebra with a signature, which consists of the join and relative complement operations, and which validates the axioms (1-10). The single-composition form can be referenced as "Boolean monoid" to distinguish it from "Boolean ring". The Boolean monoid form places the Stone's GBAs into another domain of research, that of group theory and group generalizations, and this form "deserves" a name – this was chosen to be "extent algebra".

The extent algebra is a third form of Stone's GBA – a form which, in the terminology used by Stone, has only one “composition operation”, the operation of union. The extent algebra has in signature the symbol for relative complement, but this should be considered “inverse” to “composition”. The extent algebra is a form of Stone’s GBA, where the operation of intersection is missing in signature – it can be defined through the operations in signature, but for the intended uses of this algebra, intersection is irrelevant.

The issue of reducing the number of operations in the signature of Boolean algebra was addressed in the 19th century by Huntington, who managed to reduce the signature to only the disjunction and negation. In Boolean algebra the conjunction can be defined through disjunction and negation due to De Morgan’s laws. But in Stone’s GBA, De Morgan’s laws do not make sense because the complement (the counterpart of negation) is missing and, thus, the intersection (the counterpart of conjunction) cannot be defined.

Notice, that for Stone’s GBA, the set theoretic terminology sounds more appropriate than the logical terminology, and this is in contrast with Boolean algebra, for which the logical terminology is more appropriate (this is the “algebra of logic”).

In extent algebra, the intersection can be defined through union and set difference in three different manners, according to the formula (23) or the formulas (24) and (25). In this paper, the definition according to formula (23) is preferred for the reasons immediately explained. Each of (24) and (25) can serve as a definition of intersection through “non-symmetric” expressions and this is due to “quasi-commutativity” property (26). However, there are generalizations which do not validate (26) and in such generalizations, (24) and (25) define different operations. Hence the definition (23) is more appropriate.

5.1.1 The intuition behind extent algebra

The intuition which the author of this paper puts behind Stone’s GBA is that this is the algebra of *measurable quantities*: what is measured “*extends*” in space or time or, to put it in other words, this is the domain of definition of the function of measure (see also [4]). This intuition behind Stone's GBA is supported by Stone's choice of the examples of such algebras which, in his own words, are "all Lebesgue or Borel measurable sets of finite measure in n-space".

The word "*extent*" used in the term "extent algebra" is intended to reflect the intuition described above: what is measured "*extends*" in space and time; whence the choice of the word "extent". Somebody may prefer the term "extension" to the term "extent", and accordingly, use the term "extension algebra", but the word "extension" is more likely to call after itself connective "of", which is not intended here.

In definitions of measure, the measure function is presupposed to be defined over subsets of a set and to satisfy the "countable additivity" condition formulated in terms of "union of disjoint sets". The notion of disjointness of two sets is naturally treated in terms of set difference (e.g. like this: the sets a and b are disjoint iff $a-b=a$ holds; strangely or not, the symmetric condition $b-a=b$ then automatically holds). Thus, the intersection is irrelevant to the notion of measure and the extent algebra is most suitable form among Stone's GBAs on the role of domain of definition for the function of measure.

According to the tradition of 20th century to express all notions in terms of set theory, the notion of measure is regularly defined as a function over sets. The approach to treat all objects as sets of "points" is often characterized as "pointy" – hence modern measure theory, topology, and analytic geometry are sometimes referenced as "pointy measure theory", "pointy topology", and "pointy geometry". This treatment is rather artificial, since in practice what one measures are the material bodies, intervals of time, etc. – things which are not treated as necessarily consisting of "points". The approach which gives up treating any object as a set of points is often characterized as "point-free" (or "point-less"). Among the first point-free approaches to measure is the paper [10], where the domain of definition of the measure function is considered to be a Boolean algebra. In this paper, the domain of definition of a measure function is regarded as an extent algebra, and this probably completes the expansion of point-free approach intended to cover the measure theory.

5.1.2 Extent algebra as a semi-Boolean algebra

In [11], the semi-Boolean algebras are introduced as algebraic structures of two dual to each other kinds:
– lower semilattices, the principal *ideals* of which are Boolean algebras, structures called *subtraction algebras*;
– upper semilattices, the principal *filters* of which are Boolean algebras, structures called *implication algebras*.

Because the two kinds of algebra are dual to one another, it is enough to give the axioms of the algebras of one kind. The axioms of subtraction algebra are given below:

$$a - (b - a) = a, \tag{29}$$

$$a - (a - b) = b - (b - a), \tag{30}$$

$$(a - b) - c = (a - c) - b. \tag{31}$$

These axioms are valid in the extent algebra as one can see from (17), (26) and (18) above. Thus, the extent algebras form a subclass of the class of subtraction algebras (for precision sake, one should say that the *reducts* of extent algebras, limited only to the operation of subtraction are subtraction algebras).

In [11] it is shown that there are subtraction algebras which do not have the join for some pairs of elements. On the other hand, all extent algebras are lattices. Thus, the extent algebras form a *proper* subclass of substraction algebras.

The fact that each extent algebra is also a subtraction algebra offers the first elucidation of the extent algebras' "structure": the extent algebras are lattices, the principal ideals of which are Boolean algebras.

5.2 On correlation of extent algebras and Stone's generalized Boolean algebras

The correlation between the two forms of GBA is elucidated by the following two theorems from [4]:

Theorem 1. For an extent algebra A, the algebra with A's support and a signature consisting of the constant "0", the operator "$\triangle$" defined in (22), and the operator "$\cdot$" defined in (23) is a Boolean ring (which is not necessarily unitary).

Theorem 2. The axioms of extent algebra are valid in a GBA, where $a-b$ is defined as the complement of "$a \cdot b$" in the principle ideal generated by a.

In relation with Theorem 2, notice that a principle ideal in an extent algebra is a Boolean algebra as mentioned in the previous section.

6 On metrologic algebras

The metrologic algebras are intended to serve for the algebraization of set theories without a primitive ordered pair in signature, i.e. those set theories where the ordered pair is defined. However, these algebras are also suitable for algebraization of set theories with Quine atoms.

The intuition behind the manner of axiomatization used in this paper can be easier described for a finite set theory with atoms like this: any finite set is the value of a term in the signature of metrologic algebra for an interpretation of variables as atoms.

The term "metrologic" reflects the intuitive meaning of these algebras as "algebras of quantity" ([5]). There are two kinds of quantity – "measurable" and "countable". Whereas the measurable quantities, referenced here as "extents", are explicated as "extent algebra", the countable quantities (referenced in [5] as "counts"), form mono-unary algebras. Metrologic algebras combine both kinds of quantity.

6.1 Metrologic algebras versus ZF-algebras

The Zermelo-Fraenkel algebras ("ZF-algebras") were introduced to serve as models of ZF set theory in the widely accepted algebraization of set theory [6]. To put it short, a ZF-algebra is a complete sup-lattice, equipped with an additional unary operation called "successor operation".

For a ZF-algebra A, the following two statements are true:

(a) A is a model of ZF set theory, iff A is a free ZF algebra;

(b) If A is a free algebra, then A is a "large algebra" – i.e. its support is necessarily a proper class.

Given the free ZF-algebra V, one can "recover" the membership relation between sets from the ZF-algebra structure by setting

$$a \in b \leftrightarrow a \leq b,$$

where "$\leq$" is the relation of partial order in the sup-lattice.

One cannot treat a complete lattice as a universal algebra in the customary treatment of this notion. Namely, a universal algebra is defined as an algebraic structure with a certain signature, which is a set of symbols of n-ary operations, where n is a non-negative integer – one says that the signature of universal algebras has symbols of only "finitary operations". The operation of supremum is an infinitary operation and, thus, a ZF-algebra is not a universal algebra. On the other hand, the metrologic algebras are universal algebras by definition, and this makes them more convenient objects than ZF-algebras.

7 On the order algebra

Since the prefix denotation of a binary operation looks like "o(x, y)", one can treat the denotation of an ordered pair "(x, y)" as denoting the result of application of an operation – an operation without a symbol denoting it, an operation with an "empty operator". In this paper, this operation is treated as an explication of the "association operation" used in the A3 approach to data ([7], [8]).

The algebra with empty operator as sole operation symbol in signature, and with the statement (12) as its sole axiom, is called here "order algebra". The statement (12) can be equivalently represented as a conjunction of two statements indicated below:

$$(a, b) = (a', b) \to a = a', \tag{32}$$

$$(a, b) = (a, b') \to b = b'. \tag{33}$$

These statements are quasi-identities and, thus, the class of the order algebras is a quasi-variety – a fact, which has many consequences. In particular, this class is closed under taking subalgebras, direct products, ultra-products, and this brings order algebra into the scope of regular algebraic research.

8 Conclusions

From the results presented in this paper, one can draw the conclusions:

1. The aggregate theory offers a partial explication of the mechanism of A3 approach, and namely:

$\ast$ The atomification operation is partially explicated by successor operation; it explains the notion of Quite atom as an object a which satisfies the identity $a^\circ = a$, but it does not explain what is a urelement;

$\ast$ The aggregation operation which forms a set from its elements is partially explained by the union operation for a set which consists of Quine atoms only, since these are also singletons;

$\ast$ The association operation is well explicated by the operation of forming the ordered pairs.

2. The aggregate theory describes an algebraic quasi-variety.

3. The metrologic algebras form an algebraic quasi-variety, since the axiom (11) is a quasi-identity and the axioms (1-10) are identities.

4. The extent algebras form an algebraic variety.

References

[1] A. Aho, J. Hopcroft, J. Ullman. *Data Structures and Algorithms*, Addison-Wesley, 1983.

[2] N. Bourbaki. *Elements of Mathematics: Theory of sets.* Hermann, Paris, 1968.

[3] L. Cory. *Nicolas Bourbaki and the Concept of Mathematical Structure.* Synthese, vol. 92, No. 3 (1992), pp. 315–348.

[4] I. Drugus. *Generalized Boolean Algebras as Single Composition Systems for Measure Theory.* Proceedings of the 4th Conference of Mathematical Society of Moldova (CMSM-2017), June 28 – July 2, 2017, Chisinau, Moldova, (2017), pp. 75–78.

[5] I. Drugus. *Towards an Algebraic Explication of Quantity.* Proc. Conf. on Mathematical Foundations of Informatics MFOI2017, November 9-11, 2017, Chisinau, Moldova, (2017), pp. 65–70.

[6] A. Joyal and I. Moerdijk. *Algebraic Set Theory.* Cambridge University Press, Cambridge 1995.

[7] I. Drugus. *A Wholebrain approach to the Web.* Proc. of Web Intelligence and Intelligent Agent Technology Conference, Silicon Valley, IEEE, (2007), pp. 68–71.

[8] I. Drugus. Universics: a Common Formalization Framework for Brain Informatics and Semantic Web. In: Web Intelligence and Intelligent Agents. InTech Publishers, Vucovar (2010), pp. 55–78.

[9] M. H. Stone. *Postulates for Boolean Algebras and Generalized Boolean Algebras.* American Journal of Mathematics, vol. 57, no 4 (1935), pp. 703–732.

[10] A. Horn, A. Tarski. *Measures in Boolean algebras*, Trans. Amer. Math. Soc., vol. 64 (1948), pp. 467–497.

[11] J. C. Abbott. *Semi-Boolean Algebras.* Matematicki Vesnik, vol. 4 (19), (1967), pp. 177–198.

Ioachim Drugus[1]

[1]Institute of Mathematics and Computer Science
Email: ioachim.drugus@math.md

Proceedings of the Conference on Mathematical Foundations of Informatics
MFOI'2018, July 2-6, 2018, Chisinau, Republic of Moldova

A Concept for a Decision Support Framework for the Management of Complex Mass Casualty Situations at Distribution Points

Constantin Gaindric, Svetlana Cojocaru,
Stefan Pickl, Sorin Nistor, Iulian Secrieru,
Olga Popcova, Doina Bein, Diana Cimpoesu

Abstract

Very often disasters result in mass casualty situations, which trigger a complexity of decisions to be made at collection points and advanced medical posts by local health services. The decision-making process becomes complicated because of significant bleeding into the peritoneal, pleural, or pericardial spaces may occur without visible warning signs.

We propose the design of an innovative decision support framework for the management of mass casualty situations at collection points via an artificial intelligence based multilayered approach, aimed to support decision-makers, dealing in a disaster area with a considerable number of casualties and have limited resources (ambulances, available nearby medical centers, and personnel).

Keywords: decision support, distribution management, mass casualty situations, on-site triage, medical ultrasound, information technologies, reachback.

1. Introduction - Mass Casualty Situations as Security Problem

Our world continues to experience man-made, technological and natural disasters at a faster pace than before. When a disaster of any type happens, it has an immediate, *negative impact* on countries. Acts of terrorism, fire, industrial accidents, earthquakes, landslides, floods, hurricanes, tsunami, public disorder, and communication failures – all of these events suddenly disrupt the natural flow of everyday life and cause widespread damage to human lives, economy and environment.

Mass casualty situations, as security problem, place medical services under added pressure [1]. In a disaster area, an enormous number of casualties, sustaining life injuries, can die before hospitalization, if healthcare services do not provide a full qualitative rapid aid. As a result of disasters about 2 million people die annually in the world, more than 200 million suffer trauma of diverse severity, consequently about 10 million people remain disabled. According to some studies, the number of deaths after an earthquake would be reduced by 70%, if casualties are provided medical care timely within an hour after catastrophe, by 40%, if assistance is given within three hours, and by 10% if assistance is given within six hours [2].

So, the triage and evacuation of casualties are the most important elements in disaster management chain, especially in mass casualty situations.

2. Problem Description - Complex Desaster Management Chain

In the case of a disaster or an emergency situation with a large number of casualties, many people simultaneously require urgent medical assistance and evacuation from the impact zone in a short period of time. Inevitably, these exceed the local available medical capabilities and resources.

As a rule, at a safe distance away from the immediate threat, *casualty collection points* are established [3]. Casualties are brought to these locations and rapidly triaged.

Medical Triage

Medical triage [4] is a complex process of identification and differentiation of casualties in homogeneous groups according to the severity and nature of injuries, and the degree of medical assistance.

It determines sequence, mode and evacuation destination depending on available medical capabilities and resources, as well as specific circumstances imposed by the impact.

The basic aim of medical triage is to ensure the provision of medical assistance in minimum time and to the largest number (ideally – to all) of casualties of the disaster. The triage officer's strategy is to rapidly sort casualties into several priority groups, taking into account the severity of each case. The commonly accepted classification [5] is the following:

Multilayered Approach

Priority 1 (Red) – *Absolute emergency*. Casualties with serious and very serious injuries, illnesses, intoxication or contamination, compromising vital functions, who require immediate stabilization measures, as well as priority evacuation in assisted medical transport conditions.

Priority 2 (Yellow) – *Relative emergency*. Casualties with serious or moderate injuries, illnesses, intoxication or contamination, with retained vital functions, but with the risk of developing life-threatening complications immediately ahead. They require urgent medical assistance, but not immediate one.

Priority 3 (Green) – *Minimal emergency*. Casualties with minor injuries, illnesses, intoxication or contamination, no life-threatening, which can be treated later, usually in outpatient conditions. They can be evacuated in non-specialized transport or independently.

Triage begins immediately at the casualty collection points and is performed using visible vital signs, such as respiratory rate, pulse, blood pressure, state of consciousness, etc. It should be performed rapidly, as soon as possible, to save as many lives and in a very short period of time, because the number of casualties usually considerably exceeds the capacities of the medical personnel.

Advanced medical posts (AMP) can be set up along the evacuation routes, where the triage decisions made at the previous stage are implemented and further specified [6].

Evacuation - Operations Research (OR) based transportation to a medical center

At the AMP, casualties, whose condition can worsen by potential complications, receive immediate life-saving measures aimed at stabilizing them and preparing them for further evacuation (transportation to a medical center).

Being loaded into appropriate available ambulances or other means of currently available transport (passenger cars, vans, buses, trucks, etc.), casualties are evacuated to the closest treatment facilities (usually hospitals) or ambulance exchange points. Ambulances are primarily used to evacuate casualties suffering from severe and moderately severe injuries that need continuous medical support. Depending on the situation, all casualties are either taken to one single treatment facility or distributed among several of them.

So, mass casualty situations are characterized by the complexity of decisions to be made at collection points and AMP. Unlike clinicians, emergency doctors and paramedics do not have the time to perform a comprehensive and time-consuming evaluation of an organ and body system. In most cases they are focused on free fluid determination because 40% of all trauma-related deaths are due to exsanguination.

Emergency ultrasound has been accepted as an important initial screening tool in disaster medicine because it helps in free fluid detection, allowing finding the cause timely in order to treat it before the victim decompensates.

Ultrasound-competent doctors or paramedics, using portable ultrasound scanners, are able to sample known sonographic windows within minutes, evaluating the peritoneal, pericardial and pleural spaces for free fluid and for a pneumothorax. This way they obtain information that will help emergency crews to perform more accurate re-triage and make more efficient therapeutic decisions, important for saving lives of casualties.

3. State of the Art

Innovative Decision Support System via a new disasterApp

Statistics show that in case of natural disasters, catastrophes and accidents about 70% of affected persons need specific healthcare support fast. For example, 35.2% of patients with severe liver damage die at incident site, 30.6% – on the way to hospital, 11.8% – in stationary, and only 0.4% were healed [7]. Establishment of casualty collection points in a limited time interval, at a safe distance from immediate disaster threats, in a location that allows free access to **transport and communication networks**, is recognized as one of the first important issues in disaster management. Failing to give proper attention to this task could lead to increased human mortality.

Prioritizing between patients is recognized as a very demanding task, and a simulation model for education, research and quality assurance in disaster medicine is proposed [8]. It provides a base for decisions on all levels and all components of the chain of response (including individual patient management) and shows all consequences of such decisions. This simulation model shows differences in accuracy and outcome between two main principles of triage – anatomical and physiological triage – for different categories of staff with different levels of competence and experience, providing a base for discussion when and where to use different methods.

An exercise using mobile application **disasterApp** aimed at replacement of paper-based triage system with digital triage tool is described [9]. It is focused just on storing current casualty records (where they are, what their status is, what medication they have received, and where they are transported to) and collected data replication.

A mobile-based system for supporting emergency triage in the emergency care process for mass casualty incidents is described [10]. This system collects the patient's emergency data throughout the whole emergency care process through a mobile application and data transfer mechanism. It has the capacity to present the survival curve to the triage officer, helping him/her to make triage and transportation decisions.

This system offers an alternative injury assessment tool based on the vital signs data of the injury patient. With the help of this system, the triage officer can more directly and comprehensively learn about each patient's situation and deterioration without additional operations at the incident site.

A spatial tool to support decision making, able to avoid significant processing delays by utilizing precomputed driving times from each location on the road network to each hospital in the study area, is developed [11]. But it is very difficult to estimate the exact travel time after a catastrophe, and what capacities will have hospitals at that time.

Even the authors say: “Unfortunately, incorporating real time traffic data is more complicated, as to do so would significantly extend the time required for computer data processing”.

Use of emergency ultrasound diagnostics at the disaster site is aimed at determining the level of urgency in order to save lives and to prevent any complications for people at risk. Ultrasound [12] was identified as the most sensitive and specific in patients with penetrating chest wounds or in hypotensive blunt abdominal trauma patients (sensitivity and specificity nearly 100%). Ultrasound was performed by relief teams after the 1988 Armenian earthquake as a primary screening procedure in 400 of 750 injured in mass casualty situations patients admitted to a large hospital within 72 h of the event [13]. The average time spent on evaluation of a single patient was approximately 4 minutes.

Traumatic injuries of the abdomen were detected in 12.8% of the patients, with few false-negative (1%) and no false-positive examinations. At present, there are several known protocols for ultrasound examination in case of mass casualty situations: FAST (Focused Assessment with Sonography for Trauma), EFAST (Extended Focused Assessment with Sonography for Trauma), and CAVEAT. FAST is based on the assumption that the majority of clinically significant abdominal injuries result in hemoperitoneum [14]. The standard FAST protocol is directed to detection of fluid in the pericardial and peritoneal spaces. EFAST is the extended version of FAST which includes looking for pneumothorax [15]. So, EFAST, being a systematic protocol, in short time can aid the doctor in performing correctly the patient diagnostics, in situations where every minute counts. CAVEAT [16] is the concept of a comprehensive ultrasound examination in the evaluation of chest, abdomen, vena cava, and extremities in acute triage. According to [17] the completion of the entire CAVEAT protocol by a proficient sonographer will take approximately 5 min longer, than the performance of the traditional FAST examination.

Routine use of FAST ultrasound as a secondary triage tool is described [18]. For example, 20 delayed (Yellow) patients with evidence of hemoperitoneum were identified using portable ultrasound technology, expediting evacuation to definitive care. However, only 30% of these patients subsequently underwent an operative intervention within 24 hours of arrival. Both over- and under-triage are mentioned as significant problems.

Any computer-aided tools, aimed to assist the medical triage process, use methods for quantification and estimation of vital functions affection degree based on a system of describing parameters. Traditionally these techniques are based on scoring systems or analysis of quantified imaging characteristics. Most authors of medical scorings (in particular, [19] and [20]) use statistical analysis (such as the Pearson or Spearman correlation methods) for their creation.

Many support systems for ultrasound examination in hospital conditions are known. But after carefully reviewing available systems we are not able to identify any system that uses portable ultrasound scanners and assist diagnostics under field conditions. This strongly supports the proposed decision support framework, which would provide reliable assistance, easy to use under mass casualty conditions (intense time-pressure and limited medical resources).

4. Proposed Approach - OR Based Optimal Solution

There exist decision support systems for ultrasound clinical examination that help physician in establishing the correct diagnostics in a medical center. But the specific character of mass casualty situations in which medical decisions need to be done fast, without an extensive evaluation, requires an approach, different from the traditional clinical one.

Casualty evacuation without coordination may lead to increased number of deaths. Therefore guidance for rapid transportation (with regard to triage categories, needed/available ambulances and human resources and destination hospitals capabilities) is needed.

Our proposed solution is focused on *reducing the number of victims* by implementing emergency ultrasound in injury assessment at disaster site using portable ultrasound scanners, and *offering easy-to-use computer-aided tools* for mobile devices, which will help to perform triage (based on vital signs) and more accurate re-triage of casualties with injuries of thorax and abdomen (taking into account level of urgency determined by emergency ultrasound).

The new system will suggest efficient therapeutic decisions (life-threatening interventions and emergency diagnostics). Furthermore, it will assist the coordinated evacuation of the injured persons.

Evacuation Logistics to Optimize the Corridors: Medical Analytics
The main proposed objectives are the following:

Obj. 1 – Fast assessment of triage priority for casualties at collection points.
Obj. 2 – Coarse grained medical evaluation before transportation.
Obj. 3 – Fine grained medical evaluation with artificial AI based DSS.
Obj. 4 – Placement and transportation systems for evacuating casualties.

Our distinct solution assists the main functions of casualty collection points: triage and coordinated evacuation of the injured persons. We take into account and integrate the existing approaches, recognized by medical community regarding medical triage [6] and use of emergency ultrasound in injury assessment [14, 15, 21], as well as from evacuation logistics to optimize the corridors [22].

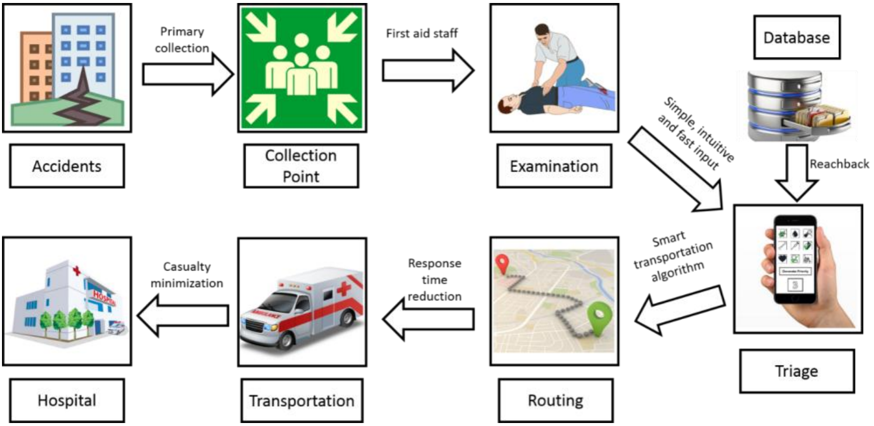


Figure 1. Multilayered Service Oriented Approach (SOA) based on Artificial Intelligence (AI)

The SOA based decision support framework will include:

- **Database:** Knowledge base of diagnostics for pathologies, anomalies and injuries of thorax and abdomen, including those leading to free fluid appearance in abdominal cavity and for a pneumothorax based on ultrasound features.
- **Coarse Grained:** A scoring based on decision rules that allows re-assessment of triage priorities for casualties;
- **Medical Guidance:** Protocol for rapid examination of the abdominal and thoracic cavity using emergency ultrasound and suggestions for life-threatening interventions;
- **Fine Grained:** Inference algorithm to confirm/refute supposition of free fluid presence in abdominal cavity and for a pneumothorax;
- **Efficient Transportation:** Smart algorithms for organization of casualties transportation taking into account selected scenarios specific for concrete situation, affected medical infrastructure, triage categories and available ambulances.
- **REACHBACK:** Holistic Operations Research approaches for efficient and intelligent knowledge based decision making.

  IRIS Integrated Reachback Information System
  More specifically, the so-called IRIS (Integrated Reachback Information System) approach focuses on the development of a technical platform that seeks to support the effective and efficient application and integration of soft and hard OR techniques within a distributed environment.

The following positive impacts are expected on security issues:

Intelligent Assessment

1. Fast registration and triage priority assessment, using the proposed DSS framework.
2. EFAST implementation and suggestions for life-threatening interventions, will imply accurate casualty triage re-assessment and more effective emergency therapy before further transportation and will minimize over- and under-triage.

3. Suggestions for emergency diagnostics, given during transportation, will enhance further clinical treatment which could be embedded in Reachback Operations supported by executive units (see below).
4. Assistance in evacuating casualties will help in efficient distribution of the available resources.

Executive Control Towers

Mass casualty situations are characterized by the urgency for quick and efficient decisions. More and more complex data analytics is required within that complex decision process. In order to support this analytic procedure the authors design an intelligent reachback framework where an additional control tower concept is embedded.

Furthermore, we developed a new app which is connected with that executive control unit.

Both innovative aspects support an intelligent assessment within such a complex triage process.

5. Conclusions and Future Work

The proposed decision support framework for management of mass casualty situations at collection points via an artificial intelligence based multilayered approach will guide the triage of casualties with injuries of thorax and abdomen, taking into account the level of urgency determined by emergency ultrasound, and will suggest efficient therapeutic decisions and assist in the coordination of casualties evacuation.

It will enhance the response procedures at collection points during disaster situations by implementing the best practices/protocols and technologies from emergency diagnostics [23-24], medical informatics [25-26] and operations research [27-29].

The decision support framework will improve triage accuracy and will make a more orderly evacuation process in mass casualty situations. Additionally, it will allow to make casualty re-assessment (at the discretion of sonographer and based on time availability) before a subsequent transportation, thus minimizing over- and under-triage cases.

References

[1] W. Smith. *Triage in mass casualty situations.* Continuing Medical Education, vol. 30, no. 11 (2012), pp. 413-415.

[2] V.V. Nikonov, A.E. Feskov, Eds. *Emergency medicine. Selected clinical lectures*, 3rd ed., vol. 1, Donetsk: Publisher Zaslavsky A.Yu., 2008. (in Russian)

[3] J.E. McGovern. *Mass Casualty Evacuation and Patient Movement.* http://emergencymedicine.health.pitt.edu/sites/default/files/4.9%20Mass%20Casualty%20Evacuation%20and%20Patient%20Movement_0.pdf

[4] S. Adhikari, M. Blaivas, M. Lyon, S. Shiver. *Transfer of real-time ultrasound video of FAST examinations from a simulated disaster scene via a mobile phone.* Prehospital and Disaster Medicine, vol. 29, issue 03 (2014), pp. 290-293.

[5] J. Boer, M. Dubouloz. *Handbook of Disaster Medicine.* CRC Press, 2000.

[6] Gh. Ciobanu, M. Pîsla, et al. *National Guide for Medical Triage in the Mass Casualty Incidents and Disasters.* The Ministry of Health, National Scientific and Practical Center for Emergency Medicine, Republican Center for Disaster Medicine, Chisinau, 2010.

[7] I.I. Sahno, V.I. Sahno. *Disaster Medicine (Organizational issues)*. Moscow, 2002. (in Russian)

[8] K.L. Montán, et al. *Development and evaluation of a new simulation model for education, research and quality assurance in disaster medicine.* https://gupea.ub.gu.se/bitstream/2077/38009/5/gupea_2077_38009_5.pdf

[9] D. Meibner, B. Erb, et al. *Mobile Triage Management in Disaster Area Networks Using Decentralized Replication*, 2016.

[10] Y. Tian, T.S. Zhou, Y. Wang, M. Zhang, J.S. Li. *Design and development of a mobile-based system for supporting emergency triage decision making*. J Med Syst. (2014), 38:65.

[11] O. Amram, N. Schuurman, S.M. Hameed. *Mass casualty modelling: a spatial tool to support triage decision making.* International Journal of Health Geographics (2011), 10:40.

[12] G.S. Rozycki, R.B. Ballard, D.V. Feliciano, J.A. Schmidt, S.D. Pennington. *Surgeon-performed ultrasound for the assessment of truncal injuries: lessons learned from 1540 patients*. Ann Surg., vol. 228(4) (1998), pp. 557-567.

[13] A.E. Sarkisian, R.A. Khondkarian, N.M. Amirbekian, N.B. Bagdasarian, R.L. Khojayan, Y.T Oganesian. *Sonographic screening of mass casualties*

for abdominal and renal injuries following the 1988 Armenian earthquake. J Trauma, vol. 31 (1991), pp. 247-250.

[14] American Institute of Ultrasound in Medicine, American College of Emergency Physicians. *AIUM practice guideline for the performance of the focused assessment with sonography for trauma (FAST) examination.* J Ultrasound Med., vol. 33(11) (2014), pp. 2047-2056.

[15] A.W. Kirkpatrick, M. Sirois, K.B. Laupland, D. Liu, K. Rowan, C.G. Ball, et al. *Hand-held thoracic sonography for detecting post-traumatic pneumothoraces: the Extended Focused Assessment with Sonography for Trauma (EFAST)*. J Trauma., vol. 57(2) (2004), pp. 288-295.

[16] D.E. Hogan, J.L. Burstein, Eds. *Disaster Medicine.* Second edition, Lippincott Williams&Wilkins, a Wolters Kluwer business, 2007.

[17] S.P. Stawicki, J.M. Howard, J.P. Pryor, D.P. Bahner, M.L. Whitmill, A.J. Dean. *Portable ultrasonography in mass casualty incidents: The CAVEAT examination.* World J Orthop, vol. 1(1) (2010), pp. 10-19.

[18] M.D. Sztajnkrycer, A.A. Sztajnkrycer, A. Luke. *FAST ultrasound as an adjunct to triage using the START mass casualty triage system: a preliminary descriptive system.* Prehosp Emerg Care, vol. 10(1) (2006), pp. 96-102.

[19] C.G. Child, J.G. Turcotte. *Surgery and portal hypertension.* Child CG. The liver and portal hypertension. Philadelphia: Saunders (1964), pp. 50-64.

[20] R.N. Pugh, I.M. Murray-Lyon, J.L. Dawson, et al. *Transection of the oesophagus for bleeding oesophageal varices.* British Journal of Surgery, vol. 60(8) (1973), pp. 646-649.

[21] A. Petris, I. Costache, A. Tiron, G. Tatu-Chitoiu, D. Cimpoesu. *Focussed ultrasound examination of the patient in shock: diagnosis, evaluation, monitoring*. Advances in Cardiology 2010, T3 Info Publishing House, Bucharest (2010), pp. 324-341.

[22] K. Kylaheiko, D. Cisic, P. Komadina. *Application of transaction costs to choice of transport corridors.* No. 0004001. EconWPA, 2000.

[23] D. Cimpoesu, L. Rotaru, A. Petris, et al. *Current protocols and guidelines in emergency medicine.* “Gr. T. Popa” UMF Iasi Publishing House, Iasi, 2011.

[24] D. Cimpoesu. *The role of protocols and medical judgment in emergency medical dispatching*. Medical Dispaching Conference, October 2014, Prague.

[25] C. Gaindric, S. Cojocaru, O. Popcova, S. Puiu, Iu. Secrieru. *Emergency-SonaRes: A System for Ultrasound Diagnostics Support in Extreme Cases.* In: Chapter 18 in Springer book "Improving Disaster Resilience and Mitigation - IT Means and Tools" (2014), pp. 283-292.

[26] C. Tambala, Iu. Secrieru. *Portal hemodynamics disorders severity in liver cirrhosis assessment by duplex ultrasound.* Scientific medical journal “Curierul medical”, vol. 59, no. 1 (2016), pp. 37-40.

[27] A. Bordetsky, S. Pickl, B. Reynolds. *Introduction to Network DSS: Decision Support in the Collaborative Environment of Mobile Social and Sensor Networks Minitrack.* 2013 46th Hawaii International Conference on System Sciences (HICSS), IEEE, 2013.

[28] M. Dehmer, F. Emmert-Streib, S. Pickl, A. Holzinger, Eds. *Big Data of Complex Networks*. CRC Press, 2016.

[29] D. Bein, Y. Wen, S. Phoha, B.B. Madan, A. Ray. *Distributed network control for mobile multi-modal wireless sensor networks*. Journal of Parallel and Distributed Computing 71, no. 3 (2011), pp. 460-470.

Constantin Gaindric[1], Svetlana Cojocaru[2], Stefan Pickl[3], Sorin Nistor[4], Iulian Secrieru[5], Olga Popcova[6], Doina Bein[7], Diana Cimpoesu[8]

[1]Institute of Mathematics and Computer Science, Chisinau, Moldova
E-mail: constantin.gaindric@math.md

[2]Institute of Mathematics and Computer Science, Chisinau, Moldova
E-mail: svetlana.cojocaru@math.md

[3]Institute for Theoretical Computer Science, Mathematics and Operations Research, Bundeswehr University Munich
E-mail: stefan.pickl@unibw.de

[4]Institute for Theoretical Computer Science, Mathematics and Operations Research, Bundeswehr University Munich
E-mail: sorin.nistor@unibw.de

[5]Institute of Mathematics and Computer Science, Chisinau, Moldova
E-mail: iulian.secrieru@math.md

[6]Affiliation: Institute of Mathematics and Computer Science, Chisinau, Moldova
E-mail: oleapopcova@yahoo.com

[7]Affiliation: California State University, Fullerton
E-mail: dbein@fullerton.edu

[8]Affiliation: University of Medicine and Pharmacy Gr.T.Popa
E-mail: dcimpoiesu@yahoo.com

Proceedings of the Conference on Mathematical Foundations of Informatics
MFOI'2018, July 2-6, 2018, Chisinau, Republic of Moldova

Optimal algorithm for optimization problems with special restrictions

Eugeniu Gârlă

Abstract

The paper analyses a class of nonlinear optimization problems with special restrictions. We propose a concept for solving the auxiliary problem, for which we calculate complexity. We also assess the maximum number of elementary operations and describe the optimal algorithm for performing numerical calculations with complexity O(nm^2, N). In practical situations, if m << n, then the value of m^2 is much smaller than the value of n, thus it can be considered as a constant, resulting in the complexity O(n^2). In other words it's almost the same as for the approximation methods.

Keywords: computer science, information technologies, algorithm complexity, optimization methods

Nonlinear optimization problems, as well as those in optimal control and economy are matters of extreme values with additional conditions, characterized in that the number of variables is very high. It is well known that most optimization models operate with the reverse for some matrix. If the reverse is found out, then the solution would find the following basic mathematical operations, but the calculation effort required for reverse is very high; there are enormous stability issues, while there is no solution in analytical form, therefore, most methods are iterative methods, i.e. iteration to iteration builds the following sequence

$$f\left(x^0\right) \geq, \ldots, \geq f\left(x^{k-1}\right) \geq f\left(x^k\right) \geq f\left(x^{k+1}\right), \geq \ldots,$$

$x^{k+1} = x^k + \alpha_k p_k$, where α_k - number, p_k - vector, and the calculation process ends, if the difference between two consecutive values of *x* satisfy the inequality $\left| x^{k+1} - x^{k} \right| < \varepsilon,\ \varepsilon$ - given small number and iteration here including ancillary problem solving - key issue, and effectiveness which depends heavily on the initial problem. Based on these considerations, the research made by the author [1,2] focused on the idea of selecting classes of large problems that cause some cases, but covering much of the practical problems and which also may be the proposed effective numerical computation schemes. Whether the initial nonlinear optimization problem in the form (model *PG*)

$$\min\ f(x) \qquad\qquad (1)$$

$$\begin{cases} g_j(x) = 0,\ j = \overline{1,m}.....(2) \\ 0 \le x_i \le a_i, \mathrm{i} = 1, n......(3) \end{cases}$$

the auxiliary problem (*PA*) in the model is a quadratic programming problem with unitary matrix

$$min\left(f_1(p) = 1/2\left\| p \right\|^2 + \left(d, p \right) \right) \qquad\qquad (4)$$

$$Hp = h, \quad H = \begin{bmatrix} h_{11} \dots h_{1n} \\ \dots . \\ h_{m1} \dots h_{mn} \end{bmatrix} \qquad\qquad (5)$$

$$0 \le p_i \le a_i, i = 1, \dots, n. \qquad\qquad (6)$$

where $f\left(x\right), g_j\left(x\right)$ - functions of *x*, continue along with their derivatives, *d*, *h* – vectors, H - matrix form of Jacobean restrictions (2), where $h_j\left(x\right) \equiv g_j^{'}\left(x\right)$ meaning the derivative and the gradient component *i* of gradient $h_j(x)$, denoted respectively with h_{ji}. We cannot separately distinguish here a system of linear equations, the same does not address in detail the conditions necessary of extremum, *PG* convergence model, as such, the stability of the solution, the problem of accumulation of errors, computer techniques etc. (for theoretical and practical results see, for example [1]). Next, we shall examine the complexity of the model and in case of the most difficult operations - auxiliary complexity of the problem, other operations of the model are not essential as time and space. It should be noted that in (2), it is assumed that $m << n$, in other words, restrictions of type "=" are very few, true size of the problem depends essentially on *n*. For (2) - (3) it is defined the projection operator:

$$P = G^T \left(GG^T\right)^{-1} G,$$

which forms the matrix design of restrictions set by an algorithm. The operator *P* has a number of remarkable properties, for which often constitutes the basic method of auxiliary problem solving. But if large issues examined here, if you would proceed front, then it would include restrictions of rectangular type " ≤" which would make it more difficult to reverse of $\left(GG^T\right)^{-1}$ than his very large matrix $\left(m+2\times n, m+2\times n\right)$. Operator *PG* projection model, however, has managed to reduce to calculation matrix inversion

$$R^{-1} = \left(HH^T\right)^{-1},$$

where *H* is composed only of restrictions of type "=" - matrix tape and restrictions of "≤" are *H* considered adding to *H* a column after the formula, or taking out from *H* a column according to a formula similar, in other words, it was shown that consideration of a restriction optimization procedure is equivalent to a column or removing restrictions that column matrix type "=". For the auxiliary problem, it has been demonstrated that in step *k* there is:

- always a non-null *p* direction: $p^{k+1} = p^k - \alpha^k \left(I - P\right)\left(p^k + d\right)$
- calculation formula of Lagrange multipliers:

$$\lambda^{m+j}{}_{k} = \left\{ p^{k} + d \right\}_{i} + \left(H^{T}{}_{j}, \lambda_{k} \right)$$

- solution after a <u>finite</u> number of steps ≤ N , N - constant
- number α^{k} , found after an <u>analytical formula</u>
- method of <u>reverse</u> of $\left(HH^{T}\right)$.

In order to avoid calculating excessive $\left(HH^{T}\right)^{-1}$, in iterative processes frequently applied a common recurrence formula of Woodbury, or the particular case of it - formula Sherman-Morrison with modifications that allow the calculation of the inverse of the new matrix to which was added a column (vector H_{j}) and a row (vector $H^{T}{}_{j}$) without proceeding to its reverse, but using the original matrix R^{-1} and vector product operation.

In the process of calculation however, it is taken full advantage of the decomposition of *P* into the matrix product performing sequentially from the right to left matrix multiplication operation of the vector, the number of elementary operations decreasing significantly. Besides this big advantage, there is one related to stability, because only matrix inversion operation generates more instability. Or, matrix subject to inversion is subject to constant and much smaller dimensions compared to those that are multiplying. Initially, it should be noted that for the calculation of the matrix product $\left(HH^{T}\right)$ we shall use $m^{2} \times (2 \times n - 1)$ elementary operations. We have shown that [1], for multiplying 4 matrixes, optimal bracketing is done by $2 \times n \times m - m +$ $2 \times m \times m - m + 2 \times n \times m - n$ multiplications and sums, or the components of *P* have exactly the same size, thus, minimizing the subspace requires the same number of operations. So, overall, in solving the *(PA)* will be used at most

$$N * \left(m^{2} \times (2 \times n - 1) + 2 \times m \times m - m + 2 \times n \times m - n + n \times (2 \times m - 1) \right)$$

elementary operations. Obviously $N = 1$ if restrictions of the type "≤" are missing. For the reasons stated above it is not taken into account the matrix inversion, which does not depend on the parameter n, but only m; it always has the same constant size - $m \times m$, $m << n$ and the complexity of the operation is $O(m^3)$, while the vector multiplications of the same considerations have linear complexity $O(mn)$, equivalent to $O(n)$. Finally

$$N * \left[2\times m^2 \times n + 4\times m\times n - 2\times n - m - m^2\right].$$

Turning now to *(PA)* (4) - (6) formal N depends on the number of restrictions such as "≤" but because they relate to parts of the variable x, it follows that there are no less than $2\times n$ and N, therefore, depends on n. In the evaluation of N, there can be distinguished several cases. From the foregoing, it follows that the auxiliary problem is rational to introduce initially restrictions; in this case, in step k of interest there is only the situation when some of them become passive, then coming out to the surface of the cuboid. If it turns out that corresponding Lagrange multipliers are non-negative, they fulfil the minimum conditions and numerical computation process has ended. But if the Lagrange multiplier is negative, then the matrix component is brought back to matrix H. It is important that this process is monotonous, functional, decreasing step by step, and each step necessarily terminates after a finite number of operations of finding the minimum. These conclusions arise from the fundamental theorem demonstrating that the operator projection minimizes a quadratic form with symmetric matrix and defined positively in most n steps. But mathematical rigors above essentially improve these results, demonstrating that in the case of unitary matrix, the minimum is reached in one step, reducing the number of elementary operations considerably, which made the developed optimal algorithm to be declared as optimal. Thus, the complexity of this algorithm can be written as

$$O(nm^2, N).$$

In addition to these statements, the respective component, once returned to the matrix determines the Lagrange multiplier appropriate to retain throughout iteration the mark of non-negative in the calculation, which has been proven also by analytical formulas which excludes returning a second time to the surface of the respective component, while maintaining the numerical value of this component within the cuboid, through appropriate choice of step α^k. So, $N \leq 2n$ for lower restrictions on variables, analogically, the same estimation applies to the above restrictions, consequently, overall $N \leq 4n$. The observed trend in setting of indices for the situation

$$J_{activ}(p^0{}_k) \diagup \{j\}\ ,\ J_{pasiv}(p^0{}_k) \cup \{j\}\ ,$$

namely passive indices crowd always broadens and inclusion is strict

$$J_{pasiv}(p^0{}_{k+1}) \supset J_{pasiv}(p^0{}_k),$$

a process which can take only a finite number of times, combined with decreased functionality, it means that the sets of indices $J(p^0{}_k)$ cannot be repeated; as a result - convergence algorithm in solving the problem in a finite number of auxiliary operations.

Of course, such an assessment *PG* optimal algorithm complexity starts to form special restrictions. Therefore, comparing it with other algorithms requires some generalizations and asymptotic estimations. It is known, for instance, that for large optimization problems, trying to solve them using the most popular method - the simplex method often faces difficulties because the number of iterations in this case, increases exponentially compared to the size of the problem. Subsequently, other algorithms have been proposed, within the meaning of polynomial complexity but far from being functional and, therefore, the simplex method still remains the most widely used. Increasing the size of the problem and its passage in large class of problems automatically turns very difficult problem to solve, namely the numerical calculations. It is therefore very important to propose methods and models that would be in some sense “immune” to the extent of the problem. Of course such methods and models are an exception, as most depend essentially on the “curse” of optimization problems. The *PG* model is exactly a happy event, and the minimum number of elementary operations needed to find its solution allows the declaration as optimal algorithm for nonlinear optimization problems with special restrictions, the number of elementary operations being

$$4n * \left[2\times m^2\times n+4\times m\times n-2\times n-m-m^2\right].$$

It should be noted that if (4) - (6) *m* tends to *n* , the model practically turns into a classic linear algorithm, the complexity becomes $O(n^4)$. If we compare this result with the complexity of a record in this field, given in the form $O(nm^5log^2m)$,), m,n - meaning the matrix dimensions restrictions and similar conditions increase up to $O(n^6log^2n)$, there is a decrease in the complexity of the algorithm described above by more than two orders with additional built in algorithm that takes out easily as well as the restrictions on the sign variables. Adding to that, in practical situations, if *m* is less than *n*, $m << n$, then the value of m^2 is much smaller than the value of *n*, thus can be considered as a constant, resulting in the complexity $O(n^2)$, in other words it's almost the same as for the approximation methods.

References

[1] E. Gârlă. *Optimal algorithm for optimization problems with special restrictions.* Economica, no.3 (97), 2016, pp.126-140. ISSN 1810-9136 https://doaj.org/.

[2] E. Gârlă. *The concept of elaboration for an algorithm with optimal complexity.* ROMAI Journal, vol. II, 2017, pp. 255-259.

Eugeniu Gârlă[1],

[1]ASEM
E-mail: *eugeniugarla@yahoo.com*

Diachronic Topic Modelling of Newspapers Articles

Daniela Gifu and Diana Trandabăț

Abstract

Due to historical reasons, it is not rare that related languages become arbitrarily distant and different (including distribution of topics) for a period of time, only to become closer for another period of time. Traditionally, the degree of topic similarity was assets by sociolinguistics on the basis of their expertise starting with of the effect of many aspects governed by some set of rules governing the Society. However, it is hardly possible to cover a large amount of data only by human effort. We present a methodology of diachronic investigation on newspapers articles which determines the significant changes in the distribution of topics that best reflects the difference between historical regional interests.

Keywords: topic similarity, LDA, SVM, diachronic corpora, related languages

1. Introduction

This paper describes a statistical methodology for identifying diachronic topic modelling of newspapers articles in four distinct regions that belong to Greater Romanian. A large corpus (over 4 million lexical tokens), chronologically ordered since the second decade of the 19th century, was developed, structured in independent collections of publications corresponding to Moldavia, Wallachia, Transylvania, and Bessarabia. The results of this contrastive analysis highlight the significant changes in the distribution of topics that best reflects the difference between historical regional interests.

The level of cohesion between epochs is computed using SVM *Support Vector Machine*) classification. The methodology presented here is language independent and it offers a basis for future large-scale studies, having a large impact on reducing the amount of human effort required by socio-historical linguistic analysis of language.

The paper is organized as follows: section 2 presents a brief review of relevant literature, section 3 depicts the data set and topic method based on machine learning algorithms. Finally, the survey conclusions are given in section 4.

2. Related Work

Many previous works [Leech *et al.*, 2009; Davies, 2013] have focused on the linguistic interpretation of the statistical results. Their hypotheses were based on the ways language changes without considering their causes. It has been established that some genetically related languages have a high degree of similarity to each other [Gooskens *et al.*, 2008; Delmestri & Cristianini, 2010].

The development and use of software for natural language processing (NLP) highlight the defining aspects of the Romanian printing press (morphological and syntactic analysis, semantic analysis and, more recently, pragmatic analysis) that has many similarities to that of Bessarabia on the time axis that we have chosen. The rich literature tells its own story regarding the usefulness of technology and information services [Carstensen *et al.*, 2009; Jurafsky & Martin, 2009; Manning & Schütze, 1999; Cole *et al.*, 1998; Popescu & Strapparava, 2013/2014, Gîfu, 2015/2016].

Until now, the Romanian diachronic phenomenon was analysed using various methods. One of them relies on the comparison of writing styles according to various indices: text features [Gîfu *et al.*, 2016], textual formality [Eggins & Martin, 1997], and textual styles [Biber, 1987]. Another one is based on machine learning approach to explore the patterns that govern the lexical differences between two lexicons [Gîfu & Simionescu, 2016].

3. Data and Method

Below, a methodology of diachronic investigation on newspapers articles is presented in order to determine the significant changes in the distribution of topics, reflecting the difference between historical regions.

3.1. Data set

The initial corpus RODICA (*ROmanian DIachonic Corpus with Annotations*), including 4 million lexical tokens, from 6550 original pages), chronologically ordered and covering the period from the second decade of the XIXth century till our days, structured in four independent collections of publications[1], corresponding to Moldavia – 414763 words, Wallachia – 533148, Transylvania – 2579778 words, and Bessarabia[2] – 496855 words, was used in the experiments described in this section.

3.2. Method

The method is based on a large data set, having the following objectives: (1) detection of epoch for the analysed period (1817-2015) and the three Romanian regions (Moldavia, Wallachia and Transylvania); the LDA model is used to identify topics, which are used to border epochs; (2) language similarities is then estimated for each of the epochs detected, in pairs, between Bessarabia and each of the three Romanian provinces, using SVM model.

[1] Bessarabia (Basarabia reînoită; Curierul; Candela; Deșteptarea; Viața economică din Bălți; Solidaritatea; Ehos; Buletinul Arhiepiscopiei Chișinăului; Cuvânt moldovenesc; Ardealul; Basarabia; România nouă; Sfatul țării; Democratul Basarabiei; Glasul Basarabiei; Luminătorul; Dreptatea; Basarabia Chișinăului; Literatura și artă; Moldova Socialistă; Jurnal; Contrafort; Jurnal de Chișinău; Moldova suverană; Ziarul de gardă); Moldavia (Albina românească; Convorbiri literare; Curierul. Foaia intereselor generale; Constitutionalul; Moldova Socialistă; Scânteia; Noutatea; Deșteptarea; Bună ziua, Iași; Ziarul de Vrancea; Monitorul de Vaslui; Evenimentul regional al Moldovei; Imparțial); Transylvania (Organulu Luminarei; Gazeta de Transilvania; Gazeta Transilvaniei; Telegrafulu Românu; Foaia pentru Minte Anima și Literatură; Telegraful român; Transilvania; Federațiunea; Gura Satului; Albina; Telegraful Românu; Familia; Aradu; Patria; Chemarea tinerimei române; Dreptatea; Aradul; Curierul creștin; Vatra românească; Echinox; Adevărul de Cluj; Făclia; Monitorul de Cluj; Bihoreanul); Wallachia (Curier românesc; Buletin. Gazeta oficială; România; Curierul românesc; Pressa, România liberă; Românulu; Timpul; Literatorul; Albina; Deșteptarea. Foaie pentru popor; Adeverul; Curierul artelor; Dimineața; Universul; Viitorul; Curentul; Universul literar; Adevărul; Adevărul literar și artistic; Scânteia; Romania literară; Dimineața copiilor; Evenimentul zilei; Gândul; Ziua; Ziua news; Ziua veche).

[2] Note, that this corpus includes also the Bessarabian texts transliterated from Cyrillic into the Latin script, given that for the period 1944-1989, there were practically no Bessarabian texts in the Latin script.

Figure 1 sketches the general architecture: after an initial epoch detection, features are extracted, which are used to train and then evaluate language similarities based in SVM model.

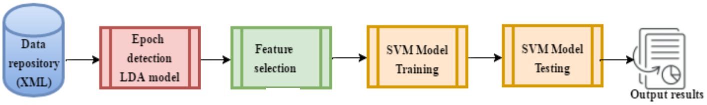


Figure 1. Architecture for the epoch-based language similarity study.

In the Preprocessing phase (Figure 1), diachronic vectors involving each word of the corpus have been built, each slot being a pair

```
<no.occ, year>
```

that records a number of occurrences and a specific year.

The lack of a POS-tagger (that would include a lemmatiser) for old forms had to be substituted by a manual phase, in which a lemma is assigned to the collection of variants of an old form. Then, <no.occ, year> pairs have been summed up for all vectors forms corresponding to the same lemma.

An example of typical vectors of diachronic occurrences, pairing on one side of Bessarabia and on the other any one of Moldavia, Wallachia and Transylvania, looks like this:

```
    <pair>
    w=pace
    Bessarabia = 14 pace 2 1809 2 1918 1 1919 2 1920 2 1929 1
1990 1 1998 1 2009 2 2014
    Moldavia = 3 pace 1 1897     1 1919     1 1990
    </pair>
    <pair>
    w=pace
    Bessarabia=14 pace 2 1809 2 1918 1 1919 2 1920 2 1929 1
1990 1 1998 1 2009 2 2014
    Wallachia = 26 pace 3 1877 15 1878 1 1880 1 1919 1 1928 3
1989 1 1990
    </pair>
    <pair>
    w=pace
    Bessarabia=14 pace 2 1809 2 1918 1 1919 2 1920 2 1929 1
1990 1 1998 1 2009 2 2014
    Transylvania = 742 pace 1 1865 1 1868 14 1877 2 1880 1
1919 230 1949 493 1950
    </pair>
```

As some topics of interest change over time, the distribution of words in newspapers accurately reflects this phenomenon. Thus, by employing a suite of statistical tests, we can determine not-random changes in the word distribution. We tested whether two samples come from the same statistical population or not. We were also interested to see whether there is a large variance with respect to the mean or not, and if the ratio of change from year to year shows an upward or a downward trend.

For a very large corpus, like Google books for example, one can choose an arbitrary set of topics to investigate, but in our study we had a very limited amount of data. Thus, we needed first to indentify a number of topics that are best represented in our corpus, which was done by applying the LDA (Latent Dirichlet Allocation) algorithm. LDA helped us to select, in an iterative process, 10 best represented topics. For each of them, some representative words have been chosen.

An example: Topic → literatura (En: literature) with the corresponding class of words:

`balada, baladei, baladelor, scrisu, scrisulu, scris, satira, satire, caragiale, expozitia, expozitie, publicu, publicului, naratiune, naratiunea, condei, carturar, scrib, scriitori, scriitor, poet, poetilor, poetulu, poetului`

Subsequently, for each of these words, all morphological variants were manually generated, including old forms, and, in order to simplify computations, all occurrences in the corpus were replaced by their corresponding lemmata. The small dimension of the corpus did not allow a secure computation of topics in each Romanian province. As such, the documents from all three provinces have been put together to make up a corpus for the whole Romanian press.

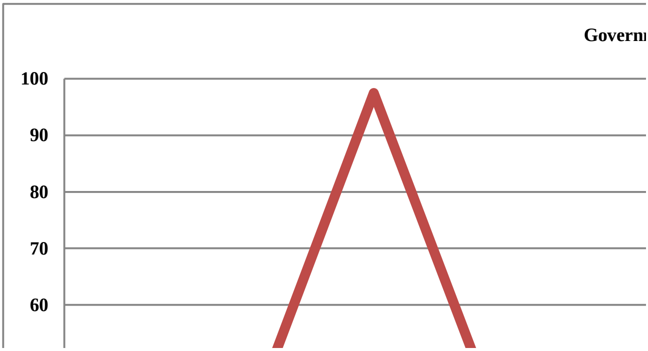
The computation of the number of occurrences of words in epochs corresponding to topics and their weights is done by the following algorithm:

```
for all topics T {
foreach word w in a document d from the period x {
if w in not in topic T {next};
if w in T #occT++;}}
if (#occT <2 ) {next};
if (#occT < 4) {p = (#occ-1) * 1.5; next};
if (#occT < 6) {p =(#occ-1)*2.5; next};
```

```
else {p = (#occ-1)*.5; next};
```

As such, a topic that, for instance, contains in a certain epoch two frequent words, raises also the weight of the words that have been noticed less frequently.

For illustration, below two topics are plotted: *government* and *justice*.



Once topics have been identified, segmentation in epochs could now be computed. This was done, this time, for each Romanian province in part, to facilitate comparison between Bessarabia and each Romanian province in turn. Table 1 shows the segmentation in epochs identified in each of the historical provinces for the established topics.

Table 1 Epochs, as determined by the newspapers' language, in the Romanian provinces

Transylvania	Wallachia	Moldavia	Romania per total
1847-1860	1829-1848	1829-1844	1829-1860
1865-1895	1865-1884	1868-1897	1865-1897
1919-1935	1911-1947	1919-1940	1911-1947
1949-1986	…- 1989	… - 1989	1949-1989

Considering these epochs as categories, we have built an SVM classifier over the whole corpus, with the intention to study the similarity of language between the Romanian written press in Bessarabia and that used in each of the three Romanian provinces, per epochs. According to this idea, the similarity was checked for each news (document) from the target corpus (i.e. Bessarabian) against the collection of documents belonging to each epoch and Romanian provinces. This will give an accurate indication whether there is indeed a similarity, over the epochs, between the source and the target idioms, or the model will assign a more or less random epoch to the target news. In Table 2, the similarity results for each epoch separately are presented.

Table 2 Similarity (as SVM classifier confidence) of Bessarabian documents with those in each Romanian province, per found epochs

Transylvania epoch	Bessarabia	Wallachia epoch	Bessarabia	Moldavia epoch	Bessarabia
1847-1860	52%	1829-1848	61%	1829-1844	85%
1865-1895	57%	1865-1884	64%	1868-1897	73%
1919-1935	63%	1911-1947	69%	1919-1940	75%
1949-1986	59%	NA	-	NA	-
2003-2015	73%	1990-2014	84%	1991-2014	82%

4. Conclusions

We investigated the problem of diachronic topic similarity between the newspapers articles focused on the relation between Romanian (including the historical regions: Moldova, Transylvania, and Wallachia) and Bessarabian. The method is based on statistical analysis of words distributions over epochs reflected on the Romanian printing press and a statistical classifier, SVM, for each epoch. The methodology is language independent and offers an objective quantification of the similarity degree between old Romanian variants.

As further work we plan to investigate the semantic similarity between related languages by employing a deep learning approach as well.

Acknowledgments. This project has received funding from two grants of the Romanian National Authority for Scientific Research and Innovation, UEFISCDI, project number PN-III-P2-2.1-BG-2016-0390, contract 126BG/2016 and project number PN-III-P1-1.2-PCCDI-2017-0818, contract 73PCCDI/2018 within PNCDI III, and partially by the README project "Interactive and Innovative application for evaluating the readability of texts in Romanian Language and for improving users' writing styles", contract no. 114/15.09.2017, MySMIS 2014 code 119286.

References

[1] A. A. Waksman. *Permutation Network.* Journal of the ACM, vol. 15, no 1 (1968), pp. 159-163. D. Biber. *A textual comparison of British and American Writing*. American Speech, (62) (1987), pp. 99-119

[2] K.-U. Carstensen, C. Ebert, S. Jekat, H. Langer, and R. Klabunde (eds.). *Computerlinguistikund Sprachtechnologie: Eine Einführung*. Spektrum Akademischer Verlag (2009).

[3] R. Cole, J. Mariani, H. Uszkoreit, G. V. Battista Varile, A. Zaenen, A. Zampolli, V. Zue (eds.) *Survey of the State of the Art in Human Language Technology*. Cambridge University Press (1998).

[4] M. Davies. *Recent shifts with three non-finite verbal complements in English: Data from the 100-million-word Time corpus (1920s-2000s).* In: Aarts, Close, Leech and Wallis (eds.) The verb phrase in English: Investigating recent linguistic change with corpora, Cambridge: Cambridge University Press (2013), pp. 46-67.

[5] A. Delmestri, N. Cristianini. *String Similarity Measures and PAM-like Matrices for Cognate Identification*. Bucharest Working Papers in Linguistics, 12(2) (2010), pp. 71-82.

[6] P. Diaconescu. *Elemente de istorie a limbii române literare moderne*. Partea I. Probleme de normare a limbii române literare moderne (1830–1880), Bucureşti (1974).

[7] S. Eggins, J. R. Martin. *Genres and Register of Discourse*. In: Dijk, T.A.v. (ed.) Discourse as Structure and Process (Discourse Studies – A Multidisciplinary Introduction), Vol. 1, pages 231–232. Sage Publications, London, UK (1997).

[8] D. Gîfu, M. Dascălu, Ș. Trăușan-Matu, L. Allen. *Time Evolution of Writing Styles in Romanian Language* at the 17th International Conference on

Intelligent Text Processing and Computational Linguistics, CICLing 2016, 3-9 Apr. 2016, Konya, Turkey.

[9] D. Gîfu, R. Simionescu. *Tracing Language Variation for Romanian* at the 17th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2016, 3-9 Apr. 2016, Konya, Turkey.

[10] D. Gîfu. *Contrastive Diachronic Study on Romanian Language*. In: Proceedings FOI-2015, S. Cojocaru, C. Gaindric (eds.), Institute of Mathematics and Computer Science, Academy of Sciences of Moldova (2015), pp. 296-310.

[11] D. Gîfu. *Diachronic Analysis Using a Statistical Model*. In: Proceedings of the Conference on Mathematical Foundations of Informatics, MFOI-2016, Cojocaru, S. and Gaindric, C. (eds.), Institute of Mathematics and Computer Science, Academy of Sciences of Moldova, Chișinău, 2016, pp. 208-221

[12] C. Gooskens, K. Beijering & W. Heeringa. *Phonetic and lexical predictors of intelligibility*. International Journal of Humanities and Arts Computing 2 (1-2) (2008), pp. 63-81.

[13] D. Jurafsky, J. H. Martin. *Speech and Language Processing*. Prentice Hall, 2nd edition (2009).

[14] G. Leech, M. Hundt, C. Mair, and N. Smith. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press (2009).

[15] C.D. Manning, H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press (1999).

[16] O. Popescu and C. Strapparava. *Behind the Times: Detecting Epoch Changes using Large Corpora*. In International Joint Conference on Natural Language Processing, Nagoya, Japan, 14-18 October 2013, pp. 347-355.

[17] O. Popescu, C. Strapparava. *Time corpora: Epochs, opinions and changes*. Knowledge-Based Systems (2014).

[18] R. Simionescu. *UAIC Romanian Part of Speech Tagger*, resource on nlptools.info.uaic.ro, “Alexandru Ioan Cuza” University of Iași (2011).

Daniela Gîfu[1,2] and Diana Trandabăț[1]

[1]Faculty of Computer Science, “Alexandru Ioan Cuza” University of Iași
[2]Institute of Computer Sicence, Romanian Academy – Iasi Branch

e-mail: daniela.gifu@info.uaic.ro

dtrandabat@info.uaic.ro

New Vigenere cipher modification

Eugene Kuznetsov

Abstract

The main aim of this work is to show a new modification of the Vigenere cipher based on the algebraic construction of orthogonal pair. This (fourth) modification improves statistical properties of the cipher-text. The necessary information from the theory of algebraic systems (quasigroups, Latin squares, orthogonal tables, transversals etc.) is provided. The new method of orthogonal pairs generation is constructed and studied.

Keywords: Vigenere cipher, quasigroup, loop, transversal, orthogonal tables.

1 Introduction

This work is dedicated to one of the most important aspects of information security software – encryption methods. There are many reliable encryption algorithms now, but most of them have a low speed of work. In this paper we will continue discussion about modifications of the famous Vigenere cipher. The classical cipher is not of interest today, because there are simple hacking methods by the help of modern computer technologies. But the principles laid down in it, potentially allow us to create quick and at the same time robust ciphers.

The aim of this work is a modification of the Vigenere cipher, in order to improve the statistical properties of the cipher-text obtained after operation. The basis of modification of the classic cipher is an encryption method by bigrams. Its essence lies in the fact that the original message is divided into pairs and each pair of symbols according to a certain law (special sequence table or tables) is encrypted in some other pair of symbols.

The work deals with the (fourth) modification of the classic Vigenere cipher. The necessary information from the theory of algebraic systems (quasigroups, transversals, Latin squares, orthogonal tables etc.) is provided below. Using the properties of these algebraic systems the modification of the cipher is constructed and studied.

2 Modification of Vigenere cipher

2.1 Classical Vigenere cipher

Vigenere cipher is a multi-alphabet advanced encryption system. The idea of the cipher is to use as the key the text of an unencrypted message or encrypted text. The Trithemius table was taken as the basis of the table in its simplest form, which subsequently was dubbed as the Vigenere's table.

Vigenere's table consists of the alphabet shifted cyclically to the left by one character, but other permutations are available too. Additionally, the first line may be a randomly mixed alphabet.

The encryption process is as follows: plain text (which must be encrypted) is written in a line with no spaces. Next, you must determine the key. Vigenere proposed to use as a key the plain text itself, adding to the top of the key a random selected symbol. But as a key it is possible to use any other sequence of characters equal in length to the plaintext.

To produce the cipher-text we take the first letter of the plaintext as an index of the row in a table Vigenere and the letter standing beneath – as a column index. At the intersection of the pair of tables write out the character of the cipher-text. Then repeat these steps for each of the remaining characters.

In order to decrypt the plaintext, you must know the cipher-text and the key. Take the first letter of the key, define the corresponding column in the Vigenere's table and run through it from top to bottom, until you meet the first character of the cipher-text. Once the desired character is met, we write a letter indicating this line, so we get the first character of the plaintext. We do the same steps for the remaining characters of the key and the cipher-text.

In practice, in the programming of the encryption algorithm it is not necessary to have the Vigenere's table in memory, since the encryption algorithm can be represented by some algebraic formula based on such specific algebraic structures, as an orthogonal pair of quasigroups, loops, etc.

2.2 Procedures for encryption and decryption

The encryption procedure by bigrams is similar to the encryption process of the classical Vigenere's cipher, only the first bigram symbol is taken from the first table and the second bigram symbol is taken from the second table (instead of a key sequence, as it was done in the classic Vigenere's cipher). In other words, if we take the table of pairs resulting in the superposition of two orthogonal tables mentioned above, then the plaintext bigram (x, y) corresponds to the encryption bigram (a, b), which is located at the intersection of the x-th row and y-th column. This procedure is repeated sequentially for all bigrams of the encrypted text.

Latin square in the algorithm described above can be changed to another Latin square. The sequence of these squares (or its generation by any algebraic method) is defined by the secret key (or by periodic key sequence). It is easy to see that the statistical hacking algorithms stop working when the number of squares becomes substantially greater than 2.

To eliminate hack statistical methods, several different tables instead of a single one can be used. Then it is obvious that if more different tables to be used, then statistics of a source text will be violated stronger.

3 Algebraic concepts

Hacking classic Vigenere's cipher strongly relies on the presence of a codeword and its length. Therefore, if we save (slightly modified) an encryption method by bigrams, but to refuse from the code word, then the usual method of hacking will not act.

Definition 1. ***Latin square*** *of order n is a square table $n \times n$, where each row and each column contains numbers from 1 to n, and each number is found exactly once.*

Definition 2. *A system $< E, \cdot >$ is called a* ***left (right) quasigroup*** *if the equation $(a \cdot x = b)$ (the equation $(y \cdot a = b)$) has exactly one solution in the set E for any fixed $a, b \in E$. If for some element $e \in E$ we have*

$$e \cdot x = x \cdot e = x \quad \forall x \in E,$$

then a left (right) quasigroup $< E, \cdot, e >$ is called a ***left (right) loop*** *(the element $e \in E$ is called a* ***unit****). A left quasigroup $< E, \cdot >$ that is simultaneously a right quasigroup is called simply a* ***quasigroup****. Similarly, left loop which is simultaneously a right loop is called a* ***loop****.*

From the algebraic viewpoint Latin square is a "multiplication table" of a quasigroup.

If we take two arbitrary Latin squares, the resulting pair of tables may have the property of orthogonality. That is, when overlapping these tables, we obtain a table of pairs of symbols in which each pair of symbols appears exactly once. Algebraically this orthogonal property is described by the following definition.

Definition 3 (see [1])**.** *The operations $A(x, y)$ and $B(x, y)$ on a set E are called orthogonal (or forming an* ***orthogonal pair****), if a system*

$$\begin{cases} A(x, y) = a \\ B(x, y) = b \end{cases}$$

has a unique solution in a set $E \times E$ for any fixed pair $(a, b) \in E \times E$.

Definition 4. *Let G be a group and H be its subgroup. Let $\{H_i\}_{i\in E}$ be the set of all left (right) cosets in G to H, and we assume $H_1 = H$. A set $T = \{t_i\}_{i\in E}$ of representatives of the left (right) cosets (by one from each coset H_i and $t_1 = e \in H$) is called a* ***left (right) transversal*** *in G to H. If a left transversal T is simultaneously a right one, it is called a* ***two-side transversal****.*

On any left transversal T in a group G to its subgroup H it is possible to define the following operation (*transversal operation*) :

$$x \overset{(T)}{\cdot} y = z \quad \overset{def}{\Longleftrightarrow} \quad t_x t_y = t_z h,\, h \in H.$$

Definition 5. *If a system $< E, \overset{(T)}{\cdot}, 1 >$ is a loop, then such left transversal $T = \{t_x\}_{x\in E}$ is called a* ***loop transversal****.*

Lemma 6. *A system $< E, \overset{(T)}{\cdot}, 1 >$ is a left loop with the two-sided unit 1.*

At last, remind how any two left transversals T and P in a group G to its subgroup H are connected .

Lemma 7 (see [3])**.** *Let $T = \{t_x\}_{x\in E}$ and $P = \{p_x\}_{x\in E}$- be left transversals in G to H. Then there is a set of elements $\{h_{(x)}\}_{x\in E}$ from H such that:*

1. $p_x = t_x h_{(x)}\ \forall x \in E$;

2. $x \overset{(P)}{\cdot} y = x \overset{(T)}{\cdot} \hat{h}_{(x)}(y)$.

3.1 Morphisms of quasigroups and loops

Definition 8 (see [1])**.** *A mapping $\Phi = (\alpha, \beta, \gamma)$ (α, β, γ are permutations on a set E) of the operation $< E, \cdot >$ on the operation $< E, \circ >$ is called an* ***isotopy*** *if*

$$\gamma(x \cdot y) = \alpha(x) \circ \beta(y) \quad \forall x, y \in E.$$

If $\Phi = (\gamma, \gamma, \gamma)$, then such an isotopy is called an ***isomorphism****. If $\Phi = (\alpha, \beta, id)$, then such an isotopy is called a* ***principal isotopy****.*

Definition 9 (see [2])**.** *A mapping $\Phi = (\alpha, B, \gamma)$, where α, γ are permutations on E and $B = B(x, y)$ is a right invertible operation on E ($B(x, y) = \varphi_x(y)$, φ_x is a permutation on E $\forall x \in E$), is called a* ***right crossed isotopy*** *(RC-**isotopy**) of operations $< E, \cdot >$ and $< E, \circ >$ if*

$$\gamma(x \circ y) = \alpha(x) \cdot B(x, y) \quad \forall x, y \in E.$$

It is obvious that any isotopy is both RC-isotopy and LC-isotopy simultaneously.

It is easy to show that the orthogonality of operations A and B is equivalent to the fact: the following mapping

$$\Theta = \begin{pmatrix} (1,1) & ... & (x,y) & ... \\ (A(1,1), B(1,1)) & ... & (A(x,y), B(x,y)) & ... \end{pmatrix}$$

is a permutation on the set $E \times E$. The following is true.

Lemma 10. *Let $< E, \cdot, e >$ be a left loop. Then RC-isotop $< E, \circ, e^{'} >$ of the left loop $< E, \cdot, e >$ (by RC-isotopy $T = (\alpha, B, \gamma)$) is a loop $\Longleftrightarrow$ the operations $(\cdot)^{(\alpha, id, id)}$ and B^{-1} are orthogonal.*

3.2 Communication between transformations of transversals and morphisms of transversal operations

Let G be some fixed group and H be its proper subgroup. Consider further the permutation representation $\widehat{G}$ of the group G (note that $\widehat{G} \cong G$, $\widehat{H} \cong St_1(\widehat{G})$).

According to Lemma 7, any two left transversals $T = \{t_x\}_{x \in E}$ and $P = \{p_x\}_{x \in E}$ in G to H are connected with the help of some RC-isotopy (id, B, id) of their transversal operations $< E, \overset{(T)}{\cdot}, 1 >$ and $< E, \overset{(P)}{\cdot}, 1 >$ (where $B(x, y) = \widehat{h}_{(x)}(y)$). It means that if we fix some left transversal T_0 in G to H, then we will receive all other left transversals in G to H from T_0 with the help of RC-isotopy. Moreover, any loop transversal

P in G to H may be received from T_0 with the help of such RC-isotopy (id, B, id) (where $B(x,y) = \widehat{h}_{(x)}(y)$) that the operations $< E, \overset{(T_0)}{\cdot}, 1 >$ and $B^{-1}(x,y) = \widehat{h}^{-1}_{(x)}(y)$ are ortogonal (according to Lemma 10).

If we consider the case $G = S_n$ and $H = St_1(S_n)$, as it is described above, it is possible to express all loops of order n as the RC-isotopies (id, B, id) of some loop $< E, \overset{(T_0)}{\cdot}, 1 >$ of order n, and the operation $< E, \overset{(T_0)}{\cdot}, 1 >$ is orthogonal to the operation $B^{-1}(x,y) = \widehat{h}^{-1}_{(x)}(y)$.

Further we will demonstrate one special case of RC-isotopy of a fixed loop transversal T_0 in G to H, which gives as a result a loop transversal in G to H again. The research will be done by the following scheme:

$$< E, \overset{(T_0)}{\cdot}, 1 > \xleftrightarrow{\Phi} < E, \overset{(P)}{\cdot}, 1 >$$

$$\updownarrow$$

$$T_0 = \{t_x\}_{x \in E} \xrightarrow{\Phi^*} P = \{p_x\}_{x \in E}$$

$$\updownarrow$$

$$p_x = t_x h^{(\Phi)}_{(x)}$$

$$\updownarrow$$

$$\Theta_{(\Phi)} = \begin{pmatrix} --- & (x,y) & --- \\ --- & (x \overset{(T_0)}{\cdot} y, (\widehat{h}^{(\Phi)}_{(x)})^{-1}(y)) & --- \end{pmatrix},$$

where $\Theta_{(\Phi)}$ - is a permutation on a set $E \times E$, corresponding to orthogonal operations $< E, \overset{(T_0)}{\cdot}, 1 >$ and $B^{-1}(x,y) = \widehat{h}^{-1}_{(x)}(y)$.

The above mentioned special case of RC-isotopy corresponds to isomorphism of transversal operations.

Lemma 11. *Let $T = \{t_x\}_{x \in E}$ and $P = \{p_x\}_{x \in E}$ be loop transversals in G to H, and its transversal operations $< E, \overset{(P)}{\cdot}, 1 >$ and $< E, \overset{(T)}{\cdot}, 1 >$ are isomorphic. A permutation Θ on $E \times E$ corresponds to the orthogonal operations ” $\overset{(T)}{\cdot}$ ” and $B^{-1}(x,y)$ (see above), and can be expressed* *in the following form (for some $h_0 \in H$): $\forall x, y \in E$*

$$\Theta = \begin{pmatrix} \ldots & (x, y) & \ldots \\ \ldots & (x \overset{(T)}{\cdot} y, h_0^{-1}(h_0(x) \overset{(T)}{\backslash} h_0(x \overset{(T)}{\cdot} y))) & \ldots \end{pmatrix}.$$

References

[1] V. Belousov. *Foundations of quasigroup and loop theory.* Moscow, Nauka, 1967 (in Russian).

[2] V. Belousov. *Cross isotopies of quasigroups.* Quasigroups and their systems: Mat. issled., vol. 113. Kishinev, Shtiintsa, 1990, pp. 14–20 (in Russian).

[3] E. Kuznetsov. *Transversals in groups. 1. Elementary properties.* Quasigroups and related systems, 1994, vol. 1, No. 1, pp. 22–42.

[4] E. Kuznetsov, S. Novoseltsev. *A modification of Vijener's cipher by the methods of non-associativity algebra.* – ASADE-2007, Abstracts, Chisinau, August 21-23, 2007, 86.

[5] A.J. Menezes, P.C. van Oorschot, S.A. Vanstone. *Handbook of Applied Cryptography,* 2002.

[6] S. Singh. *The Evolution of Secret Writing // The Code Book – The Secret History of Codes & Code-breaking.* – London: Forth Estate, 2000, pp. 3–14.

Eugene Kuznetsov[1]

[1] Institute of Mathematics and Computer Science
Address: MD-2028, Academiei str., 5, Chisinau, MOLDOVA
E–mail: `kuznet1964@mail.ru`

Algebraic representation of generalized boolean algebras

Eugene Kuznetsov, Vladimir Izbash, Olga Izbash

Abstract

The main aim of this work is to demonstrate some algebraic description and representations of generalized boolean algebras. The necessary information from the theory of algebraic systems (semigroups, semirings, etc.) is provided.

Keywords: boolean algebras, generalized boolean algebras, semigroup, semiring, idempotent.

1 Introduction

The work on laying down the foundations of the measure theory was motivated by the discovery of unmeasurable figures on plane or unmeasurable bodies in space. In [2], the notion of measure is defined as an additive function with a boolean algebra as its domain of definition. This treatment of measure complies with the intuition of measure as an additive function, but due to the fact that a boolean algebra has its top element, it is limited to objects of a limited measure. Such a definition of measure is good for probability theory, where the probability of an event is limited by 1.

Stone [3] introduced the “generalized Boolean algebras” and gave as the most representative examples of this notion, the algebras of Lebesgue and Borel measurable sets. The main result of [1] is the introduction of an alternative form of generalized boolean algebra – the form of a monoid (one binary operation with its inverse), called “extension algebra”.

In this paper, generalized Boolean algebras are treated as special commutative algebras with disjunction and conjunction as their fundamental operations. There is yet another reason for this representation – one residing in the nature of mathematics as a science about measure. In this paper the generalized Boolean algebras are viewed as algebras of measurable objects – a view supported by the fact that the main interesting examples of generalized boolean algebras are the Lebesgue or Borel measurable sets of finite measure in n-space.

2 Definitions and preliminary propositions.

2.1 Boolean and generalized boolean algebras.

It is of some interest to examine the possibility of introducing a system with double composition which possesses most of the peculiar properties of Boolean algebras without containing a unit element. The analogy with the theory of abstract rings suggests that the direct or indirect postulation of the unit element should be avoided; and examples from the theory of classes – for instance, the example of all finite subclasses of a given infinite class, or the example of all Lebesgue or Borel measurable sets of finite measure in n-space – indicate the existence of interesting algebras without unit. We shall therefore introduce the following definition:

Definition 1. *A generalized Boolean algebra $A = (A, \vee, \wedge)$ is a system with double composition which satisfies the further postulates:*

Postulate 1. $a \vee b = b \vee a$*;*

Postulate 2. $a \wedge (b \wedge c) = (a \wedge b) \wedge c$*;*

Postulate 3. $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$*;*

Postulate 4. There exists an element 0 *in* A *such that* $a \vee 0 = a$ *for every element* a *in* A*;*

Postulate 5.1. If there exists an element 0 *with the property required by Postulate 4 and if* a *and* b *are elements of* A *such that* $b \wedge a = a$*, then there exists at least one such element* 0*, independent of* a *and* b*, for* *which the simultaneous equations* $x \vee a = b,\ x \wedge a = 0$ *have a solution in* A*;*

Postulate 5.2. If there exists an element 0 *with the property required by Postulate 4 and if* a *and* b *are elements of* A *such that* $a \wedge b = a$*, then there exists at least one such element* 0*, independent of* a *and* b*, for which the simultaneous equations* $x \vee a = b,\ a \wedge x = 0$ *have a solution in* A*;*

Postulate 6.1. $a \vee a = a$*;*

Postulate 6.2. $a \wedge a = a$.

Any element such as that postulated in Postulate 4 is called a *zero*. The solution x of the system in Postulate 5.1 (Postulate 5.2) is called a *left (right) relative complement* of the element a to b.

Definition 2. *A Boolean algebra* $A = (A, \vee, \wedge)$ *is a system with double composition which satisfies the further postulates:*

Postulate 1. $a \vee b = b \vee a$*;*

Postulate 3.1. $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$*;*

Postulate 3.2. $(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c)$*;*

Postulate 4. There exists an element 0 *in* A *such that* $a \vee 0 = a$ *for every element* a *in* A*;*

Postulate 5. If there exists an element 0 *with the property required by Postulate 4, there exists at least one such element* 0 *to which corresponds a fixed element* e *in* A *such that the simultaneous equations* $x \vee a = e,\ x \wedge a = 0$ *have a solution for every element* a *in* A*;*

Postulate 6.1. $a \vee a = a$*;*

Postulate 6.2. $a \wedge a = a$.

Any element such as that postulated in Postulate 4 is called a zero; and any element such as the element e postulated in Postulate 5 is called a *unit*. The solution x of the system in Postulate 5 is called a *complement* of the element a to b.

2.2 Algebraic interpretation of boolean and generalized boolean algebras.

From [3], Theorems 46-50 and Lemmas 11 and 12, one can obtain the following theorem for generalized Boolean algebras.

Theorem 3. *Let a system $A = (A, \vee, \wedge)$ be a generalized Boolean algebra. Then the following properties are fulfilled:*

1. The system $A = (A, \vee)$ is a commutative idempotent monoid with the neutral element 0;

2. The system $A = (A, \wedge)$ is a commutative idempotent semigroup;

3. Both distributive laws take place:
$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$;
$(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c)$.

Analogical theorem may be obtained for Boolean algebras [3].

Theorem 4. *Let a system $A = (A, \vee, \wedge)$ be a Boolean algebra. Then the following properties are fulfilled:*

1. The system $A = (A, \vee)$ is a commutative idempotent monoid with the neutral element 0;

2. The system $A = (A, \wedge)$ is a commutative idempotent monoid with the neutral element e;

3. Both distributive laws take place:
$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$;
$(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c)$.

3 Representation of generalized boolean algebras through boolean algebras.

In [4] the following correspondence was shown.

Theorem 5. *1) Let a system $A = (A, \vee, \wedge)$ be a generalized Boolean algebra. Let us define binary operations $\cdot$ and $+$ on the set A by the following way:*
$x \cdot y = x \wedge y$,

and $x + y$ is a relative complement of the element $x \wedge y$ up to the element $x \vee y$. Then the system $A = (A, +, \cdot)$ is a boolean algebra.

2) Let a system $A = (A, +, \cdot)$ be a boolean algebra. Let us define binary operations $\wedge$ and $\vee$ on the set A by the following way:

$x \wedge y = x \cdot y,$

and

$x \vee y = x + y + x \cdot y.$

Then the system $A = (A, \vee, \wedge)$ is a generalized Boolean algebra.

4 Algebraic representation of generalized boolean algebras

Using terminology and some results from [5], one can obtain another characterization of generalized boolean algebras.

Definition 6. *A system $A = (A, +, \cdot)$ with two binary operations $\cdot$ and $+$ is called a semiring if the following conditions hold:*

1. A system $A = (A, +)$ is a commutative semigroup;

2. A system $A = (A, \cdot)$ is a semigroup;

3. Both distributive laws take place:

$a \cdot (b + c) = (a \cdot b) + (a \cdot c);$

$(a + b) \cdot c = (a \cdot c) + (b \cdot c).$

Definition 7. *Let system $A = (A, +, \cdot)$ be a semiring. It is called*

- a semiring with zero 0, if there exists an element $0 \in A$ such that $0 + a = a + 0 = a$ and $0 \cdot a = a \cdot 0 = 0$ for every $a \in A$,

- a semiring with unit 1, if there exists an element $1 \in A$ such that $1 \cdot a = a \cdot 1 = a$ for every $a \in A$,

- a commutative semiring, if the identity $a \cdot b = b \cdot a$ holds for every $a, b \in A$,

- a multiplicative idempotent semiring, if the identity $a \cdot a = a$ holds for every $a \in A$,

- an additive idempotent semiring, if the identity $a + a = a$ holds for every $a \in A$,

- an idempotent semiring, if it is both multiplicative idempotent and additive idempotent semiring,

- a distributive semiring, if the identity $a + b \cdot c = (a + b) \cdot (a + c)$ holds for every $a, b, c \in A$.

Now we can reformulate Theorems 1 and 2 in the following way.

Theorem 8. *Let a system $A = (A, \vee, \wedge)$ be a generalized Boolean algebra. Then the system $A = (A, \vee, \wedge)$ is a commutative idempotent semiring with zero.*

Theorem 9. *Let a system $A = (A, \vee, \wedge)$ be a Boolean algebra. Then the system $A = (A, \vee, \wedge)$ is a commutative idempotent semiring with zero and unit.*

Definition 10. *Let ρ be an equivalence relation on an arbitrary semiring $A = (A, +, \cdot)$. It is called a congruence if for every $a, b, c \in A$ we have an implication*

$$a \rho b \quad \Rightarrow \quad (a + c)\rho(b + c),\ (a \cdot c)\rho(b \cdot c),\ (c \cdot a)\rho(c \cdot b).$$

Definition 11. *A semiring C with zero 0 is called an 0-extension of a semiring A by semiring B, if there exists a congruence ρ on C such that $[0]_\rho \cong A$ and $C/\rho \cong B$.*

According to Propositions 2.3 and 2.6 from [5], we obtain the following result.

Theorem 12. *Let A be a generalized Boolean algebra. Then*

1. A is a 0-extension of some Boolean algebra by suitable idempotent semiring;

2. A is a 0-extension of some Boolean algebra by distributive semiring.

References

[1] I. Drugus. *Generalized Boolean Algebras as Single Composition Systems for Measure Theory.* Proceedings of the 4th Conference

of Mathematical Society of Moldova (CMSM4), June 28 - July 2, 2017, Chisinau, Republic of Moldova, 2017, pp. 75–78.

[2] A. Horn, A. Tarski. *Measures in Boolean algebras*, Trans. Amer. Math. Soc., vol. 64 (1948), pp. 467–497.

[3] M.H. Stone. *Postulates for Boolean Algebras and Generalized Boolean Algebras.* American Journal of Mathematics, vol. 57(1935), no 4, pp. 703–732.

[4] M.H. Stone. *The theory of representations for Boolean algebras.* Trans. Amer. Math. Soc., vol. 40(1936), pp. 37–111.

[5] E.M. Vechtomov, A.A. Petrov. *Multiplicative idempotent semirings.* Fundamental and applied mathematics, vol. 18(2013), no 4, pp. 41–70. (in Russian)

Eugene Kuznetsov[1], Vladimir Izbash[2], Olga Izbash[3]

[1]Institution Institute of Mathematics and Computer Science, Academy of Sciences, MOLDOVA
Address: MD-2028, Academiei str., 5, Chisinau, MOLDOVA
E–mail: kuznet1964@mail.ru

[2]Institution Institute of Mathematics and Computer Science, Academy of Sciences, MOLDOVA
Address: MD-2028, Academiei str., 5, Chisinau, MOLDOVA
E–mail: vladimir.izbas@math.md

[3]Institution Institute of Mathematics and Computer Science, Academy of Sciences, MOLDOVA
Address: MD-2028, Academiei str., 5, Chisinau, MOLDOVA
E–mail: olga.izbas@math.md

Tree-like Refutation Search and Model Elimination Method

Alexander Lyaletski

Abstract

Calculi of so-called literal trees are constructed for solving the problem of refutation search in classical first-order logic without equality. Their soundness and completeness are proved. The connection of one of these calculi with a certain modification of the model elimination method is established, which gives the possibility to prove the soundness and completeness of the both modification and model elimination method and better understand the operation scheme of the model elimination method.

Keywords: classical first-order logic, clause, resolution, model elimination method, literal tree, calculus, inference, refutation, completeness, soundness.

1 Introduction

Investigations in automated theorem proving gave rise to the appearance of various machine methods for refutation search in classical first-order logic. Among them, the famous resolution method [1] should be mentioned first of all. A little later, the model elimination method was appeared [2]. Despite the fact that they use a similar resolution technique for producing their new objects, they operate with different kinds of them: the resolution method — with usual clauses and the model elimination method — with clauses containing so-called framed literals whose role in refutation search is not especially clear from the content side, which leads to a misunderstanding of the features of the model elimination method.

This research is intended to fill a gap in this misunderstanding by means of using the notion of a so-called literal tree and establishing its connection with the notion of a clauses with framed literals. As a result, we become able to prove the soundness and completeness of a certain modification of the model elimination method on the basis of the soundness and completeness of so-called literal tree calculi and better understand the operation scheme of the model elimination method.

2 Preliminaries

We consider classical quantifier-free first-order logic in its clause form, using $\neg$ for denoting negation and $\vee$ for denoting disjunction, and refer to [3] for all resolution notions and to [4] for modern vision of the model elimination method and its relation to other computer-oriented methods for refutation search.

We assume a reader to be familiar with the first-order semantic notions, in particular, with satisfiability (see, for example, [3]).

Remind some of the main definitions and notions of our clause logic; at that, the notions of term, atomic formula, and formula are considered to be known.

A *ground* expression (i.e. a formula or term) is an expression without free variables.

The result of the renaming of some or all variables in an expression under consideration is called its *variant*.

An *inference* in any calculus under consideration is defined as a sequence of well-formed expressions, every of which is either a variant of an "input" (preliminary given) expression or a variant of an expression deduced from some of the previous ones according to a rule of the calculus.

If A is an atomic formula, then both A and $\neg A$ are called *literals*.

If L is a literal, then $\tilde{L}$ denotes its *complement*, which is defined as $\neg A$ in the case, when L is an atomic formula A, and as A in the case, when L is $\neg A$.

A formula of the form $L_1 \vee \ldots \vee L_n$ is called a *clause*, where

$L_1, \ldots, L_n$ are literals. (That is a clause can be considered as an ordered multi-set in terminology of [5].) The clause without literals is denoted by $\sharp$ and called the *empty clause* (considered as an unsatisfiable formula).

Remark. Presentation of a clause as a formula permits to distinguish different occurrences of the same literal in a clause. Also note that, for the purpose of this paper, it is important that any clause presents a "linear" expression when reading it from left to right, which determines a linear order over its literals according to the way of this "reading".

In what follows, IS always denotes an initial set of clauses called *input clauses* and our nearest purpose is to construct a refutation technique in the tree form for the establishing of the unsatisfiability of the set IS.

A reader is considered to be familiar with the notions of a substitution, unifier and most general unifier (mgu), which are defined in the same way as in [1] (see also [3]).

A substitution is said to be *ground* if and only if all the terms substituted for its variables are ground.

If W denotes an expression, including a tree, and σ is a substitution, then the result of application σ to W is understood in the sense of [3] and is denoted by $W{*}\sigma$. For any set Ex of expressions, $Ex{*}\sigma$ denotes the result of the application σ to every expression from Ex.

Also following to [1] and [3], we introduce the operation of multiplication of substitutions μ and λ as an operation, denoted by $*$, that possesses the property: $W(\lambda{*}\mu) = (W{*}\lambda){*}\mu$ for any expression W.

In our research we use the following forms of the resolution rule.

Resolution rule R (cf. [3]). Let clauses C_1 and C_2 be of the forms $C_1' \vee L$ and $C_2' \vee E \vee C_2''$ respectively, where L and E are literals and C_1', C_1', and C_2'' are (possibly, empty) clauses. If there exists the mgu σ of the set $\{E, \tilde{L}\}$, then the clause $(C_1' \vee C_2' \vee C_2'') * \sigma$ called a *resolvent* is said to be deduced by the *resolution rule* R from C_1 and C_2.

Remark. Usually, the *factor rule* (see, for example, [3]) is necessary for providing the completeness of a deduction system with a resolution rule in the form from [3], but for our purpose it is superfluous, because the calculi under consideration allow it to be abandoned.

Besides, we use the following clause form of the Herbrand theorem taken from [3].

Herbrand theorem. A finite set S of clauses is unsatisfiable if and only if there exist a finite set S' of variants of clauses from S and a ground substitution σ of terms from the Herbrand universe for S for all the variables from S' such that $S'{*}\sigma$ is unsatisfiable.

3 Refutation search in literal tree calculi

Our further considerations and results are connected with so-called *literal tree calculi.*

A tree is understood in the usual sense and consists of *nodes, root, leaves*, and *branches.* A tree not containing any nodes is denoted by Δ and called the *degenerative tree.*

Well-formed expressions of LC are trees with nodes labeled by literals or by the symbol $\diamondsuit$. We often identify a node with its label and suppose any tree to grow "from top to bottom".

As in the case of expressions, any renaming of variables in a literal tree is called its *variant.*

If σ is a substitution and Tr is a tree, then $Tr{*}\sigma$ is the result of the application of σ to all the leaves of Tr.

Each input clause $C = L_1 \vee \ldots \vee L_n$ from an initial set IS, where $L_1, \ldots, L_n$ are literals, induces an *initial tree* Tr_0 *w.r.t.* C according to the following: Tr_0 consists of the root labeled by $\diamondsuit$ and of n its "heirs" labeled by $L_1, \ldots, L_n$, when looking through "heirs" from left to right.

A literal tree is called *closed* in the case if any its leaf has $\sharp$ as its label.

Let $Tr_0, \ldots, Tr_m$ be a sequence of literal trees, where Tr_0 is an initial tree w.r.t. some $C \in IS$ and Tr_{i+1} is a variant of a tree deduced from a literal tree preceding to Tr_{i+1} by a inference rule of a calculus under consideration. Then $Tr_0, \ldots, Tr_m$ is called an *inference of* Tr_m

from Tr_0 in this calculus.

We solve the refutation search problem as an *inference search of a certain object* in a literal tree calculus containing resolution-type rules incorporated into literal trees.

Input Extension (IE) rule. Let $C = E_1 \vee \ldots \vee E_j \vee \ldots \vee E_k$ be a clause from IS, Tr a literal tree not having common variables with C, and L a label of its the rightmost leaf distinguished from $\sharp$, where $E_1, \ldots, E_k$ are literals. Suppose there exists the mgu σ of the set $\{E_j, \tilde{L}\}$, and that Tr' is constructed from Tr by adding $k-1$ "heirs" to the node with L that are labeled from left to right by E_1, …, E_{j-1}, $E_{j+1}, \ldots$, E_k, respectively (if $k = 1$, the only "heir" $\sharp$ is added to the node with L). Then the tree $Tr * \sigma$ is said to be deduced by the *input extension (IE)* rule from Tr w.r.t. C.

Contrary Closing (CC) rule. Let Tr be a literal tree and Br the rightmost branch of Tr containing a leaf Lf labeled by a non-empty literal L. Let Br contains such a node labeled by a literal E, that there exists the mgu σ of the set $\{E, \tilde{L}\}$. If Tr' is constructed from Tr by adding only one "heir" $\sharp$ to the node with L, then the tree $Tr * \sigma$ is said to be deduced by the *contrary closing (CC)* rule from Tr.

Chain Deleting (CD) rule. Let Tr be a literal tree, and Br be a branch of Tr containing leaf Lf labeled by $\sharp$. Let Ch denote the maximal part of Br, such that Ch is ended by $\sharp$ and each node of Ch with a label distinquished from $\sharp$ has only one successor. If Tr' denotes the result of a deletion of Ch from Tr, Tr' is said to be deduced by *chain deleting (CD)* rule from Tr.

These three inference rules produce two literal tree calculi, one of which denoting by LC contains the IE and CC rules, while the second denoting by $LC^{\sharp}$ contains the IE, CC, and CD rules.

In what follows, we assume the *CD rule to be automatically applied in $LC^{\sharp}$* after any application of both the IE and CC rules. in this connection, the rules IE and CC, satisfying this convention will be denoted by $IE^{\sharp}$ and $CC^{\sharp}$, respectively.

Lifting lemma. Let $Tr_1, \ldots, Tr_m$ be literal trees and σ – a ground substitution such that $Tr_1{*}\sigma, \ldots, Tr_m{*}\sigma$ is an inference of $Tr_m{*}\sigma$ from $Tr_1{*}\sigma$ in the LC ($LC^\sharp$) calculus and each tree $Tr_i{*}\sigma$ does not contain variables ($i = 1, \ldots, m$). Then there exists in LC ($LC^\sharp$) an inference $Tr'_1, \ldots, Tr'_m$ of Tr'_m from Tr'_1 such that $Tr'_1 = Tr_1$ and for some substitutions $\lambda_1, \ldots, \lambda_m$ $Tr'_1{*}\lambda_1 = Tr_1{*}\sigma$, …, $Tr'_m{*}\lambda_m = Tr_m{*}\sigma$.

Proof. This lemma can be easily proved by induction on m using the definitions of the IE, CC, and CD rules and the mgu property, according to which for any unifier λ of a set Ex of expressions the following equality holds: $\lambda = \mu{*}\lambda'$, where μ is the mgu of Ex and λ' sone substitution. □

Let Tr be a literal tree and $L_1, \ldots, L_n$ its literals from all leaves of Tr that are distinguished from $\sharp$ and written in the order of looking through the leaves from left to right. Then $lc(Tr)$ is the clause $L_1 \vee \ldots \vee L_n$ called the *local clause image* of Tr.

Additionally, we consider that by definition $lc(Tr) = \sharp$ if and only if Tr is a closed tree, Tr consists of only the root, or $Tr = \Delta$.

For the case of the clause logic without equality, we have the following result.

Theorem 1. *Let IS be an initial set of clauses and C such a clause from IS that the set $IS \backslash \{C\}$ is satisfiable. The set IS is unsatisfiable if and only if in the calculus LC ($LC^\sharp$) there exists an inference of a closed tree (the degenerative tree Δ) from the initial tree w.r.t. C.*

Proof. Note that this theorem is sufficient to prove only for the LC calculus, since $lc(Tr) = lc(Tr')$ for any literal tree Tr and a tree Tr' being the result of the application of the CD rule to Tr. Besides, we can restrict us by proving this theorem for the case of IS containing only ground clauses due to the following reason.

The above-given Herbrand theorem gives us the possibility to reduce the establishing of the unsatisfiability of IS to the establishing of the unsatisfiability of a certain set of ground clauses while the lifting lemma permits to generalize the proved soundness and completeness of LC and $LC^\sharp$ for sets of ground clauses on the case of sets of arbitrary clauses. (If Tr is a closed (degenerative) tree, then and only then we can say the same about $Tr{*}\sigma$ for any substitution σ.)

So, let IS is an initial set of ground clauses. At once note that without loss of generality we can consider that the following convention is satisfied: the inference rules of our calculi are always applied in a tree under consideration to the leaf node of its leftmost branch until a leaf with $\sharp$ appears (if it can be done).

Soundness. Let $Tr_1, \ldots, Tr_m$ be an inference of Tr_m from Tr_1 w.r.t. C and Tr_m a closed tree (for which $lc(Tr_m) = \sharp$). Obviously, it is sufficient to prove by induction on j ($j = 1, \ldots, m$) that $lc(Tr_j)$ is a logical consequence of IS.

If $j = 1$, then $lc(Tr_j) = C$ and $lc(Tr_j)$ is a logical consequence of IS, because $C \in IS$.

Suppose $j > 1$ and $lc(Tr_1), \ldots, lc(Tr_{j-1})$ are logical consequences of IS. Then Tr_j is the result of the application of the IE or CC rule.

If Tr_j is the result of the IE application, then $lc(Tr_j)$ is the result of the application of the resolution rule R to $lc(Tr_{j-1})$ and some clause from IS. Due to the soundness of R we obtain the desirable result.

If Tr_j is the result of the CC application, then, taking into account the above-accepted convention, $lc(Tr_j)$ can be considered as the result of the application of the R rule to $lc(Tr_{j-1})$ of the form $L \ \vee \ C_1$ and some $lc(Tr_k)$ ($1 \leq k < j-1$) of the form $\tilde{L} \ \vee \ C_2$, where L is a literal and C_1 and C_2 are clauses. Again, due to the soundness of R, we obtain what we wanted to prove.

Completeness. Because $IS \backslash \{C\}$ is a satisfiable set, we can suppose that IS is a minimal unsatisfiable set of clauses. Besides, since for any tree Tr the result of the application of the CD rule is Δ if and only if Tr is a closed tree, it is sufficient to prove the completeness of LC.

The completeness of LC is proved by induction on the difference r between the number of all literals occurred in clauses from IS and the number of all clauses in IS.

If $r = 0$, then $IS = \{L, \tilde{L}\}$ due to the minimality and unsatisfiability of IS (L is a literal) and, therefore, an arbitrary initial tree Tr_1 for this IS can consist of only the node and its unique "heir" labeled by

L or $\tilde{L}$. Independent from the labeling by L or $\tilde{L}$, the IS rule can be applied to Tr_1 producing a closed tree, which was to be proved.

Suppose $r > 0$, $C \in IS$, and $C = L_1 \vee \ldots \vee L_k$ $(k \geq 1)$. Prove the existence of an inference of a closed tree w.r.t. C by induction on k $(L_1 \vee \ldots \vee L_k$ are literals).

Let $k = 1$, that is $C = L_1$. Then the initial tree Tr_1 w.r.t. C consists of the root and its unique leaf node labeled by L_1.

Because of the minimality and unsatisfiability of IS, IS contains a clause $C' = \tilde{L_1} \vee D$, where D denotes a non-empty clause of the form $E_1 \vee \ldots \vee E_n$ for some literals $E_1 \ldots, E_n$ (that is $n \geq 1$).

If we consider the set $IS' = (IS \setminus \{C'' \mid C'' \in IS$ and C'' contains $L_1\}) \cup \{D\} \cup \{L_1 \mid$ if $\tilde{L_1}$ occurs at least in one clause from $IS \setminus \{C'\}\}$, then it is easy to prove that IS is unsatisfiable if and only if IS' is unsatisfiable.

The difference between the number of all literals occurred in clauses from IS' and the number of all clauses in IS' is less than r. Hence, by the induction hypothesis, there exists a literal tree inference $Tr'_1, \ldots, Tr'_k$ from Tr'_1 w.r.t. D, in which Tr'_k is a closed tree (here $k \geq 2$, since D is a non-empty clause).

Using $Tr'_1, Tr'_2, \ldots, Tr'_k$, let us construct a sequence of literal trees $Tr_0, Tr_1, Tr_2, \ldots, Tr_m$ in the following way.

Declare as Tr_0 the initial tree w.r.t. C (it consists of the root and its unique node labeled by L_1). Then we can apply the IE rule to Tr_0 w.r.t. C' and deduce the tree denoted by Tr_1, all leaves of which contain the same labels that Tr'_1 has and only them. That is $lc(Tr_1) = lc(Tr'_1)(= D)$.

Now, let us suppose that we have constructed $Tr_0, Tr_1, \ldots, Tr_j$ $(j \geq 1)$ on the basis of $Tr'_1, \ldots, Tr'_i$ $(1 \leq i < k)$ possessing the following *property* ($\sharp$) (which is true for Tr_1 and Tr'_1 by the construction of Tr_1):

Property ($\sharp$). Let T'_j be the subtree of Tr_j with the node labeled by L_1 as its root and T_j be the result of the replacing of L_1 at the root of T'_j by $\Diamond$. Then after removing all the leaves labeled by $\sharp$ in T_j and Tr'_i we receive the same literal tree.

Consider all possible cases of the deduction of Tr'_{i+1} from Tr'_i and continue the inference $Tr_0, Tr_1, \ldots, Tr_j$ by adding to it (i) one new literal tree (denoted below by Tr_{j+1} and possessing the property ($\sharp$)) or (ii) two new literal trees (denoted below by Tr_{j+1} and Tr_{j+2}, the last of which possesses the property ($\sharp$)). For this, let us consider the leftmost branch of Tr'_i labeled by a literal, say, E, distinguished from $\sharp$ and "participating" in the application of an inference rule in deducing Tr'_{i+1}.

The literal E cannot be L_1, since IS' contains only one clause, namely, L_1, with L_1 occurring in it.

Let E be $\tilde{L_1}$. Then only the IE rule can be applied w.r.t. L_1 to the node with E (remind that $L_1 \in IS'$). As a result, the leaf with $\sharp$ adds to Tr'_i as the "heir" of the node with E, producing Tr'_{i+1}.

This implies that the CC rule can be applied to Tr_j w.r.t. the branch containing $\tilde{L_1}$ and L_1 being the label of the unique "heir" of the root of Tr_j. The result of this application can be taken as the tree Tr_{j+1} mentioned above in (i) (it possesses the property ($\sharp$) by the way of its deducing from Tr_j possessing the property ($\sharp$)).

Let E be distinguished from $\tilde{L_1}$. First, suppose that Tr'_{i+1} is the result of CC application to Tr'_i. Hence, the leftmost branch containing the leaf with E also contains a node with $\tilde{E}$ as its label. The literal E is distinguished from both L_1 and $\tilde{L_1}$. According to the property ($\sharp$), we obtain that the leftmost branch of Tr_j, not containing $\sharp$ as the label of its leaf, contains $\tilde{E}$ and E (being a leaf label in Tr_j). That is the CC rule can be applied to Tr_j leading to the deducing of the required tree Tr_{j+1} (i.e. a tree possessing the property ($\sharp$)).

Now, suppose that Tr'_{i+1} is the result of the IE application to Tr'_i w.r.t. $D' \in IS'$. If $D' \neq D$, then $D' \in IS$ and, due to satisfying the property ($\sharp$) for Tr_j and Tr'_i, the IE rule can be applied to Tr_j w.r.t. D' in such a way that the result of this application is the required literal tree Tr_{j+1}.

For $k = 1$ it remains to consider the case the IE application to Tr'_i w.r.t. $D' \in IS'$, when $D' = D$ ($= E_1 \vee \ldots \vee E_n$, where $n \geq 1$). That is the literal distinguished from $\sharp$ and being the label of the leaf node of the leftmost branch of Tr'_i is one of the literals $\tilde{E}_1, \ldots, \tilde{E}_n$, say, for definiteness, that it is $\tilde{E_n}$. This means that the IE rule applied to Tr'_i producing Tr'_{i+1} with new (w.r.t. Tr'_i) leaves having $E_1, \dots, E_{n-1}$ as their labels.

Since $\tilde{L_1} \vee D \in IS$, then, due to satisfying the property ($\sharp$) for Tr_j and Tr'_i, the IE rule can be applied to Tr_j w.r.t. C' producing the tree Tr_{j+1} with $\tilde{L_1}, E_1, \dots, E_{n-1}$ as the labels of new (w.r.t. Tr_j) leaves. In Tr_{j+1}, $\tilde{L_1}$ is the label of the leftmost branch with a leaf node, the label of which is distinguished from $\sharp$. We can apply the CC rule to Tr_{j+1} w.r.t. $\tilde{L_1}$ and, as it is obvious, deduce Tr_{j+2}, which along with Tr_{j+1} can be considered as trees mentioned above in (ii).

The consideration of the case of $k = 1$ (the induction basis) is completed.

Let $k > 1$. (Remind that $C = L_1 \vee L_2 \vee \dots \vee L_k$.)

Consider $IS_1 = \{L_1\} \cup (IS \backslash \{C\})$ and $IS_2 = \{L_2 \vee \dots \vee L_k\} \cup (IS \backslash \{C\})$. Because of the unsatisfiability of IS these sets are unsatisfiable.

Let IS'_1 and IS'_2 denote the minimal unsatisfiable sets of IS_1 and IS_2 respectively. Since $IS \backslash \{C\}$ is a satisfiable set, $L_1 \in IS'_1$ and $L_2 \vee \dots \vee L_k \in IS'_2$. By the induction basis and hypothesis, there exist an inference $Tr'_1, \dots, Tr'_{m_1}$ w.r.t. L_1 and an inference $Tr''_1, \dots, Tr''_{m_2}$ w.r.t. $L_2 \vee \dots \vee L_k$, in which Tr'_{m_1} and Tr''_{m_2} are closed trees.

Construct a sequence of literal trees $Tr_1, \dots, Tr_{m_1}, Tr_2, \dots, Tr_{m_2}$ in the following way.

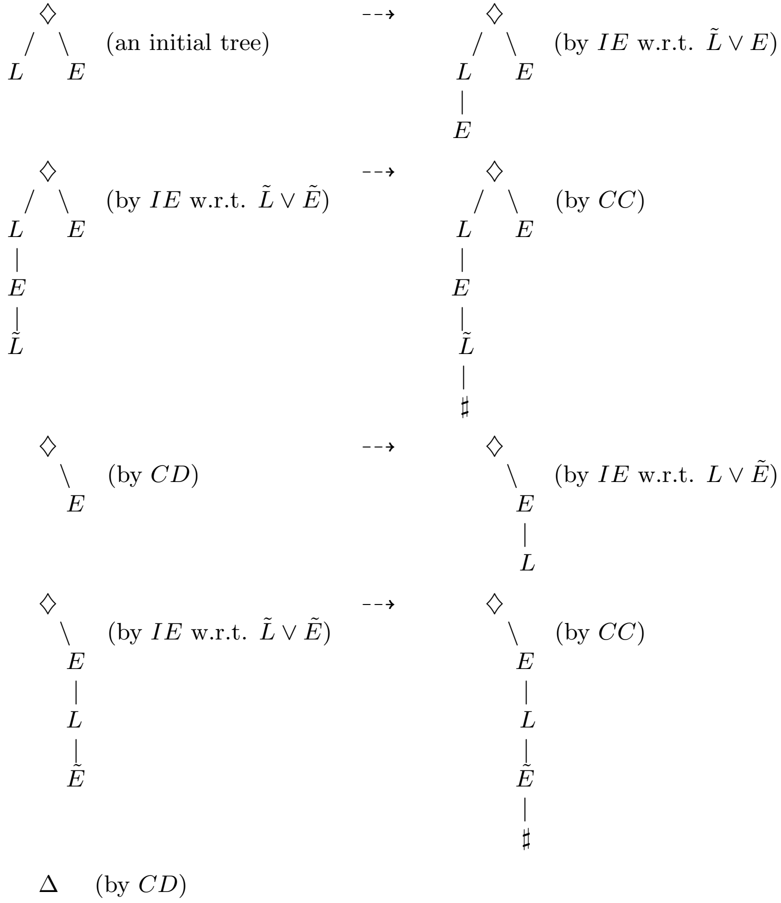
Let $Tr_1, \dots, Tr_{m_1}$ be the trees sequence, in which for each $i = 1, \dots, m_1$, Tr_i is constructed from Tr'_i by adding $(k-1)$ leaf nodes labeled by $L_2, \dots, L_k$ to the root of Tr'_i, when looking through the leaves from left to right.

We see that $Tr_1, \dots, Tr_{m_1}$ is an inference in the LC calculus and if we remove from Tr_{m_1} the subtree with the root labeled by L_1, then we receive the initial tree Tr''_1 w.r.t. $L_2 \vee \dots \vee L_k$. Therefore, it is obvious that on the basis of the induction hypothesis the inference $Tr_1, \dots, Tr_{m_1}$ can be prolonged in the direction of the construction of the above-mentioned sequence $Tr_1, \dots, Tr_{m_1}, Tr_2, \dots, Tr_{m_2}$ being an inference in LC, in which Tr_{m_2} is a closed tree.

The theorem is proved. □

Give a simple example of the tree-like inference search.

Let $IS^* = \{L \vee E, L \vee \tilde{E}, \tilde{L} \vee E, \tilde{L} \vee \tilde{E}\}$, where E and L are literals. The following sequence of literal trees is an inference of Δ in the $LC^\sharp$ calculus, which proves the unsatisfiability of IS^*.



(an initial tree)
(by IE w.r.t. $\tilde{L} \vee E$)
(by IE w.r.t. $\tilde{L} \vee \tilde{E}$)
(by CC)
(by CD)
(by IE w.r.t. $L \vee \tilde{E}$)
(by IE w.r.t. $\tilde{L} \vee \tilde{E}$)
(by CC)
Δ (by CD)

This example also demonstrates that the presence of two inference rules IE and CC is necessary for providing the completeness of the both LC and $LC^\sharp$ calculi.

Additionally note that the LC and $LC^\sharp$ calculi can be considered as specific extensions of the input [3] and SLD [6] resolutions *providing the soundness and completeness of refutation search in the general case.*

4 Literal trees and model elimination method

The technique of refutation search proposed for the literal tree calculi can be easily transformed into the model elimination method considered here in a form distinguished from its original treatment from [2] (see also [3] or [4]). This form denoted by ME demonstrates that the model elimination method can be considered as a method actually operating with "linear writings" of literal trees.

The model elimination method deals with clauses, possibly containing so-called *framed literals* appearing in refutation search. In order to distinguish them from usual (*unframed*) literal, they will be underlined by a ruled line. At that, note that all the literals of any clause from an initial set IS are declared as *unframed literals.*

Remind that the inference rules of the original model elimination method are: a specific resolution rule that can be considered as a special form of the resolution rule extended by the so-called "merging left" rule, reduction rule, and usual factor rule, which, as it is easy to prove, is redundant for providing the completeness of the model elimination method. (Note that this redundancy will automatically follow from the considerations given below.) That is why our treatment of the ME method contains only two inference rules given below.

Model Elimination (ME) resolution. Let $C_1 = C_1' \vee E \vee C_1''$ be a clause from an input set IS and $C_2 = L \vee C_2'$, where L and E are unframed literals and C_1', C_1'', and C_2'' are (possibly, empty) clauses. If there exists the mgu σ of the set $\{E, \tilde{L}\}$, then the clause $C_1'{*}\sigma \vee C_1''{*}\sigma \vee \underline{L{*}\sigma} \vee C_2'{*}\sigma$, called a *MD-resolvent* with $L{*}\sigma$ as a *framed literal*, is said to be deduced by the *model elimination (ME) resolution* from C_1 and C_2 (w.r.t. C_1).

Reduction (RD) rule. Let C be a clause of the form $L \vee C_1' \vee \underline{E} \vee C_2' \vee C_3'$ ($L \vee C_1' \vee C_3' \vee \underline{E} \vee C_2'$), where E is a framed literal,

L is an unframed literal and C'_1, C'_2, and C'_3 are such clauses that for $C'_3 \neq \sharp$ its first literal is unframed and C'_1 and C'_2 do not contain unframed literals (and C'_1 does not contain unframed literals). Suppose that there exists the mgu of $\{L, \tilde{E}\}$. Then $C'_3{*}\sigma$ $(C'_3{*}\sigma \vee \underline{E}{*}\sigma \vee C'_2{*}\sigma)$ is said to be deducible from C by the *reduction rule RD*.

A sequence of clauses $C_1, \ldots, C_m$, containing, maybe, framed literals, where C_1 is a variant of a clause from IS and for $i = 2, \ldots, m$ C_i is the result of the application of one of the ME and RD rules to variants of clauses preceding to C_i, is called an *inference of C_m in the ME method w.r.t. C_1*.

Let us look through all the nodes of a tree Tr "from left to right and from bottom to top". Let $L_1, \ldots, L_k$ be a sequence Q of labels of all its nodes, except the root, constructed by the following way: every label from Tr appears in Q only in the case of the first passing through its node in the above-given order. Then the clause $L_1 \vee \ldots \vee L_k$ is called a *total clause image* of Tr and denoted by $tc(Tr)$.

For the degenerative tree Δ we postulate that $tc(\Delta) = \sharp$.

Theorem 2. *Let IS be an initial set of clauses and C such a clause from IS that the set $IS \backslash \{C\}$ is satisfiable. The set IS is unsatisfiable if and only if in the ME method there exists an inference w.r.t. C of the empty clause $\sharp$.*

Proof. Let IS be an unsatisfiable set. According to Theorem 1, there exists an inference $Tr_1, \ldots, Tr_m$ in $LC^\sharp$ w.r.t. C, in which $Tr_m = \Delta$. Thus, $tc(Tr_1) = C$ and $tc(Tr_m) = \sharp$.

Consider Tr_i $(1 < i < m)$. This literal tree can be deduced from Tr_{i-1} by the $CC^\sharp$ rule or by the $IE^\sharp$ rule w.r.t. $D \in IS$.

It is easy to check that independent from which rule was applied to Tr_{i-1}, the RD rule (for the case of $CC^\sharp$) or the ME rule (for the case of $IE^\sharp$) can be applied to $tc(Tr_{i-1})$ in such a way that the result of this application will coincide with $tc(Tr_i)$. Hence, $tc(Tr_1), \ldots, tc(Tr_m)$ is an inference of $\sharp$ w.r.t. C constructed according to the ME method.

Now, let us suppose that $C_1, \ldots, C_m$ be an inference w.r.t. C $(C_1 = C)$ constructed according to the ME method, in which $C_m = \sharp$.

It is not difficult to construct an algorithm, which, when looking through $C_1, \ldots, C_m$ from left to right, converts each C_i ($i = 1, \ldots, m$) to a literal tree Tr_i, satisfying the following properties:

(i) $lc(Tr_i)$ is the result of deleting in C_i all its framed literals (in particular, we have that at the first step $lc(Tr_1) = C_1 = C$);

(ii) if C_{i+1} is deduced from C_i by the ME rule w.r.t. $D \in IS$ (by the RD rule), then Tr_{i+1} is deduced from Tr_i by the $IE^\sharp$ rule w.r.t. D (by the $RD^\sharp$ rule). Since, by the construction, $Tr_m = \Delta$, we conclude on the basis of Theorem 1 that IS is an unsatisfiable set. □

For the above-given set IS^* we can construct the following inference of $\sharp$ by the ME method, which reestablishes the unsatisfiability of IS^*:

$L \vee E$ (an input clause), $E \vee \underline{L} \vee E$ (by ME w.r.t. $\tilde{L} \vee E$), $\tilde{L} \vee \underline{E} \vee \underline{L} \vee E$ (by ME w.r.t. $\tilde{L} \vee \tilde{E}$), E (by RD), $L \vee \underline{E}$ (by ME w.r.t. $L \vee \tilde{E}$), $\tilde{E} \vee \underline{L} \vee \underline{E}$ (by ME w.r.t. $\tilde{L} \vee \tilde{E}$), $\sharp$ (by RD).

Corollary. The usual model elimination method provides sound and complete refutation search in classical first-order logic.

Proof. It is easy to check that the ME resolution rule and RD rule are special forms of the resolution rule extended by the "merging left" rule and the reduction rule belonging to the usual model elimination method, which provides the truth of the corollary. □

5 Conclusion

This research demonstrates that the consideration of computer-oriented calculi based on tree-like structures can lead not only to the appearance of new methods for inference search in classical first-order logic, but also can shed light on the operation scheme of some of the famous automated theorem-proving methods, such as, for example, the model elimination method. This fact can be explained by the observation that the tree-like approach gives more freedom for the incorporation of separate logical tricks in well-known deductive technique and, therefore, is of an additional interest for research on automated reasoning.

References

[1] J.A. Robinson. *A machine-oriented logic based on resolution principle.* Journal of the ACM, vol. 12, no.1 (1965), pp. 23–41.

[2] D.V. Loveland. *Mechanical theorem proving by model elimination.* Journal of the ACM, vol. 15, no. 2 (1968), pp. 236–251.

[3] C. Chang C. and R. Lee. *Symbolic logic and mechanical theorem proving.* Academic Press Inc., Orlando, FL, USA (1997), 332 pp.

[4] R. Letz, G. Stenz. *Model elimination and connection tableaux procedures.* Handbook of Automated Reasoning (Ed. by A.Robinson and A.Voronkov). Elsevier Science Pub. (2001), pp. 2017–2116.

[5] Bachmair, L., Ganzinger, H. *Resolution theorem proving.* Handbook of Automated Reasoning (Ed. by A.Robinson and A.Voronkov). Elsevier Science Pub. (2001), pp. 19–99.

[6] R. Kowalski and D. Kuehner. *Linear resolution with selection function.* Artificial Intelligence, vol. 2 (1971), pp. 227–260.

Alexander Lyaletski

Institute of Mathematics and Computer Science
Address: 5, Academiei street, Chisinau, Republic of Moldova, MD 2028
E–mails: `forlav@mail.ru`

Some generalisations of Markovsky algorithm on i-invertibile groupoids

Nadeghda Malyutina, Alexandra Scherbacova,
Victor Shcherbacov

Abstract

We propose modification of Markovsky crypto-algorithm [2, 4] based on n-ary groupoids [1] that are invertible on at least one place.

Keywords: left quasigroup, n-ary quasigroup, i-invertibile groupoid, Markovsky algorithm

AMS: 20N15, 05B15, 94A60

We continue researches of applications of n-ary groupoids that are invertible on i-th place in cryptology [5].

It is clear that Markovsky crypto-algorithm which is based on binary or n-ary quasigroup has better "mixing properties" than the proposed algorithm.

But from the other side it is well known [7] that binary (n-ary) quasigroup used in Markovsky algorithm is its key. It is clear that number of i-invertible n-groupoids (number n is fixed) is more than number of n-ary quasigroups (number n is fixed).

Definition. n-Ary groupoid (Q, f) is called invertible on the i-th place, $i \in \overline{1,n}$, if the equation $f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) = a_{n+1}$ has a unique solution for any elements:

$a_1, \dots, a_{i-1}$, $a_{i+1}, \dots,$ $a_n, a_{n+1} \in Q$ [1].

In this case operation ${}^{(i,n+1)}f(a_1, \dots, a_{i-1}, a_{n+1}, a_{i+1}, \dots, a_n) = x_i$ is defined in a unique way and we have:

$$\begin{aligned} f(a_1, \dots, a_{i-1}, {}^{(i,n+1)}f(a_1, \dots, a_{i-1}, a_{n+1}, a_{i+1}, \dots, a_n), \\ a_{i+1}, \dots, a_n) = a_{n+1}, \\ {}^{(i,n+1)}f(a_1, \dots, a_{i-1}, f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n), \\ a_{i+1}, \dots, a_n) = x_i. \end{aligned} \tag{1}$$

A translation of i-invertible n-ary groupoid (Q, f) $(n > 2)$ will be denoted as $T(a_1, \dots, a_{i-1}, -, a_{i+1}, \dots, a_n)$, where $a_i \in Q$ for all $i \in \overline{1, n}$ and

$$T(a_1, \dots, a_{i-1}, -, a_{i+1}, \dots, a_n)x = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$$

for all $x \in Q$.

From the definition of i-invertible n-ary groupoid (Q, f) it follows that any translation of the groupoid (Q, f) is a permutation of the set Q. In the next lemma we suppose that $i = n$. It is clear that next lemma is true for any other value of variable i.

Lemma. If ${}_fT(a_1, \dots, a_{n-1}, -)$ is a translation of an i-invertible n-groupoid (Q, f), then

$${}_fT^{-1}(a_1, \dots, a_{n-1}, -) = {}_{(n,n+1)f}T(a_1, \dots, a_{n-1}, -).$$

Proof. In the proof we omit the symbol f in the notation of translations of the groupoid (Q, f). We have

$$\begin{aligned} & T^{-1}(a_1, \dots, a_{n-1}, -)(T(a_1, \dots, a_{n-1}, -)x) = \\ & T^{-1}(a_1, \dots, a_{n-1}, -)f(a_1, \dots, a_{n-1}, x) = \\ & {}^{(n,n+1)}f(a_1, \dots, a_{n-1}, f(a_1, \dots, a_{n-1}, x)) \overset{(1)}{=} x. \end{aligned} \tag{2}$$

Algorithm. Let Q be a non-empty finite alphabet and k be a natural number, $u_i, v_i \in Q$, $i \in \{1, ..., k\}$. Define an n-ary groupoid (Q, f) which is invertible on n-th place. It is clear that groupoid $(Q, {}^{(n,\,n+1)}f)$ is defined in a unique way.

Take the fixed elements $l_1^{(n^2-n)/2}$ $(l_i \in Q)$, which are called leaders.

Let $u_1u_2...u_k$ be a k-tuple of letters from Q, $a, b, c, d, \dots$ are natural numbers.

$$\begin{aligned}
&v_1 = T^a(l_1, l_2, \dots, l_{n-1}, u_1),\\
&v_2 = T^b(l_n, l_{n+1}, \dots, l_{2n-3}, v_1, u_2),\\
&\dots,\\
&v_{n-1} = T^c(l_{(n^2-n)/2}, v_1, \dots, v_{(n-2)}, u_{n-1}),\\
&v_n = T^d(v_1, \dots, v_{n-1}, u_n),\\
&v_{n+1} = T^e(v_2, \dots, v_n, u_{n+1}),\\
&v_{n+2} = T^t(v_3, \dots, v_{n+1}, u_{n+2}),\\
&\dots
\end{aligned} \tag{3}$$

Therefore we obtain the following ciphertext $v_1v_2 \dots v_k$.

Taking into consideration Lemma we can say that the deciphering algorithm can be constructed similar to the deciphering Algorithm given in [4, 5].

Example. We construct ternary groupoid (R_3, f), $R_3 = \{0, 1, 2\}$, which is defined over the ring $(R_3, +, \cdot)$ of residues modulo 3 and which is invertible on the third place. We define ternary operation f on the set R_3 in the following way: $f(x_1, x_2, x_3) = \alpha x_1 + \beta x_2 + x_3 = x_4$, where $\alpha 0 = 2, \alpha 1 = 2, \alpha 2 = 0$, $\beta 0 = 1$, $\beta 1 = 1$, $\beta 2 = 1$.

Below $T_{2,0}1 = 2$ means that $f(2, 0, 1) = 2$ and so on.

We have:

$T_{0,0}0 = 0$, $T_{0,0}1 = 1$, $T_{0,0}2 = 2$, $T_{0,1}0 = 0$, $T_{0,1}1 = 1$, $T_{0,1}2 = 2$,
$T_{0,2}0 = 0$, $T_{0,2}1 = 1$, $T_{0,2}2 = 2$, $T_{1,0}0 = 0$, $T_{1,0}1 = 1$, $T_{1,0}2 = 2$,
$T_{1,1}0 = 0$, $T_{1,1}1 = 1$, $T_{1,1}2 = 2$, $T_{1,2}0 = 0$, $T_{1,2}1 = 1$, $T_{1,2}2 = 2$,
$T_{2,0}0 = 1$, $T_{2,0}1 = 2$, $T_{2,0}2 = 0$, $T_{2,1}0 = 1$, $T_{2,1}1 = 2$, $T_{2,1}2 = 0$,
$T_{2,2}0 = 1$, $T_{2,2}1 = 2$, $T_{2,2}2 = 0$.

In this case ${}^{(3.4)}f(x_1, x_2, x_4) = x_3 = 2 \cdot \alpha x_1 + 2 \cdot \beta x_2 + x_4$.

Check. $f(x_1, x_2, x_3) = f(x_1, x_2, {}^{(3.4)}f(x_1, x_2, x_4)) = \alpha x_1 + \beta x_2 + 2 \cdot \alpha x_1 + 2 \cdot \beta x_2 + x_4 = x_4$.

$^{(3.4)} f(x_1, x_2, f(x_1, x_2, x_3)) = 2 \cdot \alpha x_1 + 2 \cdot \beta x_2 + \alpha x_1 + \beta x_2 + x_3 = x_3$.
We propose the following elements $l_1 = 2, l_2 = 0, l_3 = 2$ as leader elements. In Algorithm 3 we put $a = 1$, $b = 2$, $c = 1$, $d = 2$ and so on.

In this case open text 2 0 1 1 2 1 is transformed into the following crypto-text:
$T^1_{l_1,l_2} u_1 = f(l_1, l_2, u_1) = f(2, 0, 2) = 0 = v_1$,
$T^2_{l_3,v_1} u_2 = f(2, 0, f(2, 0, 0)) = f(2, 0, 1) = 2 = v_2$,
$T^1_{v_1,v_2} u_3 = f(0, 2, 1) = 1 = v_3$,
$T^2_{v_2,v_3} u_4 = f(2, 1, f(2, 1, 1)) = f(2, 1, 2) = 0 = v_4$,
$T^1_{v_3,v_4} u_5 = f(1, 0, 2) = 2 = v_5$,
$T^2_{v_4,v_5} u_6 = f(0, 2, f(0, 2, 1)) = f(0, 2, 1) = 1 = v_6$.

We obtain the following crypto-text 0 2 1 0 2 1.

We have the following deciphering procedure. Notice that in conditions of this example the following fact is true: $T^{-1}(x, y, -) = T^2(x, y, -)$.
$T^2_{l_1,l_2} v_1 = f(l_1, l_2, f(l_1, l_2, v_1)) = f(2, 0, f(2, 0, 0)) =$
$f(2, 0, 1) = 2 = u_1$,
$T^1_{l_3,v_1} v_2 = f(2, 0, 2) = 0 = u_2$,
$T^2_{v_1,v_2} v_3 = f(0, 2, f(0, 2, 1)) = f(0, 2, 1) = 1 = u_3$,
$T^1_{v_2,v_3} v_4 = f(v_2, v_3, v_4) = f(2, 1, 0) = 1 = u_4$,
$T^2_{v_3,v_4} v_5 = f(v_3, v_4, f(v_3, v_4, v_5)) = f(1, 0, f(1, 0, 2)) =$
$f(1, 0, 2) = 2 = u_5$,
$T^1_{v_4,v_5} v_6 = f(0, 2, 1) = 1 = u_6$.
Therefore we have the following open text: 2 0 1 1 2 1.

References

[1] V.D. Belousov. *n-Ary Quasigroups.* Stiintsa, Kishinev, 1971. (in Russian).

[2] S. Markovski, D. Gligoroski and S. Andova. *Using quasigroups for one-one secure encoding*, Proc. VIII Conf. Logic and Computer Science "LIRA'97", Novi Sad, 1997, pp. 157–167.

[3] V.A. Shcherbacov. *Quasigroup based crypto-algorithms*, 2012. arXiv:1201.3016.

[4] V.A. Shcherbacov. *Elements of Quasigroup Theory and Applications.* CRC Press, Boca Raton, 2017.

[5] V.A. Shcherbacov, N.N. Malyutina. *Role of quasigtoups in cryptosystems. Generalization of Markovsky algorithm.* International Conference on Mathematics, Informatics and Information Technologies dedicated to the Illustrious Scientists Valentin Belousov, April 19 - April 21, 2018, Bălţi, Communications, 2018, pp. 88–89.

[6] M. Vojvoda. *Cryptanalysis of one hash function based on quasigroup.* Tatra Mt. Math. Publ., 29:173–181, 2004. MR2201663 (2006k:94117).

[7] M. Vojvoda. *Stream ciphers and hash functions - analysis of some new design approaches.* PhD thesis, Slovak University of Technology, July, 2004.

Nadeghda Malyutina[1], Alexandra Scerbacova[2], Victor Shcherbacov[3]

[1]Shevchenko Transnistria State University
Email: `231003.bab.nadezhda@mail.ru`

[2]Gubkin Russian State Oil and Gas University
Email: `scerbik33@yandex.ru`

[3]Institute of Mathematics and Computer Science, Moldova
Email: `victor.scerbacov@math.md`

Proceedings of the Conference on Mathematical Foundations of Informatics
MFOI2018, July 2-6, 2018, Chisinau, Republic of Moldova

Non-commutative 6-dimensional associative algebras of two different types

Alexander Moldovyan, Nicolay Moldovyan, Victor Shcherbacov

Abstract

This paper introduces two different types of the non-commutative finite associative algebra of 6-dimensional vectors, which have been proposed as potential carriers of the hidden logarithm problem. The algebra of the first type contains global unit, whereas the algebra of the second type contains only local units. It is also introduced two new forms of the hidden logarithm problem. One of the proposed forms is characterized in using non-invertible elements and assumes using the first-type algebra. The second proposed form is characterized in using locally invertible elements and assumes using the second-type algebra.

Keywords: finite algebra; non-commutative algebra; associative multiplication; hidden logarithm problem

AMS: 16U60, 11G20, 11T71

1 Introduction

Finite non-commutative associative algebras (FNAA) are interesting for applications in the desin of the public-key cryptoschemes characterized in using the hidden logarithm problem (HLP that is also called discrete logarithm problem in hidden commutative subgroup)[2, 1]. In the literature [3, 4, 5] there are considered different FNAA defined over the finite vector spaces with dimensions $m = 2, 3,$ and 4. As regards development of the post-quantum public key cryptoschemes, the main attention was paid to the case $m = 4$ represented by the finite algebra of quaternions defined over the field $GF(p)$ [2, 6]. In the paper [6] it is shown that the HLP in the finite algebra of quaternions can be polynomially reduced to the discrete logarithm problem in $GF(p)$ and using the HLP for designing the post-quantum cryptoschemes requires looking for new carriers of the HLP.

In present paper there are considered two different 6-dimensional finite non-commutative associative algebras (FNAA) possessing different properties. There are also introduced two new forms of defining the HLP. One of the forms is characterized in using non-invertible elements of the FNAA. The second form is characterized in using locally invertible elements of the FNAA.

2 Defining m-dimensional FNAA

Suppose $\mathbf{e}_0, \mathbf{e}_1, \ldots\ \mathbf{e}_{m-1}$ are some formal basis vectors and $v_1, v_2, \ldots v_m \in GF(p)$, where prime $p \geq 3$, are coordinates of the m-dimensional vectors V that are denoted as $v_1\mathbf{e}_0 + v_2\mathbf{e}_1 + \cdots + v_{m-1}\mathbf{e}_{m-1}$ or as $(v_0, v_1, \ldots, v_{m-1})$. Terms $v_i\mathbf{e}_i$, where $i = 0, 1, \ldots, m-1$, are called components of the vector.

Addition of two vectors $A = \sum_{i=0}^{m-1} a_i\mathbf{e}_i$ and $B = \sum_{i=0}^{m-1} b_i\mathbf{e}_i$ is defined as addition of the corresponding coordinates, i.e., with the following formula

$$A + B = \sum_{i=0}^{m-1} (a_i + b_i)\,\mathbf{e}_i,$$

where $+$ denotes the addition operation of both the m-dimensional vectors and the elements of the field $GF(p)$. The multiplication operation in m-dimensional FNAAs (denoted as $\circ$) is defined with the formula

$$A \circ B = \left(\sum_{i=0}^{m-1} a_i\mathbf{e}_i\right) \circ \left(\sum_{i=0}^{m-1} b_i\mathbf{e}_i\right) = \sum_{j=0}^{m-1}\sum_{i=0}^{m-1} (a_ib_j)(\mathbf{e}_i \circ \mathbf{e}_j), \qquad (1)$$

where products of different pairs of formal basis vectors $\mathbf{e}_i \circ \mathbf{e}_j$ are to be replaced by some one-component vector in accordance with some basis vector multiplication table (BVMT), for example, Table 1 or Table 2 in the case of 6-dimensional vector space. It is also assumed that the left basis vector defines the row and the right one defines the column. Thus, the intersection of the ith row and jth column gives the value of the product $\mathbf{e}_i \circ \mathbf{e}_j$ in the form of the single-component vector $\mu\mathbf{e}_i$, where $\mu \in GF(p)$ is called structural coefficient.

For the fixed value m FNAA possessing different properties can be defined with using different BVMTs. For the given value m and fixed distribution of the basis vectors properties of the FNAAs depend on distributions of the structural coefficients and their values.

3 Six-dimensional FNAA containing the global unit element

The BVMT shown as Table 1, where $\mu \in GF(p)$, defines the 6-dimensional FNAA containing the unit element E such that for all 6-dimensional vectors V the following formula holds:

$$VE = EV = V. \tag{2}$$

The unit element E is called global since (2) holds for every element of the considered algebra.

Proposition 1. The global unit E is equal to $(1, 0, 0, 0, 0, 0)$.

Proof of this statement consists in straightforward using the definition of the multiplication operation and Table 1.

Initially the HLP was defined over FNAA with global unit as solving the following equation

$$Y = Q^w \circ G^x \circ Q^{q-w}, \tag{3}$$

where Y, Q, and G are some known vectors such that $Q \circ G \neq G \circ Q$. The vector Y can be used as public key connected with the private key (x, w) in the public key agreement protocol and in the public encryption algorithm [2, 7].

To provide security of the post-quantum cryptoschemes based on the HLP against attacks using homomorphism of the FNAA into the field $GF(p)$ one should select the element G having prime order ω that does not divide the value $p-1$ [5]. Possibility to use the vector G having the order $\omega = p$ is provided by the following proposition.

Proposition 2. An arbitrary vector $V = (1, v_1, v_2, v_3, v_4, v_5)$ has order equal to p, if its coordinates satisfy the following condition:

$$\begin{cases} a_5 = -a_4; \\ a_1^2 + a_2^2 + a_3^2 = -2a_4a_5; \quad \mu\left(a_1a_2 + a_1a_3 + a_2a_3\right) = -a_5^2. \end{cases} \tag{4}$$

Proof. Using formula (1) and Table 1, for the vector V, the coordinates of which satisfy the condition (4), one gets the following: $V^2 = V \circ V = (1, 2v_1, 2v_2, 2v_3, 2v_4, 2v_5)$. Suppose that for some integer $k \geq 3$ the following equality holds:

$$V^k = (1, kv_1, kv_2, kv_3, kv_4, kv_5)\,. \tag{5}$$

Then, using (1) and Table 1, for $V^{k+1} = V^k \circ V$ we get $V^{k+1} = (1, (k+1)v_1, (k+1)v_2, (k+1)v_3, (k+1)v_4, (k+1)v_5)$. Therefore (5) holds for arbitrary value k including the case $k = p$, i. e. $V^p = (1, pv_1, pv_2, pv_3, pv_4, pv_5) = (1, 0, 0, 0, 0, 0)$.

To prevent attacks with using homomorphisms into the field $GF(p)$, a new form of defining the HLP is of interest, that is characterized in using non-invertible elements N of the considered FNAA, which has prime local order λ and satisfies condition $G \circ N \neq G \circ N$. Notion of the local order is connected with local invertibility of some non-invertible vectors N for which there exists the bi-side local unit element E' such that the following formula holds:

$$N \circ E' = E' \circ N = N. \tag{6}$$

The minimum natural number λ such that $N^\lambda = E'$ is called local order of the vector N.

Table 1. The BVMT defining 6-dimensional FNAA with the global unit

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$
$\mathbf{e}_0$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_4$
$\mathbf{e}_1$	$\mathbf{e}_1$	$\mu\mathbf{e}_0$	$\mu\mathbf{e}_4$	$\mu\mathbf{e}_5$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_2$	$\mathbf{e}_2$	$\mu\mathbf{e}_5$	$\mu\mathbf{e}_0$	$\mu\mathbf{e}_4$	$\mathbf{e}_3$	$\mathbf{e}_1$
$\mathbf{e}_3$	$\mathbf{e}_3$	$\mu\mathbf{e}_4$	$\mu\mathbf{e}_5$	$\mu\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$
$\mathbf{e}_4$	$\mathbf{e}_4$	$\mathbf{e}_3$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_5$	$\mathbf{e}_0$
$\mathbf{e}_5$	$\mathbf{e}_5$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_1$	$\mathbf{e}_0$	$\mathbf{e}_4$

4 Six-dimensional FNAA containing only locally invertible elements

The associative non-commutative multiplication operation can be also defined with the BVMT presented as Table 2. The associativity of the multiplication can be easily proved using the formula (1) and considering fulfillment of the following condition for arbitrary three vectors A, B, and $C = \sum_{k=0}^{m-1} c_k\mathbf{e}_k$:

$$(A \circ B) \circ C = A \circ (B \circ C)\,.$$

From the vector equation

$$A \circ X = A,$$

where $X = \sum_{k=0}^{m-1} x_j\mathbf{e}_j$ is the unknown vector (the right unit element), with using Table 2 one can get the following system of six linear equations with the unknown values $x_j \in GF(p),\ j = 0, 1, \ldots, m-1$:

$$\begin{cases} a_0x_0 + \tau\mu a_3x_1 + a_0x_2 + \mu a_3x_3 + \tau a_0x_4 + \mu a_3x_5 = a_0; \\ a_1x_0 + \tau a_4x_1 + a_1x_2 + a_4x_3 + \tau a_1x_4 + a_4x_5 = a_1; \\ a_2x_0 + \tau\mu a_5x_1 + a_2x_2 + \mu a_5x_3 + \tau a_2x_4 + \mu a_5x_5 = a_2; \\ a_3x_0 + \tau a_0x_1 + a_3x_2 + a_0x_3 + \tau a_3x_4 + a_0x_5 = a_3; \\ a_4x_0 + \tau\mu a_1x_1 + a_4x_2 + \mu a_1x_3 + \tau a_4x_4 + \mu a_1x_5 = a_4; \\ a_5x_0 + \tau a_2x_1 + a_5x_2 + a_2x_3 + \tau a_5x_4 + a_2x_5 = a_5. \end{cases} \tag{7}$$

Table 2. The basis-vector multiplication table defining the 6-dimensional FNAA with local invertibility of its elements

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$
$\mathbf{e}_0$	$\mathbf{e}_0$	$\tau\mathbf{e}_3$	$\mathbf{e}_0$	$\mathbf{e}_3$	$\tau\mathbf{e}_0$	$\mathbf{e}_3$
$\mathbf{e}_1$	$\mathbf{e}_1$	$\tau\mu\mathbf{e}_4$	$\mathbf{e}_1$	$\mu\mathbf{e}_4$	$\tau\mathbf{e}_1$	$\mu\mathbf{e}_4$
$\mathbf{e}_2$	$\mathbf{e}_2$	$\tau\mathbf{e}_5$	$\mathbf{e}_2$	$\mathbf{e}_5$	$\tau\mathbf{e}_2$	$\mathbf{e}_5$
$\mathbf{e}_3$	$\mathbf{e}_3$	$\tau\mu\mathbf{e}_0$	$\mathbf{e}_3$	$\mu\mathbf{e}_0$	$\tau\mathbf{e}_3$	$\mu\mathbf{e}_0$
$\mathbf{e}_4$	$\mathbf{e}_4$	$\tau\mathbf{e}_1$	$\mathbf{e}_4$	$\mathbf{e}_1$	$\tau\mathbf{e}_4$	$\mathbf{e}_1$
$\mathbf{e}_5$	$\mathbf{e}_5$	$\tau\mu\mathbf{e}_2$	$\mathbf{e}_5$	$\mu\mathbf{e}_2$	$\tau\mathbf{e}_5$	$\mu\mathbf{e}_2$

The system of equations (7) can be represented in the following form:

$$\begin{cases} a_0\left(x_0 + x_2 + \tau x_4\right) + \mu a_3\left(\tau x_1 + x_3 + x_5\right) = a_0; \\ a_1\left(x_0 + x_2 + \tau x_4\right) + a_4\left(\tau x_1 + x_3 + x_5\right) = a_1; \\ a_2\left(x_0 + x_2 + \tau x_4\right) + \mu a_5\left(\tau x_1 + x_3 + x_5\right) = a_2; \\ a_3\left(x_0 + x_2 + \tau x_4\right) + a_0\left(\tau x_1 + x_3 + x_5\right) = a_3; \\ a_4\left(x_0 + x_2 + \tau x_4\right) + \mu a_1\left(\tau x_1 + x_3 + x_5\right) = a_4; \\ a_5\left(x_0 + x_2 + \tau x_4\right) + a_2\left(\tau x_1 + x_3 + x_5\right) = a_5. \end{cases} \tag{8}$$

It is easy to see the solutions of the last system satisfy the following two equations:

$$\begin{cases} x_0 + x_2 + \tau x_4 = 1; \\ \tau x_1 + x_3 + x_5 = 0. \end{cases} \tag{9}$$

From (9) one can write the following formula describing the set of local right-side units E_r relating to vector A:

$$E_r = \left(i,\ k,\ j,\ h, \frac{1-i-j}{\tau}, \quad -\tau k - h \right) \tag{10}$$

To get the formula for the left-side units corresponding to vector A one should consider the following vector equation:

$$X \circ A = A$$

that can be rewritten in the form of the following system of six linear equations with the unknowns $x_0, x_1, x_2, x_3, x_4, x_5$:

$$\begin{cases} \Phi x_0 + \mu \Psi x_3 = a_0; \\ \Phi x_1 + \Psi x_4 = a_1; \\ \Phi x_2 + \mu \Psi x_5 = a_2; \\ \Psi x_0 + \Phi x_3 = a_3; \\ \mu \Psi x_1 + \Phi x_4 = a_4; \\ \Psi x_2 + \Phi x_5 = a_5, \end{cases} \tag{11}$$

where $\Psi = a_0 + a_2 + \tau a_4$ and $\Psi = \tau a_1 + a_3 + a_5$.

There exists the single solution of the system (11), which defines the following formula for the left-side local unit corresponding to vector A:

$$\begin{aligned} E_l = \Bigg(& \frac{\Phi a_0 - \mu \Psi a_3}{\Phi^2 - \mu \Psi^2}, \frac{\Phi a_1 - \Psi a_4}{\Phi^2 - \mu \Psi^2}, \frac{\Phi a_2 - \mu \Psi a_5}{\Phi^2 - \mu \Psi^2}, \\ & \frac{\Phi a_3 - \Psi a_0}{\Phi^2 - \mu \Psi^2}, \frac{\Phi a_4 - \mu \Psi a_1}{\Phi^2 - \mu \Psi^2}, \frac{\Phi a_5 - \Psi a_2}{\Phi^2 - \mu \Psi^2} \Bigg). \end{aligned} \tag{12}$$

It is easy to see that the value E_l is included in the set (6). Thus, to the vector A such that

$$(a_0 + a_2 + \tau a_4)^2 \neq \mu \left(\tau a_1 + a_3 + a_5 \right)^2 \tag{13}$$

there corresponds the single bi-side local unit, i.e., every of such vectors is locally invertible. From (13) one can conclude that selecting a quadratic non-residue modulo p as the structural coefficient μ one can define full local invertibility in the considered FNAA, i.e. every element of this algebra, except zero $(0, 0, 0, 0, 0, 0)$, will be locally invertible.

Let us consider some locally invertible vector A satisfying the condition (13) and the sequence $A, A^2, ..., A^i$ (for $i = 1, 2, 3, ...$). It is easy to show that this sequence is periodic and for some integer ω the equality $V^\omega = E'$ holds, where E' is a bi-side local unit such that $V^i \circ E' = E' \circ V^i = V^i$ holds for all values i. Such value ω can be called local order of the vector A. Evidently for the given vector A the value E' can be also computed using the formula (12).

To define public-key agreement cryptoscheme based on computations in the FNAA containing only locally invertible vectors one can select vectors G and N satisfying the condition $G \circ N \neq N \circ G$ and then compute the vector $G_{E'} = G \circ E'$, where E' is the bi-side local unit relating to the vector G. Using the vectors N and G_E the public key Y is defined by the following formula:

$$Y = N^{\omega - t} \circ G_{E'}^{x} \circ N^{t}, \qquad (14)$$

where ω is the local order of the vector N; the pair of integer numbers (t, x) represents the private key (the integers $t < \omega$ and $x < g$ are to be selected at random). Finding the values (t, x) from equation (14) represents a novel form of the HLP that provides possibility to design public-key cryptoschemes based on computations in FNAAs without the global unit.

5 Conclusion

It has been introduced two different BVMT for defining 6-dimensional FNAAs: i) with and ii) without the global unit element. The first-type FNAA is interesting for designing post-quantum cryptoschemes based on the HLP defined in standard form. For the case of such FNAAs a new form of the HLP is also proposed, which is characterized in using non-invertible elements of the algebra.

To design post-quantum cryptoschemes based on computations in the second-type FNAA that contains only locally invertible vectors, it is proposed another new form of the HLP. In the case of the FNAA with the multiplication operation defined by Table 2 it is possible to obtain local invertibility of all non-zero elements.

References

[1] E. Sakalauskas, P. Tvarijonas, A. Raulynaitis. *Key Agreement Protocol (KAP) Using Conjugacy and Discrete Logarithm Problems in Group Representation Level,* Informatica, vol. 18, no. 1, (2007), pp. 115–124.

[2] D.N. Moldovyan. *Non-Commutative Finite Groups as Primitive of Public-Key Cryptoschemes,* Quasigroups and Related Systems, vol. 18, no. 2, (2010), pp. 165–176.

[3] A. A. Moldovyan, N. A. Moldovyan, V. A. Shcherbacov. *Non-commutative finite associative algebras of 2-dimension vectors,* Computer Science Journal of Moldova, vol. 25, no. 3 (2017). p. 344–356.

[4] A.A. Moldovyan, N.A. Moldovyan, V.A. Shcherbacov. *Non-commutative finite rings with several mutually associative multiplication operations*, The Fourth Conference of Mathematical Society of the Republic of Moldova dedicated to the centenary of Vladimir Andrunachievici (1917-1997), June 28 - July 2, 2017, Chisinau, Proceedings CMSM4, 2017, pp. 133–136.

[5] D. N. Moldovyan, N. A. Moldovyan. *Cryptoschemes over hidden conjugacy search problem and attacks using homomorphisms,* Quasigroups Related Systems, vol. 18, no. 2, (2010), pp. 177–186.

[6] A. S. Kuzmin, V. T. Markov, A. A. Mikhalev, A. V. Mikhalev, A. A. Nechaev. *Crypto-graphic Algorithms on Groups and Algebras,*

Journal of Mathematical Sciences, vol. 223, no. 5, (2017), pp. 629–641.

[7] D.N. Moldovyan, N.A. Moldovyan. *A New Hard Problem over Non-Commutative Finite Groups for Cryptographic Protocols.* Springer Verlag LNCS. 2010. vol. 6258. pp. 183194 / 5th Int. Conference on Mathematical Methods, Models, and Architectures for Computer Network Security, MMM-ANCS 2010 Proceedings. St.Petersburg, September 811, 2010.

Alexander Moldovyan[1], Nicolai Moldovyan[1], Victor Shcherbacov[2]

[1]St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences
Email: nmold@mail.ru

[2]Institute of Mathematics and Computer Science, Moldova
Email: victor.scerbacov@math.md

Expert Systems and Semantic Information

Alexei Y. Muravitsky

Abstract

We consider a shell for expert systems based on the notion of a recursively represented domain. According to this approach, the finite elements of the domain are the states of knowledge that a computer can have. The idea of approximation of fragments of information is realized through the structure of the domain. Possible knowledge transformers are considered as Scott-continuous functions recursively defined on the finite elements of the domain.

Keywords: report, semanticization, recursively presented finitary domain, Scott-continuous functions, recursive knowledge transformer.

1 To the concept of the expert system

What is an expert system? The question is not easy. The quality of a person as an expert has a time variable. You were an expert in the past, say, ten years ago, but you are not one of them today; the opposite is also possible. This change does not happen immediately. Knowledge of experts can increase or decrease in the process of their participation in the development or exploitation of their field of knowledge. At the individual level, if parts of the expert knowledge contradict each other, he tries to find a point of view that satisfies either all, or at least the most important of them. This moment is important for the continuation, since the new point of view of the expert can be considered as an individual development of the expert's knowledge. This is what he does himself when his decision is caused by his intuition or his mood at the moment. We call this procedure an *adjustment*. The choice is clear: either to preserve all the conflicting parts of expert assessments, or to develop a new point of view that would “satisfy” them all, or at least the most important of them. As the reader will see, we prefer the second option. Thus, we are in favor of an adjustment procedure, and not for the collection of all knowledge.

Other transformations of existing knowledge of an imaginary expert arise when he studies new problems from the outside world that attract his attention.

We can distinguish three tasks facing the expert. First of all, the results of new studies or new assessments should be coordinated in some way with the knowledge that the expert had before he received new information. Secondly, the transformations occurring at the individual level reflect the processes that function in the external world and where the expert is the subject to which these processes are applied. The expert should be well trained or have sufficient experience (or both of these qualities) to distinguish between reliable methods of transformation from unreliable ones; say, he should be able to see the lack of scientific grounds in astrology. Third, experts in research areas that are not formalized or formalized to the extent that mathematics is formalized deal with *degrees of truth* (or *degrees of credibility*), and not only with *truth* and *falsehood*. Indeed, we ourselves temporarily become experts when in the doctor's office the nurse asks us to estimate the pain that we have on a scale of 1 to 10.

The situation is different when the expert is an artificial agent (*computer* or *automated system*). It is different, but there are similarities too. There are pros and cons of such a transition. Some pros are:

- the computer is fully under our control;
- the time variable disappears; whether the automated system is a good expert or bad, it will remain so until we get another system;[1]
- working within the framework of a formal language, the com-

[1] We will not discuss machine learning here.

puter can analyze the truth values of very complex sentences and compare truth values of many sentences at the current state of computer knowledge;

- we can see an imaginary picture: our computer is placed in a stream of information where it receives reports from the outside world; reports are estimated by the values of truth; the computer transforms the states of its knowledge in accordance with each report it receives; it also does adjustment;

- in a larger picture, we can imagine the space of states of computer knowledge, so that the current state is a point in it; on this space, knowledge transformers are defined; they change knowledge of the computer from one state to another; control over the automated system is achieved through the implementation of this plan.

There are also cons:

- the computer works only with formal languages;

- knowledge transformers are rigidly defined and there are only a finite number, although this number is unlimited;

- the adjustment procedure is also rigidly defined; thus the flexibility inherent in the human mind is lost;

- the state of computer knowledge at the moment completely depends on the reports that the computer received and processed by that time; it does not depend on self-education, nor on behavioral insight;

- the expert can use reasoning, some elements of which can be based on an intuitive concept of probability.

The pros and cons mentioned above show differences. And, perhaps, the reader can find more differences after reading the subsequent sections. But, nevertheless, there are similarities. Of course, everything depends on how an automated system is organized. We are talking about a platform that will be discussed in the following sections. At this initial stage, we can note the following similarities:

- the automated system and the expert are simply experts, they have their strengths and weaknesses;
- in particular, they can make mistakes based on their adjustment procedures or their arrangements of truth values;
- the state of knowledge of both the computer and the expert depends on the incoming information and its processing;
- the expert not only accumulates knowledge, but also assimilates it; the same, according to our theoretical platform, will do the computer.

In short, our conclusion: the cooperation between an automated system and an expert is not only possible, but also desirable. We will not discuss the format of such cooperation. Instead, in the rest of this section, we will focus on the concepts that we will be dealing with, and which will be defined in the sequel.

Our basic concepts are divided into the following two categories. The first category contains concepts related to the space of states of knowledge, the second has concepts related to knowledge transformers. Both categories share the notion of the degree of truth and that of report based on a formal language. These two, in turn, lead to the problem of *information semanticization*. In our presentation, we follow the order opposite to the one just mentioned. Namely we start with a formal language and truth values.

Before we begin, we warn the reader that when there is a choice between a good theory and common practice, we choose the theory.

2 Reports as formal objects

The main units of our language, *formulas*, or *formal judgments*, will be codes of concrete judgments, not forms of judgments. That is, we use these terms in a utilitarian sense, not related to the philosophy of utilitarianism. Thus, we will not ask the question, whether a formal judgment is (universally) valid; only the truth value of judgment is important.

We employ a sentential language grounded on a denumerable set Var of ***propositional variables*** (also called ***atomic judgments***) and the three logical connectives: $\land$ (conjunction), $\lor$ (disjunction), and $\neg$ (negation). ***Formulas*** are built from variables in the usual way. The metavariables for variables are denoted by $p, q, r, \ldots$ (perhaps with subscripts); the metavariables for formulas are $A, B, C, \ldots$ (with or without subscripts). The parentheses, '(' and ')', are used as punctuation marks. The set of all formulas is denoted by Fm. It is also convenient to have a formula algebra $\mathfrak{F} = (\mathit{Fm}, \land, \lor, \neg)$.

A set of ***truth values***, $\mathfrak{T}$, is a nonempty set which contains an element $\bot$. An arbitrary (unspecified) element of $\mathfrak{T}$ is denoted by τ (with or without subscripts).

Any expression of the form $\tau\!:\!A$ is called a ***report***. Reports are the only source of information for the computer from the outside world. This can go as follows. An expert evaluates a judgment A with a value τ and sends the report $\tau : A$ to the computer. If the expert does not know what to say about A and is still required to send a report, he sends $\bot : A$. Thus, $\bot$ plays a role of the truth value *unknown*. It is also convenient to have the *unknown* among the truth values in case when one report informs the computer that A is true, symbolically $\mathbf{t}\!:\!A$, and another one that A is false, symbolically $\mathbf{f}\!:\!A$. This suggests that it is convenient, though not necessary, when $\mathfrak{T}$ contains, besides $\bot$ (unknown), also the truth values $\mathbf{t}$ (*true*) and $\mathbf{f}$ (*false*).

Now, the question arises: How should the computer react at a report $\tau : A$? Suppose that the computer is in its original state, that is, its memory is empty. If A is an atomic judgment, say, p, then the computer stores in its memory the valuation

$$s(q) := \begin{cases} \tau & \text{if } q = p \\ \bot & \text{otherwise.} \end{cases}$$

If A is a compound formula, it is natural to find all valuations which assign the value τ to A. We call this task the *semanticization of information* that is contained in A. The process of semanticization is clearly the opposite of the one we use in logic, because in logic for a given valuation s, we compute the value $s(A)$, while in semanticization, having received a report $\tau : A$, we search to find all valuations s such that $s(A) = \tau$.

3 Epistemic structure and epistemic states

From what has been said it should be clear that we have to deal with approximations of some kind. For instance, if an expert evaluates the information conveyed in A by a truth value τ_1 and another expert evaluates the same formula by τ_2, the computer has to take its decision, dealing with the two different reports $\tau_1 : A$ and $\tau_2 : A$. Thus, starting with a set $\mathfrak{T}$ of truth values, we will be implementing the idea of *information approximation*, first proposed by Dana Scott in the 1970s; see, e.g., [14]. The usefulness of this idea is twofold. On the one hand, we perceive information coming into the computer as *partial*. On the other hand, working with units of partial information, in order to be in agreement with how these units are arranged by an *approximation relation*, we will be using functions which are agreed with this relation. Scott called such functions continuous, we will call them *Scott-continuous*.

The idea of approximation (due to D. Scott) is based on the following concepts.

Let $\mathcal{P}^* = \langle \mathcal{P} \leqslant \rangle$ be a partially ordered set, or a ***poset*** for short. And let $\mathcal{D} \subseteq \mathcal{P}$ and $x \in \mathcal{P}$. We call x an ***upper bound*** of $\mathcal{D}$, symbolically $\mathcal{D} \leqslant x$, if for any $y \in \mathcal{P}$, $y \leqslant x$. Similarly, a y is a ***lower bound*** of $\mathcal{D}$, symbolically $x \leqslant \mathcal{D}$, if for any $x \in \mathcal{D}$, $y \leqslant x$. In some cases, a set $\mathcal{D}$ can have a greatest lower bound which is denoted by $\bigsqcap \mathcal{D}$, or/and a least upper bound which is denoted by $\bigsqcup \mathcal{D}$; the former is called the ***supremum*** (abbreviated often as **sup**) of $\mathcal{D}$, the latter the ***infimum*** (abbreviated as **inf**) of $\mathcal{D}$.

A set $\mathcal{D} \subseteq \mathcal{P}$ is called ***directed*** if $\mathcal{D} \neq \varnothing$ and for any $x, y \in \mathcal{D}$, there is an element $z \in \mathcal{D}$ such that $x \leqslant z$ and $y \leqslant z$.

A poset $\mathcal{P}^*$ is called ***complete*** (or is a **CPO** for short) if it has a bottom element, $\bot$ and a sup exists for each directed subset of $\mathcal{P}$.[2]

Given a poset $\mathcal{P}^* = \langle \mathcal{P}, \leqslant \rangle$, an element $x \in \mathcal{P}$ is said to be ***finite*** [5], or ***compact*** [7, 8], if for any directed set $\mathcal{D} \subseteq \mathcal{P}$,

$$(\bigsqcup \mathcal{D} \text{ exists and } x \leqslant \bigsqcup \mathcal{D}) \Longrightarrow x \leqslant d, \text{ for some } d \in \mathcal{D}.$$

The set of finite elements of a poset $\mathcal{P}^*$ is denoted by $F(\mathcal{P}^*)$.

Further, a CPO is a ***complete semilattice*** if an inf exists for each nonempty subset.

Proposition 1. *Let $\mathcal{P}^* = \langle \mathcal{P}, \leqslant \rangle$ be a complete semilattice and $x \in \mathcal{P}$ Then the set $\{e \in F(\mathcal{P}^*) : e \leqslant x\}$ is directed.*

Proposition 1 induces the following definition.

A complete semilattice $\mathcal{P}^* = \langle \mathcal{P}, \leqslant \rangle$ is a ***domain*** if for each element $x \in \mathcal{P}$,

$$x = \bigsqcup \{e \in F(\mathcal{P}^*) : e \leqslant x\}; \tag{1}$$

cf. [5, 9].[3]

We call a complete semilattice ***finitary*** if $\langle F(\mathcal{P}^*), \leqslant \rangle$ is closed under all inf's of nonempty finite sets.

An operation $\phi(x_1, \ldots, x_n)$ on a CPO is ***monotone*** if

$$a_1 \leqslant b_1, \ldots, a_n \leqslant b_n \Longrightarrow \phi(a_1, \ldots, a_n) \leqslant \phi(b_1, \ldots, b_n).$$

Given a CPO $\langle \mathcal{P}, \leqslant \rangle$, a partial operation $\phi : \mathcal{P} \to \mathcal{P}$ is ***Scott-continuous at a point*** $x \in \mathcal{P}$ if for any directed set $\{x_i\}_{i \in I} \subseteq \mathcal{P}$ the following holds:

$$x = \bigsqcup \{x_i\}_{i \in I} \Longrightarrow \phi(x) \simeq \bigsqcup \{\phi(x_i)\}_{i \in I}, \tag{2}$$

[2] In [7, 8] a CPO, as defined above, is called a *directed complete poset* (a **dcpo** for short) *with zero*, or a *pointed* **dcpo**.

[3] In our usage of the term *domain*, we follow [6]. This is the sense, in which [1] uses the term *algebraic domain*, while *domain* in our sense is called there *continuous domain*.

where $\simeq$ is Kleene's equality relation for partial functions; about this relation see [11], § 63. An operation is ***Scott-continuous*** if it is Scott-continuous at each point where this operation is defined.

Proposition 2. *If a partial operation ϕ is Scott-continuous in a CPO $\langle \mathcal{P}, \leqslant \rangle$, then ϕ is also monotone in it.*

In the sequel, we use Proposition 2 without reference.

To semanticize reports, we need the following concept.

Definition 1 (epistemic structure)**.** *An epistemic structure* (*abbreviated as* ES, *in singular or plural*) $\mathfrak{T}$ *is an algebraic system* $\langle \mathfrak{T}, \wedge, \vee, \neg, \bot, \sqsubseteq \rangle$ *of type* $(2, 2, 1, 0, 2)$ *such that the following conditions are satisfied*:

(a) $\langle \mathfrak{T}, \sqsubseteq \rangle$ *is a finite complete semilattice with a least element* $\bot$;
(b) *operations* $\wedge$, $\vee$ *and* $\neg$ *are monotone with respect to* $\sqsubseteq$.

An ES *is called* ***expanded*** *if, in addition, it has two constants* (*or* 0-*ary operations*), $\mathbf{t}$ *and* $\mathbf{f}$, *each different from* $\bot$ *and such that*

$$\text{(c)} \quad \neg\mathbf{t} = \mathbf{f} \text{ and } \neg\mathbf{f} = \mathbf{t}.$$

We say that $\mathfrak{T}$ *has a top element if there is an element* $\top \in \mathfrak{T}$ *such that* $x \sqsubseteq \top$ *for all* $x \in \mathfrak{T}$.

We note that the signature operations of any ES are Scott-continuous; this conclusion can be obtained from [5], section 8.8.

We illustrate the last notion by two well-known examples.

The first example is $\mathfrak{T}_3 = \langle \{\mathbf{t}, \mathbf{f}, \bot\}, \wedge, \vee, \neg, \sqsubseteq \rangle$, where $\bot \sqsubseteq \mathbf{t}$ and $\bot \sqsubseteq \mathbf{f}$, and the signature operations are defined as follows:

x	$\neg x$
$\mathbf{t}$	$\mathbf{f}$
$\mathbf{f}$	$\mathbf{t}$
$\bot$	$\bot$

$x \wedge y$	$\mathbf{t}$	$\mathbf{f}$	$\bot$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{f}$	$\bot$
$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$
$\bot$	$\bot$	$\mathbf{f}$	$\bot$

$x \vee y$	$\mathbf{t}$	$\mathbf{f}$	$\bot$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{t}$	$\mathbf{f}$	$\bot$
$\bot$	$\mathbf{t}$	$\bot$	$\bot$

This is Kleene's 3-valued strong logic.[4]

The other example is Belnap-Dunn's 4-valued logic $\mathfrak{T}_4 = \langle \{\mathbf{t}, \mathbf{f}, \bot, \top\}, \wedge, \vee, \neg, \sqsubseteq \rangle$, where $\bot \sqsubseteq \mathbf{t}, \mathbf{f} \sqsubseteq \top$ and the signature operations are governed by the following tables:

x	$\neg x$
$\mathbf{t}$	$\mathbf{f}$
$\mathbf{f}$	$\mathbf{t}$
$\bot$	$\bot$
$\top$	$\top$

$x \wedge y$	$\mathbf{t}$	$\mathbf{f}$	$\bot$	$\top$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{f}$	$\bot$	$\top$
$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$
$\bot$	$\bot$	$\mathbf{f}$	$\bot$	$\mathbf{f}$
$\top$	$\top$	$\mathbf{f}$	$\mathbf{f}$	$\top$

$x \vee y$	$\mathbf{t}$	$\mathbf{f}$	$\bot$	$\top$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{t}$	$\mathbf{f}$	$\bot$	$\top$
$\bot$	$\mathbf{t}$	$\bot$	$\bot$	$\mathbf{t}$
$\top$	$\mathbf{t}$	$\top$	$\mathbf{t}$	$\top$

(About the last logic see [2], § 81.)

Now, given an ES $\mathfrak{T}$, for an arbitrary report $\tau : A$, the computer searches for all ***valuations*** $s : \mathfrak{F} \longrightarrow \mathfrak{T}$ such that $s(A) = \tau$. (We employ the principle of semantic compositionality (aka 'Frege's Principle') to extend a map $s : \mathit{Var} \longrightarrow \mathfrak{T}$ to the set of all formulas; thus each valuation is a homomorphism.)

It is not necessary (and, in fact, it is impossible for the computer) to use all valuations. Therefore, we define: a valuation s is ***finite*** if the set $\mathrm{V}(s) := \{p : p \in \mathit{Var} \text{ and } s(p) \neq \bot\}$ is finite.

Given valuations s_1 and s_2, we define:

$$s_1 \preceq s_2 \overset{\mathrm{df}}{\Longleftrightarrow} s_1(p) \sqsubseteq s_2(p) \text{ for any } p \in \mathit{Var}. \tag{3}$$

The set of all valuations associated with $\mathfrak{T}$ is denoted by $\mathbf{S}_{\mathfrak{T}}$ and the latter, enriched with $\preceq$, by $\mathbf{S}^*_{\mathfrak{T}}$. We write simply $\mathbf{S}$ and $\mathbf{S}^*$ when the underlying $\mathfrak{T}$ is irrelevant. As two exceptions, we write $\mathbf{S}^*_3$ and $\mathbf{S}^*_4$ instead of $\mathbf{S}^*_{\mathfrak{T}_3}$ and $\mathbf{S}^*_{\mathfrak{T}_4}$, respectively.

Denoting $s^{\bot}(p) = \bot$, for any $p \in \mathit{Var}$, we see that $s^{\bot}$ is the least element of $\mathbf{S}^*$ with respect to $\preceq$. If an underlying $\mathfrak{T}$ has a top element, $\top$, we define $s^{\top}(p) = \top$ for any $p \in \mathit{Var}$. It is clear that $s^{\top}$ is the greatest element of $\mathbf{S}^*$ with respect to $\preceq$.

[4] A historical note about this logic can be found in [12], Footnote 7.

Definition 2 (epistemic state)**.** *An* ***epistemic state*** (*or simply a* ***state***) *is a nonempty set of valuations. A state is* ***finite*** *if it is a finite set of finite valuations. A finite state is* ***minimal*** *if the valuations in it constitute an antichain with respect to* $\preceq$*.*

An arbitrary state is denoted by $\mathcal{E}$ (maybe with subscripts). For any state $\mathcal{E}$, we define:

$$\mathrm{V}(\mathcal{E}) := \bigcup\{\mathrm{V}(s) : s \in \mathcal{E}\}.$$

It is clear that for any finite state $\mathcal{E}$, the set $\mathrm{V}(\mathcal{E})$ is finite. We denote by $\mathbf{F}$ the set of all finite epistemic states.

At this point, it should be clear that the computer can only use finite valuations and therefore only deal with finite states. However, as we shall see, for the definition of “good” transformers of knowledge, we need that the finite states be enclosed in a larger space.

The following proposition should give an idea that domain theory is good for our purposes, for it shows how “ideal/infinite” knowledge can be approximated, say in the sense of (1), by “feasible/finite” knowledge.

Proposition 3 ([12], Proposition 4)**.** *Given an epistemic structure* $\mathfrak{T}$*,* $\mathbf{S}^{*}_{\mathfrak{T}}$ *is a domain with the least element* $s^{\perp}$*, whose finite elements are the finite valuations in* $\mathfrak{T}$*.*

The last proposition only points to domain theory as a possible good tool, but does not in itself provide a good shell for expert systems, since more than one valuation can be obtained from some reports.

4 What is a knowledge space and what is it for?

It would be nice to have something like Proposition 3 for the finite states as a basis for approximating other states of some domain. The main question: What is a proper relation between states? A possible candidate for the answer is the following.

For any states $\mathcal{E}_1$ and $\mathcal{E}_2$,

$$\mathcal{E}_1 \lesssim \mathcal{E}_2 \quad \overset{\text{df}}{\Longleftrightarrow} \quad \text{for any } s_2 \in \mathcal{E}_2, \text{ there is a } s_1 \in \mathcal{E}_1 \text{ such that } s_1 \preceq s_2.^5 \tag{4}$$

Here is the motivation for (4). Assume the computer being at a state $\mathcal{E}$ receives a report $p : \tau$. How should $\mathcal{E}$ be modified? The definition (4) suggests to replace each valuation $s \in \mathcal{E}$ by

$$s'(q) := \begin{cases} s(q) \sqcup \tau & \text{if } q = p \\ s(q) & \text{otherwise,} \end{cases}$$

for any $q \in \mathit{Var}$. Thus we obtain a new state $\mathcal{E}'$ such that $\mathcal{E} \lesssim \mathcal{E}'$.

We note that (4) defines only a preorder. To improve this, given a state $\mathcal{E}$, we define:

$$\downarrow\mathcal{E} := \{\mathcal{E}' : \mathcal{E}' \lesssim \mathcal{E}\},$$

the ***principal ideal*** generated by $\mathcal{E}$. A nonempty set $\mathcal{I}$ is called an ***ideal*** if

$$\mathcal{E} \in \mathcal{I} \text{ and } \mathcal{E}' \lesssim \mathcal{E} \text{ imply } \mathcal{E}' \in \mathcal{I}.$$

We denote by $\mathbf{P}$ the set of all ideals and also define $\mathbf{P}^* := \langle \mathbf{P}, \subseteq \rangle$.

In virtue of [9], Theorem 5.1, $\mathbf{P}^*$ is a domain with $F(\mathbf{P}^*)$ equal to all principal ideals generated by the finite states,[6] that is

$$\langle F(\mathbf{P}^*), \subseteq \rangle = \langle \{\downarrow \mathcal{E} : \mathcal{E} \in \mathbf{F}\}, \subseteq \rangle.$$

It seems that we reached our goal, but got very little. Indeed, we can state that the elements of $\mathbf{P}^*$ are approximated in the sense of (1) by the elements of $F(\mathbf{P}^*)$. However, $\mathbf{P}^*$ is too complex and each elements of $F(\mathbf{P}^*)$ can contain more information than the computer has received by this time. Besides, it is unclear how to define adjustment.

Our approach to adjustment will be implemented using the following two concepts.

[5] This definition was proposed by Nuel Belnap; cf. [4, 3] or [2], § 81; see also [9], pp. 653–654.

[6] $\mathbf{P}^*$ is known as the *completion* [15] of the preorder $\langle \mathbf{F}, \lesssim \rangle$ or also as the *upper powerdomain* (or the *Smith powerdomain*) [9, 8] of $\mathbf{S}^*$.

Definition 3 (value $\mathcal{E}(A)$ of A at $\mathcal{E}$)**.** *The value of a formula A at a state $\mathcal{E}$ is defined as follows:*

$$\mathcal{E}(A) := \bigsqcap\{s(A) : s \in \mathcal{E}\}.$$

Definition 4 (operation $\mathbf{m}(\mathcal{E})$)**.** *Given a finite state $\mathcal{E}$, $\mathbf{m}(\mathcal{E})$, the* ***minimal retract*** *of $\mathcal{E}$, is the set of all minimal elements of $\mathcal{E}$. A finite $\mathcal{E}$ is called* ***minimal*** *if $\mathbf{m}(\mathcal{E}) = \mathcal{E}$.*

We denote by $\mathbf{M}$ the set of all minimal states.

Now, we define ***adjustment*** of a finite state $\mathcal{E}$ as $\mathbf{m}(\mathcal{E})$, because

$$\mathcal{E}(A) = \mathbf{m}(\mathcal{E})(A),$$

for any formula A; cf. [12], Proposition 7.

Proposition 4 ([12], (12))**.** *For any finite states $\mathcal{E}_1$ and $\mathcal{E}_2$,*

$$\mathcal{E}_1 \lesssim \mathcal{E}_2 \iff \mathbf{m}(\mathcal{E}_1) \lesssim \mathbf{m}(\mathcal{E}_2).$$

Next we define

$$\mathbf{M}^* := \langle \mathbf{M}, \lesssim \rangle$$

The advantage of $\mathbf{M}^*$ over $\langle \mathbf{F}, \lesssim \rangle$ is that the former is a poset, while the latter is only a preorder. Moreover, we state the following.

Proposition 5 ([12], Proposition 9)**.** *$\langle F(\mathbf{P}^*), \subseteq \rangle$ and $\mathbf{M}^*$ are isomorphic.*

Corollary 5.1. *The domain $\mathbf{P}^*$ is finitary.*

We see that the elements of $\mathbf{M}$ are quite observable and the relation $\lesssim$ is recursive. From a practical point of view, the computer can always work with pairs $\langle \mathbf{m}(\mathcal{E}), \mathcal{E} \rangle$ or even n-tuples (in order not to loose the information received by this time), where the first component $\mathbf{m}(\mathcal{E})$ is used only for approximation.

For some particular epistemic structures, for example, $\mathfrak{T}_3$ and $\mathfrak{T}_4$, the corresponding domains $\mathbf{P}_3^*$ and $\mathbf{P}_4^*$, respectively, allow more transparent descriptions; cf. [10, 13] for $\mathbf{P}_4^*$.

Here is the picture: the knowledge of the computer (in the form of minimal epistemic states) passes from one state to another within the set of the compact elements of $\mathbf{P}^*$. The elements of $\mathbf{P}^*$ can be regarded as the states of ideal knowledge, approximation to which is implemented in the form of (1).

$\mathbf{P}^*$ is required not only for our belief in ideals, but also for the definition of knowledge transformers. Here our choice is in favor of partial Scott-continuous operations in $\mathbf{P}^*$; even among those operations we have to focus only on the operations that are total on $\mathbf{M}^*$ and closed on this set.

5 What are knowledge transformers for and what can they be?

In this section, we focus on computable operations in finitary domains understood, in each particular case, as a state space of computer's knowledge. At the same time, we want to maintain the idea of approximation that, we believe, should be implemented in any expert system. Therefore, we restrict ourselves only to Scott-continuous operations. Thus our operations being computable are to be defined effectively on finite elements of a domain, and, since they are Scott-continuous, they can "continuously" approximate any element of the domain, even an infinite one.

With this in mind, we assume that $\mathcal{D}$ is a finitary domain, the elements of which are arranged by an approximation relation $\leqslant$, and $\mathcal{D}_0$ is the set of the finite elements of $\mathcal{D}$, that is $\mathcal{D}_0 = F(\mathcal{D})$. Also, we assume that an effective enumeration of the elements of $\mathcal{D}_0$ is fixed, with respect to which the restriction of $\leqslant$ to $\mathcal{D}_0$ is decidable. Such a domain we call ***recursively presented***.

Given a domain $\mathcal{D}$ and the set $\mathcal{D}_0$ of its finite elements, we say that a partial operation $\phi : \mathcal{D} \to \mathcal{D}$ is ***coordinated with*** $\mathcal{D}_0$ if ϕ is closed on $\mathcal{D}_0$; that is, for any $e \in \mathcal{D}_0$, if $\phi(e)$ is defined, then $\phi(e) \in \mathcal{D}_0$. We denote the set of all such operations by $\Phi_{\mathcal{D}}$

Further, given an operation $\phi \in \Phi_{\mathcal{D}}$, we define a (***finite***) ***spectrum of*** ϕ as follows:

$$G_\phi := \{(e_1, e_2) \in \mathcal{D}_0 \times \mathcal{D}_0 : e_2 \leqslant \phi(e_1)\}.$$

We observe that, if an operation ϕ is Scott-continuous, it can be recovered from its spectrum.

Proposition 6. *Let $\mathcal{D}$ be a finitary domain and $\mathcal{D}_0 = F(\mathcal{D})$. Also, let $\phi \in \Phi_{\mathcal{D}}$ be monotone on $\Lambda := \{e \in \mathcal{D}_0 : e \leqslant x\}$ and Scott-continuous at x, for some $x \in \mathcal{D}$. Then $\phi(x)$ is defined and*

$$\phi(x) = \bigsqcup \{e' : \exists e \in \mathcal{D}_0.\ (e, e') \in G_\phi,\ e \leqslant x\}. \tag{5}$$

Given a domain $\mathcal{D}$ and $\phi \in \Phi_{\mathcal{D}}$, we call ϕ $\mathcal{D}_0$-***computable*** if it is partially recursive on $\mathcal{D}_0$.

Proposition 7. *Let $\mathcal{D}$ be a recursively presented domain and $\phi \in \Phi_{\mathcal{D}}$ be Scott-continuous. If ϕ is $\mathcal{D}_0$-computable, then the spectrum G_ϕ is recursively enumerable. And if G_ϕ is recursively enumerable, then the relation $e' \leqslant \phi(e)$ on $\mathcal{D}_0$ is recursively enumerable.*

Corollary 7.1. *Let $\mathcal{D} = \mathbf{P}^*$ (Section 4) and $\phi \in \Phi_{\mathcal{D}}$. Then ϕ is $\mathcal{D}_0$-computable if, and only if, G_ϕ is recursively enumerable.*

References

[1] S. Abramsky and A. Jung. *Domain theory.* Handbook of Logic in Computer Science, vol. 3, Handb. Log. Comput. Sci., vol. 3, Oxford Univ. Press, New York, 1994, pp. 1–168.

[2] A. R. Anderson, N. D. Belnap, Jr., and J. M. Dunn. *Entailment. The logic of relevance and necessity.* Vol. 2, Princeton University Press, Princeton, NJ, 1992, With contributions by K. Fine, A. Urquhart et al, Includes a bibliography of entailment by R. G. Wolf.

[3] N. D. Belnap, Jr.. *How a computer should think.* G. Ryle (ed.) Contemporary Aspects of Philosophy, Oriel Press, 1977; reprinted as §81 of [2].

[4] N. D. Belnap. *A useful four-valued logic.* Modern uses of multiple-valued logic (Fifth Internat. Sympos., Indiana Univ., Bloomington, Ind., 1975), Reidel, Dordrecht, 1977, Episteme, Vol. 2, pp. 5–37.

[5] B. A. Davey and H. A. Priestley. *Introduction to lattices and order*, second ed., Cambridge University Press, New York, 2002.

[6] J. M. Dunn and G. M. Hardegree. *Algebraic methods in philosophical logic*, Oxford Logic Guides, vol. 41, The Clarendon Press Oxford University Press, New York, 2001, Oxford Science Publications.

[7] G. Gierz, K. H. Hofmann, K. Keimel, J. D. Lawson, M. Mislove, and D. S. Scott. *Continuous Lattices and Domains.* Encyclopedia of Mathematics and its Applications, vol. 93, Cambridge University Press, Cambridge, 2003.

[8] G. Gierz, K. Heinrich Hofmann, K Keimel, J. D. Lawson, M. W. Mislove, and D. S. Scott. *A compendium of Continuous Lattices*, Springer-Verlag, Berlin, 1980.

[9] C. A. Gunter and D. S. Scott. *Semantic domains.* Handbook of Theoretical Computer Science, Vol. B, Elsevier, Amsterdam, 1990, pp. 633–674.

[10] Y. Kaluzhny and A. Y. Muravitsky. *A knowledge representation based on the Belnap four-valued logic.* J. Appl. Non-Classical Logics, vol. 3, no 2 (1993), pp. 189–203.

[11] S. C. Kleene. *Introduction to metamathematics.*, D. Van Nostrand Co., Inc., New York, N. Y., 1952.

[12] A. Y. Muravitsky. *Domains as models for semantic information.* Fund. Inform. vol. 138, no 1-2 (2015), pp. 207–225.

[13] A. Y. Muravitsky. *Knowledge representation as domain.* J. Appl. Non-Classical Logics, vol. 7, no 3 (1997), pp. 343–364.

[14] D. Scott. *Lattice theory, data types and semantics.* Formal semantics of programming languages (Courant Comput. Sci. Sympos. 2, New York Univ., New York, 1970), Prentice-Hall, Englewood Cliffs, N. J., 1972, Prentice–Hall Series in Automat. Comput., pp. 65–106.

[15] M. B. Smyth. *Effectively given domains.* Theoret. Comput. Sci., vol. 5, no 3 (1977/78), pp. 257–274.

Alexei Muravitsky

[1]Northwestern State University
Email: alexeim@nsula.edu

Towards Representation of Free Logic as Logic of Partial Quasiary Predicates

Mykola Nikitchenko, Stepan Shkilniak

Abstract

Quasiary predicates can be defined as partial predicates over partial states (partial assignments) of variables. Such predicates have flexible arity. They are used to represent conditions in programs and program requirements, therefore logics of quasiary predicates are program-oriented logics. The main semantic idea of free logic is to allow singular terms not to denote an object. Therefore free logic has fewer existential presuppositions than classical logic. Free logic as well as classical predicate logic is based on total n-ary predicates. We demonstrate how different variants of free logic can be represented as logics of quasiary predicates. Properties of such representation are investigated. The obtained results can be useful for program verification and theory of definite descriptions.

Keywords: free logic, first-order logic, partial predicate, quasiary predicate, quasiary logic.

1 Introduction

Classical predicate logic forms a basis for new logics oriented on different applied domains. Such new logics are often obtained by weakening presuppositions of classical logic. Free logic [1, 2, 3] is one of the logics developed in this manner. Contrary to classical logic, free logic allows singular terms (variables, in particular) not to denote an object. This leads to a logic in which some laws of classical logic are violated. There are various variants of free logic determined by semantic treatment of atomic formulas and quantifiers.

Logics of quasiary predicates [4, 5, 6] have their roots in software engineering domain. The main semantic idea of quasiary logic is “double” partiality: partiality of variables assignments and partiality of predicates and functions. Such partiality is quite natural for computer programs for which program states (variables assignments) and programs themselves and their components are partial mappings. Thus, quasiary logics as well as free logic allow terms not to denote an object.

In this paper we demonstrate how free logic can be represented as logic of quasiary predicates. We simplify definitions and treat logics in model-theoretic style. Thus, *logic* L is a triple $(Fr, INT, \models)$, where

– Fr is a set of *formulas* (language of the logic),

– INT is a class of *logic interpretations* (L-interpretations),

– $\models: INT \xrightarrow{t} \mathcal{B}(\mathcal{B}(Fr) \times \mathcal{B}(Fr))$ is a *consequence relation* for sets of formulas ($\mathcal{B}(S)$ denotes the set of all subsets of S).

For any $J \in INT$ we denote $\models (J)$ as ${}_J \models$ and for any $\Phi \in Fr$ its interpretation in J we denote as Φ_J.

A set of formulas Δ is a *logical consequence* of a set of formulas Γ (denoted $\Gamma_L \models \Delta$) if $\Gamma_J \models \Delta$ for all $J \in INT$.

Let $L = (Fr, INT, \models)$ and $L' = (Fr', INT', \models')$ be two logics. We say [7] that L is a *sublogic* of L' if $Fr \subseteq Fr'$, $INT \subseteq INT'$, and for any $J \in INT$ we have that $({}_J \models) \supseteq ({}_J \models')$. Sublogic L is a proper sublogic of L' if $({}_J \models) \supset ({}_J \models')$. We say that L is *consequence isomorphic* with L' if there are bijections $b_{Fr} : Fr \xrightarrow{t} Fr'$ and $b_{INT} : INT \xrightarrow{t} INT'$ such that for any sets of formulas $\Gamma, \Delta \in \mathcal{B}(Fr)$ and L-interpretation J we have that $(\Gamma_J \models \Delta)$ iff $b_{Fr}(\Gamma)\,{}_{b_{INT}(J)} \models' b_{Fr}(\Delta)$, where $b_{Fr}(\Gamma) = \{b_{Fr}(\Phi) | \Phi \in \Gamma\}$ and $b_{Fr}(\Delta) = \{b_{Fr}(\Psi) | \Psi \in \Delta\}$.

In the sequel we consider only pure logics (without function symbols); notations not defined here are treated in the sense of [6].

2 Logic of partial quasiary predicates

To define a logic of partial quasiary predicates L^{QE} we should define its language (based on logic signature), its class of interpretations, and its consequence relation. Here we present a version called extended quasiary logic [6].

2.1 Quasiary logic signature and formulas

Let Ps be a set of *predicate symbols*, V be an infinite set of names (variables). Usually, within V a subset U of *unessential variables* is identified but here we will not go into detail [6]. A tuple $\Sigma^{QE} = (V, CEs(V), Ps)$ is called the language *signature*. Here $CEs(V)$ is the set of composition symbols which consists of symbols (for all parameters $\bar{x}, \bar{v}, x, z$) of disjunction $\vee$, negation $\neg$, renomination $R_{\bar{x}}^{\bar{v}}$, existential quantification $\exists x$, and variable unassignment composition (predicate) εz.

Given Σ^{QE}, we define inductively the language of L^{QE} – *the set of formulas* $Fr(\Sigma^{QE})$. Formulas P ($P \in Ps$) and εz ($z \in V$) are *atomic*; *composite formulas* are of the form $\Phi \vee \Psi$, $\neg\Phi$, $R_{\bar{x}}^{\bar{v}}\Phi$, and $\exists x\Phi$, where Φ and Ψ are formulas.

2.2 Quasiary logic interpretations

Let A be a set called a set of *values*. Given V and A, the class ${}^{V}\!A$ of *partial assignments* (*partial data*, *nominative sets*) is defined as the class of all partial mappings from V to A, thus, ${}^{V}\!A = V \xrightarrow{p} A$. The set of total assignments is denoted A^V.

Note that two cases can be considered: A is nonempty and A is empty. In the first case we obtain noninclusive quasiary logic and in the second case we obtain inclusive quasiary logic. Here we restrict ourselves by considering noniclusive case only.

Let $Pr_A^V = {}^{V}\!A \xrightarrow{p} Bool$ be the set of all *partial quasiary predicates* over ${}^{V}\!A$.

For $p \in Pr_A^V$ the truth and falsity domains of p are respectively $T(p) = \{d \in {}^VA \mid p(d) \downarrow= T\}$ and $F(p) = \{d \in {}^VA \mid p(d) \downarrow= F\}$.

Compositions from the set $CE(V)$ are defined by the following formulas ($p, q \in Pr_A^V$):

– $T(p \vee q) = T(p) \cup T(q)$; $F(p \vee q) = F(p) \cap F(q)$;

– $T(\neg p) = F(p)$; $F(\neg p) = T(p)$;

– $T(R_{\bar{x}}^{\bar{v}}(p)) = \{d \in {}^VA \mid r_{\bar{x}}^{\bar{v}}(d) \in T(p)\}$;
$F(R_{\bar{x}}^{\bar{v}}(p)) = \{d \in {}^VA \mid r_{\bar{x}}^{\bar{v}}(d) \in F(p)\}$;

– $T(\exists xp) = \{d \in {}^VA \mid d\nabla x \mapsto a \in T(p)$ for some $a \in A\}$;
$F(\exists xp) = \{d \in {}^VA \mid d\nabla x \mapsto a \in F(p)$ for all $a \in A\}$;

– $T(\varepsilon z) = \{d \in {}^VA \mid z$ does not have a value in $d\}$;
$F(\varepsilon z) = \{d \in {}^VA \mid z$ has a value in $d\}$.

Here $d\nabla x \mapsto a = [v \mapsto c \in d \mid v \neq x] \cup [x \mapsto a]$.

Please note that definitions of compositions are similar to strong Kleene's connectives and quantifiers.

A pair $AQE(V, A) =< Pr_A^V; CE(V) >$ is called *an extended first-order algebra of partial quasiary predicates.*

Composition symbols have fixed interpretation. We also need interpretation $I_{QE}^{Ps} : Ps \xrightarrow{t} Pr_A^V$ of predicate symbols. An L^{QE}-*interpretation* J^{QE} is a tuple $(\Sigma^{QE}, AQE(V, A), I_{QE}^{Ps})$. A class of such interpretations for various A and I_{QE}^{Ps} is denoted INT^{QE}.

2.3 Quasiary logic consequence relation

Usually, for logics of quasiary predicates an *irrefutability consequence relation* is defined [5, 6]: a set of formulas Δ is a consequence of a set of formulas Γ in an interpretation J^{QE} (denoted $\Gamma_{J^{QE}} \models_{IR} \Delta$), if

$$\bigcap_{\Phi \in \Gamma} T(\Phi_{J^{QE}}) \cap \bigcap_{\Psi \in \Delta} F(\Psi_{J^{QE}}) = \emptyset.$$

Please note that we can characterize L^{QE} as a logic of *double partiality*: its predicates are *partial* predicates over *partial* assignments.

3 Free logic

For free logic L^{FE} we use notations of the previous section.

3.1 Free logic signature and formulas

A *signature* of L^{FE} is $\Sigma^{FE} = (V, FCs(V), Ps, ar_n)$, where $FCs(V) = \{\vee, \neg, \exists x, E!z\}$ and $ar_n : Ps \xrightarrow{t} Nat$ is an arity mapping. Here $E!z$ is variable assignment predicate. (Note, that $E!z = \neg\varepsilon z$.)

A *set of formulas* $Fr(\Sigma^{FE})$ is defined in a usual way.

3.2 Free logic interpretations

There are various forms of formula interpretations (semantics) in free logic [1, 2, 3]: negative semantics, positive semantics, neutral semantics, and supervaluational semantics. In [3] they are informally characterized as follows:

– in negative semantics empty-termed atomic formulas are false,

– in positive semantics some empty-termed atomic formulas not of the form E!t are allowed to be true,

– in ordinary neutral (or nonvalent) semantics all empty-termed atomic formulas not of the form E!t are truth-valueless;

– in supervaluational semantics empty-termed atomic formulas not of the form E!t are true (or false) if all possible ways of assigning referents to those terms agree in making them true (or false).

Analyzing these semantics we can state that free logic tries to resolve the following difficulty: how to combine interpretations of empty-termed atomic formulas (i.e. predicates over partial assignments) with interpretations of such formulas as total n-ary predicates? The first two semantics (negative and positive semantics) extend atomic formulas interpretations to total predicates over all assignments; the neutral semantics accepts partiality of predicates; supervaluational semantics in a simple form accepts predicate partiality and in more complex form introduces an outer domain to represent non-existing things, thus letting predicates to be total.

For all mentioned semantics propositional compositions are defined in a traditional way, but quantifiers may obtain modified definitions. In this paper we consider only those variants of free logic which have the same interpretation of compositions as in quasiary logic.

Formal definitions of such semantics (interpretations J^{FE}) will be given in the next session.

3.3 Free logic consequence relation

Negative and positive semantics of free logics interpret atomic formulas as total predicates. In this case it is natural to consider a *classical (related to always-true predicate) consequence relation*: a set of formulas Δ is classical consequence of a set of formulas Γ in an interpretation J^{FE} (denoted $\Gamma_{J^{FE}}{\models}_{Cl}\ \Delta$), if

$$\bigcup_{\Phi\in\Gamma} F(\Phi_{J^{FE}}) \cup \bigcup_{\Psi\in\Delta} T(\Psi_{J^{FE}}) = A^V.$$

Neutral and supervaluation semantics allow partial predicates as interpretations of atomic formulas. In this case it is natural to consider an *irrefutability consequence relation*.

4 Defining free logic as quasiary logic

It is not possible to represent directly free logic as quasiary logic because of differences in their languages, interpretations, and consequence relations. Therefore for different semantics of free logic we first construct special free quasiary sublogics and then prove their consequence isomorphism with these semantics of free logic. Thus, we construct four variants of free quasiary logics: L_{-}^{FQE} for negative semantics, L_{+}^{FQE} for positive semantics, L_{0}^{FQE} for ordinary neutral semantics, and L_{s}^{FQE} for simple supervaluational semantics. Here simple supervaluational semantics means that we do not introduce a completion set to provide referents for empty terms.

We start with the definition of a *signature* of free quasiary logics which is $\Sigma^{FQE} = (V, FCs(V), Ps, ar_f)$, where $ar_f : Ps \xrightarrow{t} V^*$ is a finite arity mapping such that for any $P \in Ps$ if $ar_f(P) = (v_1, ..., v_n)$, then $v_1, ..., v_n$ are different variables.

Given Σ^{FQE}, we define inductively the language of L^{FQE} – *the set of formulas* $Fr(\Sigma^{FQE})$. Primitive formulas of the form $R^{v_1,...,v_n}_{x_1,...,x_n}P$ are *atomic* ($ar_f(P) = (v_1, ..., v_n)$, $P \in Ps$); also formulas of the form $E!z$ are *atomic*; *composite formulas* are of the form $\Phi \vee \Psi$, $\neg\Phi$, and $\exists x\Phi$, where Φ and Ψ are formulas. Note that renomination is used only in atomic formulas; such formulas in classical and free logics are presented as atomic formulas of the form $P(x_1, ..., x_n)$.

For quasiary logics we define a *standard interpretation* of predicate symbols $I^{Ps}_{FQE} : Ps \xrightarrow{t} Pr^V_A$ in such a way that predicate $I^{Ps}_{FQE}(P)$ is defined on all data in which $v_1, ..., v_n$ have values while other variables from $V \setminus \{v_1, ..., v_n\}$ are unessential for it ($P \in Ps$, $ar_f(P) = (v_1, ..., v_n)$).

A standard L^{FQE}-*interpretation* J^{FQE} is a tuple

$$(\Sigma^{FQE}, AQF(V, A), I^{Ps}_{FQE}).$$

A class of such standard interpretations for various A and I^{Ps}_{FQE} is denoted INT^{FQE}.

Theorem 1. *Logic L^{FQE} is a proper sublogic of logic L^{QE}.*

Note that logic L^{FQE} is a conservative extension (with symbol εz) of logic L^{QF} [7]. The latter logic is consequence isomorphic with classical logic.

To obtain various variants of free quasiary logics we extend the interpretations of predicate symbols taking into account properties of various semantics.

Let $\bar{v} = (v_1, ..., v_n)$. We treat list $\bar{v}$ also as a set $\{v_1, ..., v_n\}$. Now we define different subclasses of partial assignments:

– a set of partial assignments having values for variables from $\bar{v}$ is

$${}^VA^{\bar{v}} = \{d \mid d \in {}^VA, \text{ values of } v_1, ..., v_n \text{ are defined in } d\};$$

– a set of strict partial assignments for variables from $\bar{v}$ is ${}^VA^{-\bar{v}} =$
$= \{d \mid d \in {}^VA, \text{ at least one of } v_1, ..., v_n \text{ does not have a value in } d\}$.

Note that ${}^VA^{-\bar{v}} = {}^VA \setminus {}^VA^{\bar{v}}$.

A predicate $p \in Pr_A^V$ is called *finitely determined* if there is a finite set $\bar{v}$ such that for any $d \in {}^V\!A^{\bar{v}}$ p is defined and its value is equal to the value of p on $d|_{\bar{v}}$ (projection of d on variables from $\bar{v}$). The class of finitely determined predicates is denoted $PrFD_A^V$.

Such predicates (for various A and various definitions of compositions) form a semantic base for L^{FE}.

4.1 Negative quasiary logic

For this logic we should extend a standard interpretation of a predicate symbol P with arity $\bar{v}$ prescribing its values on ${}^V\!A^{-\bar{v}}$ to be false (F). This gives us a free quasiary logic with negative semantics L_{-}^{FQE}.

Negative semantics has some anomalies [3]. We indicate the following one. Consider a formula $x < y \vee x < z$. In negative semantics this formula on partial assignment $[x \mapsto 5, y \mapsto 3]$ (z does not have a value) obtains a value F. But on an extended assignment $[x \mapsto 5, y \mapsto 3, z \mapsto 7]$ it obtains a value T. This contradicts to the principle of equitonicity [4] according to which a formula value on an extended assignment should remain the same.

The main results for negative semantics are the following.

Theorem 2. *Logic L_{-}^{FQE} is a proper sublogic of L^{QE}.*

Theorem 3. *Logic L_{-}^{FQE} is consequence isomorphic with free logic with negative semantics.*

For negative quasiary logic L_{-}^{FQE} a calculus of sequent type can be constructed by methods proposed in [6].

4.2 Positive quasiary logic

For this logic we should extend a standard interpretation of a predicate symbol P with arity $\bar{v}$ prescribing its values on ${}^V\!A^{-\bar{v}}$ to be true (T). This gives us a free quasiary logic with positive semantics L_{+}^{FQE}.

Being dual to negative semantics, positive semantics also has anomalies [3], in particular, equitonicity principle is not valid.

The main results for positive semantics are the following.

Theorem 4. *Logic L_{+}^{FQE} is a proper sublogic of L^{QE}.*

Theorem 5. *Logic L_{+}^{FQE} is consequence isomorphic with free logic with positive semantics.*

For positive quasiary logic L_{+}^{FQE} a calculus of sequent type can be constructed by methods proposed in [6].

4.3 Neutral quasiary logic

For neutral quasiary logic L_{0}^{FQE} no changes in atomic formulas interpretations are made, therefore this logic coincides with logic L^{FQE}.

For these logics a calculus of sequent type was constructed in [4].

4.4 Simple supervaluational quasiary logic

In quasiary logic the idea of supervaluation was realized by construction of quiasiary predicate extension preserving its equitonicity [4]. The obtained logic L_{s}^{FQE} has the same properties as logic L^{FQE}.

Theorem 6. *Logic L_{s}^{FQE} is consequence isomorphic with L^{FQE}.*

A sequent calculus for this logic is the same as for logic L^{FQE}.

5 Conclusion

In this paper we have investigated how to represent free logic as a logic of partial quasiary predicates. We have considered four semantics for free logics: negative, positive, neutral, and simple supervaluational. For each of the formulated semantics we have constructed special quasiary logics representing corresponding free logics and briefly discussed their properties.

In the future we plan to consider other cases of free logics semantics, in particular, inclusive versions.

References

[1] K. Lambert. *Free logic: Selected essays*, Cambridge Univ. Press. ISBN 9780511039195, 2003.

[2] E. Bencivenga. *Free Logics*, in D. Gabbay and F. Guenthner (eds.), Handbook of Philosophical Logic, vol. III: Alternatives to Classical Logic, Dordrecht: D. Reidel (1986), pp. 373–426.

[3] J. Nolt. *Free logic*, in Zalta, Edward N. Stanford Encyclopedia of Philosophy. Available online https://plato.stanford.edu/entries/logic-free/

[4] M. Nikitchenko, S. Shkilniak. *Mathematical logic and theory of algorithms*, Publishing house of Taras Shevchenko National University of Kyiv, Kyiv, 2008, 528 p. (in Ukrainian).

[5] M. Nikitchenko, V. Tymofieiev. *Satisfiability in composition-nominative logics*, Central European Journal of Computer Science, vol. 2, no. 3 (2012), pp. 194–213.

[6] M. Nikitchenko, S. Shkilniak. *Algebras and Logics of Partial Quasiary Predicates*, Algebra and Discrete Mathematics, vol. 23, no. 2 (2017), pp. 263–278.

[7] M. Nikitchenko, S. Shkilniak. *Towards Representation of Classical Logic as Logic of Partial Quasiary Predicates*, in Proceedings MFOI2017, 2017, Chisinau, Moldova (2017), pp. 133–138.

Mykola Nikitchenko[1], Stepan Shkilniak[2]

[1]Taras Shevchenko National University of Kyiv, Ukraine
Email: `nikitchenko@unicyb.kiev.ua`

[2]Taras Shevchenko National University of Kyiv, Ukraine
Email: `sssh@unicyb.kiev.ua`

The challenges of shadow information economics

Serghei Ohrimenco, Grigori Borta

Abstract

The technological breakthrough of the beginning of the 21st century led to a profound transformation of all the activity domains of the society and the state. The emergence and active development of information and communication technologies marked the beginning of the formation of an information society, meaning the transition from an industrial to a service economy, where theoretical knowledge, technology and information become a commodity of mass consumption.

Keywords: information security, shadow economics, information economics.

1. Introduction

The formation of the digital (informational) economy is connected with the solution of a set of tasks aimed at the development of artificial intelligence, big data, the Internet of things, telemedicine, blockchain technologies, crypto currencies, virtual and augmented reality, and so on. To our greatest regret, the risks of new technologies are hushed up or leveled, and most often they are simply not brought into the light of the expert community.

In particular, in [1], a list of risks is provided, the following being the major ones of them:

- The threat of unauthorized access;
- Mass and/or personal surveillance;
- Markets controlled by foreign producers;
- Disappearance of old professions and loss of jobs;
- Viruses, spyware, vulnerabilities;

- Legal uncertainty, ethical issues;
- Loss of digital sovereignty and digital colonization.

2. Defining Shadow Information Economics

The authors find it necessary to focus on the phenomenon of the shadow information economy (SIE). The starting point in this direction is the shadow IT, which is described by the following definitions:

- Shadow IT are all the third-party IT solutions, including cloud applications and services beyond the control of the corporate IT department [2];
- Shadow IT is a term used to describe the situation when business units acquire, own and manage IT resources, without the help of the IT department. IT departments consider shadow IT to be ineffective, as well as a source of risk, and see it a part of their task to be an obstacle to its spread [3];
- The term "shadow information systems" (Shadow IS) refers to stand-alone software solutions or extensions of existing solutions that are not developed and controlled by the central IT department [4].

The following definitions were used as the basis for determining the SIE [5, 6, 7]:

- SIE is a specific sphere of economic activity with its inherent structure and a system of economic relations. Specificity is set by illegality, unofficiality, as well as the criminal nature of economic activity and the concealment of income;
- From the economic perspective, SIE is the sector of economic relations covering all types of production and economic activities, which in their direction, content, nature and form contradict the requirements of existing legislation and are carried out contrary to state regulation of the economy and bypassing control over it;
- From a technological point of view, SIE is all the individual and collective activity that is illegal, connected with the design, development, distribution, support and use of ICT components hidden from society.

Thus, SIE is comprised of all the illegal and hidden products and services that use and are based on information and communication technologies. The most important economic elements of this sphere are illegal economic relations, activities related to the production, distribution and use of illegal products and services.

The basis of the SIE is shadow (illegal) entrepreneurial activity, the common features of which are:

- A hidden, latent, secret nature, including all the activity that is not registered by government agencies and is not reflected in the official accounts;
- Coverage of all phases of the process of social reproduction (production, distribution, exchange and consumption);
- The parasitic nature of all processes, including but not limiting to the disclosure of software source code, monetization of leasing botnets, etc.

It is necessary to highlight the features that are characteristic of the information field of the shadow economy. Among them are the following:

1. The risk of being found out and punished for a crime committed in the sphere of the shadow digital economy is minimal in comparison with the "classical" shadow economy.
2. The initial threshold of entry is low both in terms of material and time costs. To get started, one only needs to have a computer with Internet access. Moreover, for the initial receipt of profit, there is no need for an in-depth understanding of the principles of the operation of both information technology in general and e-commerce in particular. Many tools are easily and freely available. The management interfaces of such tools are intuitive and easy to master. Personal data and credit card data can be purchased without having any technical skills.
3. In the information environment, it is much easier to find a customer or service provider due to the processes of globalization, the Internet and Darknet.
4. Compared with the "classical" money orders, transactions are much faster and more reliable, and can be made anonymously thanks to the crypto-currencies.
5. Information products and services carry fewer risks than selling, for example, weapons and drugs, while the amount of profit can be comparable.
6. Minimal risks associated with liability, including criminal liability.

3. Segmentation of SIE

Another problem that has not been solved to date is the segmentation of SIE and the identification of two areas of activity – products and services. The products in the field of SIE include but do not limit to: specialized software used to hide the traces of the presence of an attacker; worms; viruses; targeted cyberattacks; malicious software designed for extortion; generators of malicious software; software products aimed at automating the processes of committing cyber crimes; Trojans; adware, fraudware; rootkits; packers, etc. It should be noted that the structure of malicious programs of criminal nature is constantly changing.

In turn, a separate category is formed by services, including but not limiting to: search and analysis of software and hardware vulnerabilities; interception of personal data, credit cards, logins and passwords; leasing of malicious software; phishing; pharming; extortion; terrorism; piracy; lease of proxy servers, as well as encryption and hiding of Internet traffic; money laundering through information technology; the creation and lease of botnets; organization of DOS-attacks; spam delivery; carding, etc.

Particular attention should be paid to the analysis of new customized services that are represented in the TOR network to individual and corporate customers. Among them are Cybercrime-as-a-Service, Research-as-a-Service, Crimeware-as-a-Service, Cybercrime Infrastructure-as-a-Service, Hacking-as-a-Service, Rent-a-Hacker.

It should be noted that the abovementioned list of products and services is neither complete nor exhaustive due to constant and dynamic development of the components of information and communication technologies.

To our greatest regret, there is no statistical information characterizing the degree of development and penetration of the shadow digital economy now. In most sources, data from ongoing surveys is cited, which does not reflect the full picture due to constant evolution. In the Republic of Moldova information on committed information crimes is published without due analysis. In 2015, the Center for Combating Computer Crimes of Moldova investigated 43 cases of Internet money transfers using credit card details, 14 cases of unauthorized access to information systems, 6 cases of fraud, 42 cases of lechery using information technology, 14 cases of interception of information and blackmail. According to the statistics for 2003-2015, the most committed crime is the production and forgery of bank payment instruments, the second is the violation of copyright and related rights (accounting for 256 cases in the period). The third is violation of personal privacy (173 cases), violation of the secrecy of correspondence (55), and child pornography (55). Unfortunately, this data does not reflect the full picture. For example, not all banks provide information on attempts to hack their electronic payment systems [8].

According to the Global Forecast [9], the size of the cybersecurity market is expected to grow from 137.85 billion in 2017, to 231.94 billion USD by 2022, and the shortage of jobs in cybersecurity is one of the most serious problems that enterprises will face. The main characteristics are provided in the Table 1.

Table 1. Key evaluation characteristics of the cybersecurity market. Calculated by the authors based on the data provided by https://www.mnmks.com/subscribers/mnm/industry_trends/cyber_security?isguest=true

Segment	**Min**	**Max**	**CAGR**
Software-Defined Perimeter (SDP) Market	992.8 Million in 2016	4,396.1 Million by 2021	34.7%
Cognitive Analytics Market	1.84 Billion in 2017	10.95 Billion by 2022	42.9%
Enterprise AI Market	845.4 Million in 2017	6,141.5 Million by 2022	48.7%
Recommendation Engine Market	801.1 Million in 2017	4414.8 Million by 2022	40.7%
AI in Education Market	537.3 Million in 2018	3,683.5 Million by 2023	47.0%
User and Entity Behavior Analytics Market	131.7 Million in 2016	908.3 Million by 2021	47.1%
Artificial Intelligence in Healthcare Market	667.1 million in 2016	7,988.8 million by 2022	52.68%
Data Science Platform Market	19.58 Billion in 2016	101.37 Billion by 2021	38.9%
Software-Defined Networking and Network Function Virtualization	3.68 Billion in 2017	54.41 Billion by 2022	71.4%

Market			
Software-Defined Wide Area Network (SD-WAN) Market	738.9 Million in 2016	9,066.2 Million by 2021	65.11%
SDN Orchestration Market	214.7 Million in 2017	4,458.5 Million by 2022	83.4%
Network Automation Market	2.32 Billion in 2017	16.89 Billion by 2022	48.7%
Virtualized Evolved Packet Core (vEPC) Market	968.9 million in 2017	7,975.3 million by 2022	52.4%
Low Power Wide Area Network Market	1.01 Billion in 2016	24.46 Billion by 2021,	89.3%
Network Transformation Market	6.01 Billion in 2017	66.86 Billion by 2022	61.9%
Integration Platform as a Service Market	528.0 Million in 2016	2,998.3 Million by 2021	41.5%
Disaster Recovery as a Service Market	2.19 Billion in 2017	12.54 Billion by 2022	41.8%
Personal Cloud Market	12.02 Billion in 2015	80.02 Billion by 2020	46.1%
Cloud/Mobile Backend as a Service (BaaS) Market	1.32 Billion in 2015	28.10 Billion by 2020	84.2%
Integration Platform as a Service Market	528.0 Million in 2016	2,998.3 Million by 2021	41.5%
Blockchain Market	411.5 Million in 2017	7,683.7 Million by 2022	79.6%
Artificial Intelligence as a Service Market	1.52 Billion in 2018	10.88 Billion by 2023	48.2%
Blockchain Government Market	162.0 Million in 2018	3,458.8 Million by 2023	84.5%
Social Customer Relationship Management (CRM) Market	2.22 billion in 2014	17.92 billion in 2019	51.9%
Gamification Market	1.65 Billion in 2015	11.10 Billion by 2020	46.3%
3D Mapping and 3D Modeling Market	1.90 Billion in 2015	16.99 Billion by 2020	55.0%

In authors' opinion, all actions (products and services) within the framework of the SIE are ultimately connected with cybercriminal. Including but not limiting to actions like theft of a credit card and attempt to cash out money, spying, stealing personal data, laundering huge amounts with the help of special information technologies - can serve as a basis for financing terrorism, including cybernetic. The cost of confronting cyber-fraudsters is growing all over the world. The data characterizing the cost of cyber-crimes and cyber espionage, expressed as a percentage of GDP and represented in Figure 1 is very revealing.

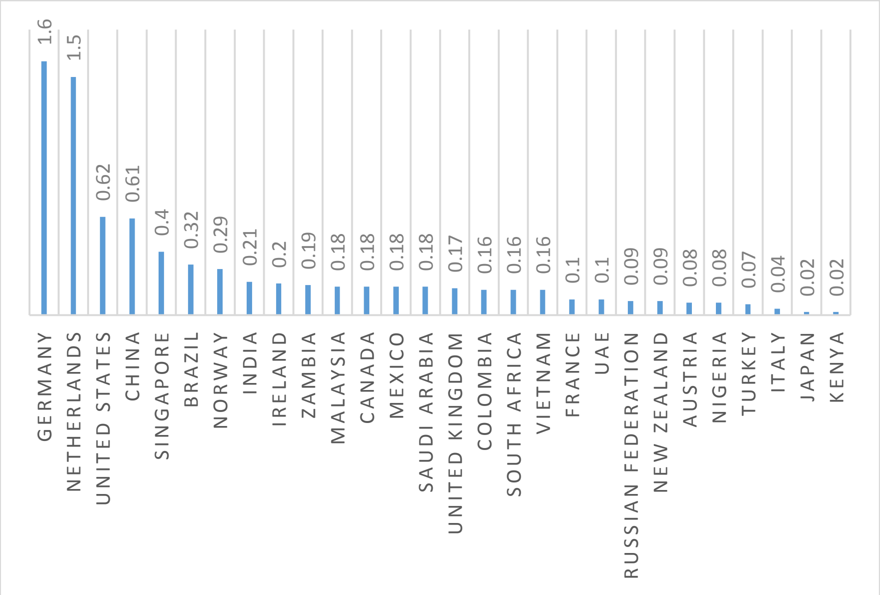


Figure 1. The cost of cybercrime and cyber espionage expressed as percent of GDP [10].

The information economy carries challenges and threats that are directly related to the expansion of the scope of digital technology and its spread to individuals:

- Opportunities for the control of digital services are reduced and opportunities for unlawful actions are increasing;

- Risks of data loss, influence on equipment operation are increased;
- Emergence of new threats associated with the explosive growth of the importance of social networks in the life of society and the development of Internet-thing technology.

The response to these threats is the training of qualified personnel and society in general to develop conscious and legitimate use of the means and methods of the digital economy. The main objectives of training personnel in the field of information security are the following:

- Constant improvement of training of personnel in the field of information security in accordance with the changing demands of the society and professional organizations and the requirements of legislation;
- Differentiate the directions of training in accordance with the nature of the activities of specialists of various profiles;
- Integrated efforts of various organizations and leading specialists in the field of education, information technology and information security;
- Use of various forms of education, involving all levels of society in the process.

We consider it necessary to pay attention to the inadequate research in the following areas: the possibility of implementing threats against medical equipment (in particular, with reference to cardiologic equipment); cryptomania - as a socio-economic phenomenon; Wetware - computer technologies integrated with a biological organism; the concept of "digital twin" (Digital Twin), etc.

5. Conclusion

In conclusion, we consider it possible to propose the development of a holistic strategy to counteract the shadow information economy.

The basic principles of this strategy may be as follows:

- Improvement of the legislative basis of economic regulation, aimed at creating conditions under which the concealment of certain types of activities or their elements, as well as any illegal activity, will become unprofitable;
- Development of cooperation at the state, regional and international levels in order to reduce the level of the shadow information economy;

- Creating jobs, reforming the taxation system, in order to tighten measures to combat money laundering, as well as to increase the fight against corruption;
- To improve the system of training of personnel capable of resisting the phenomena of the shadow information economy;
- Expanding the base of theoretical research and practical developments aimed at new groups of threats, including Internet-related things aimed at medical equipment, cryptomania, etc.;
- It is necessary to exclude the element of spontaneity in the processes of strategy development.

References

[1] Н. Касперская. *Цифровая экономика и риски цифровой колонизации*. Санкт-Петербург, 2018.

[2] Д. Орешкина. *Shadow IT в вашей сети.* http://bis-expert.ru/bdi_source/20/files/assets/basic-html/index.html#32

[3] *Every Employee Is a Digital Employee*. https://blogs.msdn.microsoft.com/jmeier/2015/08/23/every-employee-is-a-digital-employee/

[4] D. Fürstenau, M. Sandner, D. Anapliotis. *Why do Shadow Systems Fail? An Expert Study on Determinants of Discontinuation*. https://www.researchgate.net/publication/303682057Why_Do_Shadow_Systems_Fail_An_Expert_Study_on_Determinants_of_Discontinuation

[5] G. Borta. *The Dark Side of Information Economics. Economica*. An.XXIII, nr2. (92), iunie 2015, ISSN 1810-9136, Academia De Studii Economice A Moldovei, Chisinau, Moldova, p. 97-102.

[6] С. Охрименко, Г. Бортэ. *Обратная Сторона Информационного общества.* Економічна та інформаційна безпека суб'єктів господарювання: сучасний стан і тенденції розвитку: монографія. Авт. кол.: ред. кол.: Т. С. Смовженко, А. Я. Кузнецова, О. І.Барановський, О. М. Тридід, Г. М. Азаренкова та ін., К.: УБС НБУ, 2014, 386 с.

[7] С. Охрименко, А. Саркисян, Г. Бортэ. *Противостояние в информационной сфере*. //Revista militară, №1 (9) 2013, с. 53-61.

[8] В. Волков. *Генпрокуратура Молдовы: Мы предложим руководству страны создать специализированную прокуратуру по борьбе с киберпреступностью.* https://digital.report/genprokuratura-moldovyipredlozhit-rukovodstvu-sozdatspetsprokuraturu-po-borbe-skiberprestupnostyu/

[9] *How to Make More Money as a Cybersecurity Expert*. https://i.medium.com/polyswarm/how-to-make-more-money-as-a-cybersecurity expert.html

[10] Risk Nexus. *Overcome by cyber risks? Economic benefts and costs of alternate cyber futures.* http://publications.atlanticcouncil.org/cyberrisks/risk-nexus-september-2015-overcome-by-cyber-risks.pdf

Serghei Ohrimenco[1], Grigori Borta[2]

[1]Academy of Economic Studies of Moldova/ Laboratory of Information Security
E-mail: osa@ase.md

[2]Academy of Economic Studies of Moldova/ Laboratory of Information Security
E-mail: grigori.borta@gmail.com

Quality of data for mitigation of social disaster effects

Mircea Petic

Abstract

In this paper we refered to three main directions in data processing: the way of processing of the unstructured text data in order to establish the classifiers for emotional messages clustering; the approaches to obtain quality data and crowdsourcing approach to data evaluation.

Keywords: social disasters, emotion processing, quality data, crowdsourcing.

1 Introduction

The appearance of new innovations in the information technology industry (cloud computing, the Internet of Things, and social networking) led to an enormous speed of data amount [1]. Therefore, we analyzed the challenges faced by such situation and try to find the solutions for better ways of data processing [2].

A characteristic feature of contemporary society is the special role of information networks, where different signals related to disasters that may be produced or already have been produced, may occur promptly. A social disaster, as a rule, is followed by information generated in the form of news, discussions and expressions of opinion.

Our research is made within a project whose goal is related to the development of information systems oriented to ensure the security of citizens in extreme situations (natural calamities, technogenic catastrophes, etc.).

In this paper we will refer to three main directions: ways of processing of the unstructured text data in order to establish the classifiers for emotional messages clustering; the approaches to obtain qualitative data and crowdsourcing approach to data evaluation.

2 Classifiers for emotional messages clustering

There are many papers that analyse the topic of social disasters warning and decision making [3]. The main source of information for the means of preventing and mitigating the consequences of social disasters are large volumes of unstructured data accessible to global information networks: mass media, social networks, blogs, and so on [4]. Social media is considered to be a quick information propagation tool for being informed about recent human kind catastrophes [5].

One of the first steps of the research was finding the texts containing any signals about something that has occurred or is about to occur somewhere. Another step is enriching of the existent classifier list by means of internal language mechanisms that can be automatized [4].

In order the processing phase be more rapid we elaborated a Crawler based [6] application service. It searches through web news articles, downloads and extracts the texts of this news, and stores them in the database. As every news site has its own structure we should take into account its particularity.

Then we used a tool that extracts the most frequent words from the unstructured text data. For every of these words, the context, where it is present, is highlighted and rules of inflection and derivation with a high degree of accuracy are applied to generate more semantically related words, thus enriching the set of classifiers [3]. So, the procedure showed how to reduce the number of susceptible words for classifiers and to optimize the processing time.

As the Romanian language belongs to the class of inflectional ones, the process of word forming or derivation of a number of vowel or consonant alternations may occur, generating new stems. For every of these words, the context, where it is present, is highlighted and rules of inflection and derivation with a high degree of accuracy are applied to generate more semantically related words, thus enriching the set of classifiers [3].

3 Quality of data

The developed tools are useful in computational linguistic resources creation, which is of great importance in natural language processing applications. Building both large and good quality text corpora is the challenge we face nowadays [4].

Actual electronic data are very pervasive. That is why the quality of data plays a critical role in the most of all kinds of applications [7]. The need for organizations to identify, measure and manage the quality of their data is becoming increasingly vital as technology becomes more complex and organizations experience expanded information and decision demands from political, social, economic and technical influences [8].

Below we will present five factors that determine the quality of data.

The first one is **data completeness**, that is when we have data different from what we were expected to collect. For example, we extracted on Web that somewhere an earthquake has happened, but we didn't register its place.

The second factor is **data consistency** that refers to what data have been expected. Referring to the same example of the situation that we know that in the Caucasus mountains an earthquake has occurred. At the same time we find in the news reports that it was written that an earthquake has occurred in the Caucaz mountains. In this case the name of the mountain is misspelled and we can say that the data are not consistent.

The third factor, **data accuracy**, refers to the situation when the collected data are correct and exact what is should be. In the situation with the earthquake occurred in the Caucasus mountains, suppose the name is written correct, but it is said that this mountain is situated 2000 km from Erevan and not 200 km that is correct.

The forth factor is **data validity**. Validity of data is determined by whether the data measure that which it is intended to measure. When new information is needed but forms don't get changed, the data is no longer valid because it does not properly measure what it is supposed to [9].

The fifth factor is **data timeliness** that refers to the expectation of when data should be received in order for the information to be used effectively. This factor is needed to prevent bad situation and to ensure the security of citizens in extreme situations (natural calamities, technogenic catastrophes, etc.).

Poor-quality data are often associated with the source of inaccurate reporting and ill-conceived strategies in a variety of situations that appears every day.

4 Crowdsourcing approach to data evaluation

In Web 2.0, it is clear that the Web has made people available as resources to take advantage of. This trend reaches one logical conclusion when the Web helps to find specific people to bring them together and to use their work as a service by means of internet. Crowdsourcing is a strategy that combines the effort of the public to solve one problem or produce one particular thing. "Crowdsourcing" has been used in the popular press to emphasize that the workers need not to be experts but laymen or amateurs.

One example of the crowdsourcing platform is Amazon Mechanical Turk (https://www.mturk.com/). It provides an on-demand, scalable, human workforce to complete jobs that humans can do better than computers, for example, identifying objects in a photo or video, performing data de-duplication, transcribing audio recordings or researching data details. Amazon Mechanical Turk software formalizes job offers to the thousands of Workers (colloquially known as Turkers) willing to do small tasks (called Human Intelligence Tasks or "HITs").

While human subjects can be used to provide data or services in many forms, we limit our attention in this work on data quality monitoring in mitigation of social disaster effects [10].

5 Conclusion

Even though there are many scientific papers in the field of social disasters warning and decision making, however, the field requires more in-depth studies and analysis of a larger volume of computational linguistic resources in this domain.

The compilation of large-scale computational linguistic resources requires a very careful approach, as they are of value only when the quality of these resources is very good.

Verifying the quality of computational linguistic resources is an area that remains closer to human activity than computational processing. This is why the crowdsourcing approach is a solution to get qualitative data from computational linguistic resources for later use in natural language processing applications.

Acknowledgments. 17.80013.5007.01/Ua, bilateral project *Development of a toolkit for modeling strategies to mitigate social disasters caused by catastrophes and terrorism* has supported part of the research for this paper.

References

[1] X. F. Meng, & X. Ci. *Big Data Management: Concepts, Techniques and Challenges.* Journal of Computer Research and Development, vol. 50 (1), 2013, pp. 146–169.

[2] L. Cai, Y. Zhu. *The Challenges of Data Quality and Data Quality Assessment in the Big Data Era.* Data Science Journal, vol. 14: 2, 2015, pp. 1–10, DOI: http://dx.doi.org/10.5334/dsj-2015-002.

[3] Sv. Cojocaru, M. Petic, Gr. Horoş. *Tools for Texts Monitoring and Analysis Aimed at the Field of Social Disasters, Catastrophes, and Terrorism.* In: Computer Science Journal of Moldova, vol.24, no.2 (71), 2016, pp. 157–171.

[4] M. Petic, V. Cozlov. *Determining emotional classifiers for social disasters text clustering.* In: Conference on Mathematical Foundations of Informatics. Proceedings MFOI2017, November 911, 2017, Chisinau, Moldova, pp. 146–149.

[5] N. O. Hodas, et al. *Disentangling the Lexicons of Disaster Response in Twitter.* In: WWW 2015 Companion, Florence, Italy, May 1822, 2015, pp. 1201–1204.

[6] M. Najork, J. L. Wiener. *Breadth-first crawling yields high-quality pages.* In: Proceedings of the Tenth Conference on World Wide Web, Hong Kong, May 2001. Elsevier Science pp. 114–118.

[7] C. Batini, et al. *Methodologies for Data Quality Assessment and Improvement.* In: Journal ACM Computing Surveys (CSUR) Surveys Homepage archive Volume 41 Issue 3, July 2009, Article No. 16.

[8] Palmer Broderick Ch. *An approach for managing data quality.* In: Master Thesis, Faculty of Information Sciences and Engineering, University of Canberra, Australia, November 2011. 155 p.

[9] *5 Factors of High Quality Data & How They Affect Business Decisions*, https://nektardata.com/5-factors-of-high-quality-data/ - visited 01.06.2018.

[10] A. Wang, C. Hoang, M. Kan. *Perspectives on Crowdsourcing Annotations for Natural Language Processing,* Technical report, The National University of Singapore, July, 2010, 24 p.

Mircea Petic [1,2]

[1]Alecu Russo Balti State University, Republic of Moldova
[2]Institute of Mathematics and Computer Science, Chisinau, Republic of Moldova

Email: `petic.mircea@gmail.com, mirsha@math.md`

The projects management system churn prediction

Vladimir Popukaylo

Abstract

Based on the data of the project management system planfix.com, we predict customers' churn after the end of the trial version of the service. The selection of features significantly influencing the target variable has been made. To solve this problem, algorithms are used: logistic regression, decision trees, random forest; the characteristics of mathematical models are compared; preprocessing and model building is done in the programming language of R.

Keywords: data analysis, classification, logistic regression, decision trees, random forest.

1 Introduction

This article solves the problem of churn predicting of the project management system after they have finished using the trial version of the service.

This task is relevant and commercially important for any company representing services on the basis of the subscription mechanism as it is able to afford not only to classify clients, but also, with some probability to predict the subsequent behavior of customers.

The information obtained can be used to make changes in the marketing strategy, in order to increase the percentage of customers who switched to paid service packages.

Churn prediction – the widespread task of machine learning which is solved by many companies based on the analysis of their own data, which does not allow using ready-made software for these purposes.

To solve this problem, language R is used with libraries implementing various stages of data preprocessing, constructing models for classification and visualization of the results obtained.

2 Statement of the research task

The research task is to construct a predictive model that will allow predicting with certain confidence the churn of clients on the basis of their behavior during the free trial period of using the service that does not impose restrictions on the functionality of the project management system.

The experience of data analysis application for churn prediction is described in different researches [1-3]. In practice, various tools are used to solve this problem, among them the Python programming language [4], Statsoft statistical package STATISTICA [5], the Microsoft Azure cloud platform [6], IBM SPSS Statistics and the programming language for statistical data processing R [7].

To carry out this study, the programming language R was chosen, as it has a large number of additional libraries that facilitate both the preliminary analysis of data and the construction of predictive models.

When predicting the churn of clients, it is necessary to solve the classification problem, which in this case is binary. Predictors in the constructed model should be metrics describing the client's behavior on the seventh, fourteenth and twenty-first days after registration, as well as at the end of the free trial period of using the service.

The solution of this task can be divided into several stages:

1. Transformation of database tables containing both quantitative and nominative data and representing time series in a form convenient for further processing.

2. Selection of features that significantly affect the probability of customers switching to paid service packages.

3. Construction of classification models for unbalanced groups.

3 Preprocessing of data

The data for the study is a backup copy of the MySQL database containing information about two tables: ”account_event” and ”metrics_account”. The first table contains 193639 rows with information about the user ID, the date of the event, the type and subtype of the event, and optional additional background information.

The second table contains information about the metrics of the client’s activity and each line contains the date of its creation, a unique user ID, the type and name of the metric, and a numeric field containing the value of the metric. Thus, the table consists of 5 columns and 2785162 rows. The benefits for further research are metrics related to the trial period of using the service and having a type beginning with the word ”trial”.

Thus, the first step in the study was to select rows from the ”metrics_account” table that relate to the first 30 days of using the service. At this stage, 284823 lines were eliminated. For the convenience of further data processing, the columns containing the type and name of the metric were combined and replaced by one column containing the unique identifier of the metric.

The next step was to transform the table in such a way that the resulting metric names became independent columns, and their values filled the cells at the intersection with the corresponding user IDs.

However, when carrying out this transformation, it became clear that for some users there are metrics with the same names, but recorded at different times and storing different values. This circumstance is caused by the incorrect behavior of the function that stores information in the database table. To correct this situation, the data was grouped by a unique user ID and metric name, after which the lines containing the earliest date were selected from them. After the conversion was completed, a table was obtained containing 13895 rows and 133 columns.

The next step was to obtain the target variable, for which the identifiers of users were selected from the ”account_event” table for which there were events showing the transition from a free package to a paid one: ”PlanChangedToPayedFromFree”. On the basis of the data obtained, a binary vector containing the sign ”1” was formed, if the user switched to a paid package and ”0”, if this did not happen.

The last step in data preprocessing was the replacement of missing values in the summary table by zeros, which is consistent with the lack of information on this metric for this customer. Thus, an array of initial data was obtained, suitable for further processing by methods of machine learning.

4 Characteristics selection

At the stage of the primary analysis of the examined characteristics, it was revealed that from further evaluation it is necessary to discard the column containing the user's identifier, as well as eight marks characterizing the users' transitions between free and paid packages.

After that, the remaining 124 metrics were divided into binary and quantitative, based on the contained data. The analysis of binary metrics has shown that 76 indicators, in fact, are not used, that is, they contain only zeros. The remaining 48 indicators were checked for a statistical relationship with the target variable by means of an accurate Fisher test, adjusted for multiple Benjamin-Hochberg comparisons, which allowed to reject the null hypothesis that there was no connection at the significance level of 1% for all traits except the metric, which fixes the account lock for 21 days of use: ”trial_21blockedTrialExpired”.

Conducting a statistical test with a more conservative Bonferoni correction gave similar results. The analysis of the quantitative metrics relationship to the target variable using the Wilcoxon-Mann-Whitney criterion with the Benjamin-Hochberg amendment made it possible to accept the hypothesis that there are no differences in the metric groups that characterize the number of reports created on the 14th and 21st days after the new account was registered by the user:

"trial_14countReportTemplate" and "Trial_21countReportTemplate".

The application of the Bonferonni amendment also allowed us to accept the hypothesis for the metrics that show the number of reports created on day 7 and at the end of the free period.

Thus, it can be concluded that the use of a more conservative Bonferoni correction to test multiple hypotheses, in this case, gives more reliable results, since it allows determining an insignificant factor, without reference to the time of fixing its value by the system.

The next step in selecting the characteristics that affect the target variable was the decision tree. For this, the data table was divided into the training and test part, based on the target attribute in the ratio of 80 to 20. Analysis of the received tree showed that the most significant factors are those that characterize such indicators as: the number of new actions in the last 10 days, the number of new ones tasks for the last 10 days, the total number of tasks, the total number of contacts, the total number of projects and others.

In total, not depth limited tree allowed to select 30 most important features relating to the activity of users in all time intervals and used later to construct predictive models.

5 Construction of predictive models

The first model to predict the churn of users was decided to use the decision tree built in the previous step of the study. The proportion of correctly classified clients on the training sample was 94.1%, and on the test sample - 90.8%. However, considering the fact that the distribution of the target variable is strongly shifted towards the class of clients that did not switch to the paid package, we also need to calculate the F-measure, which on the deferred sample for the class marked as "0" will be 0, 966, and for the class "1" - 0, 523.

As you can see from the received results, in general, the decision tree does a good job of classifying the target group of customers, but it is practically meaningless to predict the probability of the client switching to a paid package based on it. The next step to improve the quality of forecasting was to apply randomizing procedures and build an ensemble of decisive trees. The random forest model, constructed from the factors selected by the decision tree and containing 5000 trees, correctly classifies 98.8% of the clients on the test sample, while the F-measure for each of the classes is 0.993 and 0.924, respectively. However, this accuracy is due to retraining of the model and on the delayed selection of such results it was not possible to achieve: the F-measure for the class "0" coincides with the analogous metric for the tree to the second decimal point and is equal to 0.968, and for the class "1" - 0.586 .

Thus, the total share of correctly classified customers is 94.1%. The third model for forecasting the churn was to use logistic regression. At the first step, the model was built on the parameters selected by the decision tree, after which an iterative procedure was applied, at each step of which the least significant factor was discarded, as long as the information criterion of Akaike decreased.

The result of this approach is the binary classification model, consisting of 20 predictors with quality characteristics that are not inferior to the original one. However, the proportion of properly classified customers using logistic regression is much lower than for decision-based decision trees and is 81.2%, while the F-measure for class "0" is 0.888, and for class "1" 0.414. This quality of prediction can be due to the presence of strongly correlated input variables, which is unacceptable for logistic regression. The results obtained make it possible to conclude that it is inappropriate to use the obtained model to predict the churn of customers.

The results obtained in the previous steps allow us to classify with an acceptable level of accuracy clients who are not ready to switch to paid packages after the trial version, which can allow more reliable planning of the development indicators for the next month. However, it should be noted that the list of factors that most significantly affects the target variable includes indicators that record the quantitative characteristics of the client's activity during the use of the service. At the same time, the developers of the project management system can not practically influence them, in order to reduce the churn of customers, and hence increase the profit of the organization.

In this regard, it was decided to build additionally a model based on metrics describing the use of various service functions at the end of the free trial period of using the service. To solve this problem, only the columns containing the "_enduse" substring in their name were selected from the data array.

Since all eleven received characteristics are binary, it was decided to obtain the probabilities of their influence on the target variable using the regression logistic model. The analysis of the obtained results made it possible to draw the following conclusions: - Coefficients before the factors "Integration with the social network Facebook" and "Using the configuration for billing" are not statistically significant, which does not allow drawing conclusions about their impact on the target variable. - All other coefficients are positive and significant with probability of error of the first kind in 5%.

Here are some characteristics that most strongly affect the probability of switching to paid service packages:

- If the client uses the integration with the service to schedule meetings, events and cases of "Google Calender" or Telegram-bot, his chances of subscribing to the service after a free trial period increase 5-fold compared to the client who does not use these services.

- 2.5 times much are the chances for customers using one of the following services: "Using the SMTP mail transfer protocol", "Integration with the social network VKontakte", "Using the configuration of the accounting of working hours".

- Using the service "Auto-Signature in Mail" increases the chances of switching to a paid package 2.4-fold.

- The rest of the analyzed metrics affect the chances of subscribing no more than 1.5-fold.

Thus, it can be concluded that from integration, the use of which is reflected in the data obtained for research, the greatest impact makes working with the Google calendar and the instant messaging system Telegram.

6 Conclusion

As a part of the research, it was possible to construct predictive models capable to classify with the necessary accuracy clients who will not switch to paid versions of the product after finishing the work with the trial version. The best quality in forecasting was achieved using the random forest algorithm. The application of the current implementation of the model is difficult due to the high computational complexity. The next stage of the research should be to build the most effective model for implementation in the software product. For this it is proposed to minimize the number of decision trees, so that it does not affect the quality of the classification. With the help of the logistic regression model among the parameters characterizing the usage of the functional by the user at the termination of the trial version, the most significant influencing chances of switching to paid versions of the service were selected. The obtained results may affect the direction of the company's marketing strategy in order to reduce the churn of clients of the project management system.

References

[1] A. Canale, N. Lunardon. *Churn prediction in telecommunications industry. A study based on bagging classifiers telecom.* Carlo Alberto Notebooks, 2014. https://www.carloalberto.org/assets/working-papers/no.350.pdf.

[2] A. A. Khan, J. Sanjay, M. M. Sepehri. *Applying data mining to customer churn prediction in an Internet service provider.* Int. J. Comput. Appl., vol. 9(7) (2010), pp. 8–14. http://www.ijcaonline.org/volume9/number7/pxc3871889.pdf.

[3] I. M. Mitkees, S. M. Badr, A. I. B. ElSeddawy. *Customer churn prediction model using data mining techniques.* In 13th International Computer Engineering Conference (ICENCO) (2017), pp. 262–268. doi:10.1109/ICENCO.2017.8289798e.

[4] A.A. Karyakina, A.V. Mel'nikov. *Comparison of methods for predicting the customer churn in Internet service provider companies.* Machine Learning and Data Analysis, vol. 3, no 4 (2017), pp. 250–256, (In Russian). http://jmlda.org/papers/doc/2017/no4/Karyakina2017Churn.pdf.

[5] S.V. Pal'mov. *Analysis and forecasting of the outflow of customers in telecommunication companies based on Data Mining technology.* Abstract of dissertation Ph.D. in Engineering Science. Povolzhskaya gosudarstvennaya akademiya telekommunikatsii i informatiki, Samara, (2005), 16p, (In Russian).

[6] H. Shapiro et. al. *Analyzing Customer Churn by using Azure Machine Learning.* (2017). https://docs.microsoft.com/azure/machine-learning/studio/azure-ml-customer-churn-scenario.

[7] A.V. Gruzdev. *Predictive modeling in IBM SPSS Statistics and R. Method of decision trees.*, 2016. 278p, (In Russian).

Vladimir Popukaylo[1,2]

[1]Institute of Mathematics and Computer Science ASM;
[2] T.G. Shevchenko University
Email: vsp.science@gmail.com

Cyber-Physical Systems: challenges and achievements

Volodymyr G. Skobelev, Volodymyr V. Skobelev

Abstract

In the given paper basic mathematical models and some methods intended for the design and analysis of Cyber-Physical Systems are considered. A semigroup transition system is briefly presented, and some its essential advantages relatively to the timed and time-abstract labeled transition systems designed for a hybrid automaton are outlined. Methods of resolving basic problems of analysis of a weakly initialized hybrid automaton are discussed. It is shown that some of these problems can be resolved by usual methods of Graph Theory. Some method intended to provide the correctness of the customer requirements specifications for the designed hybrid automaton is proposed. Some scheme for resolving two versions of the reachability problem for the designed hybrid automaton is presented.

Keywords: cyber-physical systems, hybrid automata, transition systems, reachability

1 Introduction

Achievements of modern information technologies have led to essential changes in all spheres of mankind activity. In particular, they have catalyzed the 4th Industrial Revolution. The last is based on the broad use of Cyber-Physical Systems (CPS).

Informally speaking, in any CPS some computer networks and built-in controllers (perhaps, with the assistance of the Person) can control considered physical processes by means of the feedbacks, i.e.

the considered physical processes conduct the computations, and the computations, in its turn, conduct the choice and the mode of the considered physical processes.

This integration of physical, computation, networking, and communication processes in any manufacturing forms some strong base for the formation of the Industry 4.0.

Since 2007 the U.S. Government has began to consider the development of CPSs as one of the main strategies [1]. Very quickly intensive research in this area have been extended worldwide. The corresponding Road Maps has been developed in the USA [2, 3] and in Europe [4, 5]. At present similar Road Maps are also developed in Russia Federation and in China.

Currently, different CPSs are used at the research of the Space, in power, military, transport, healthcare, and production spheres, for the design of modern infrastructure, for the remote control of consumer electronics, etc. The state of the art in the area of the research and applications of CPSs is characterized in [6, 7, 8].

Basic mathematical model for the research of any CPS is a hybrid automaton (HA). In what follows some problems and challenges associated with this mathematical model will be considered.

2 Models of HA

Several various definitions of an HA are known. Let us analyze them.

Definition 1. [9, 10]. Any HA is a system

$$\mathsf{H} = (V, E, X, \Sigma, event, Init, Inv, Flow, Jump), \tag{1}$$

where:

(V, E) is the finite directed multigraph (V is the set of control modes, and E is the set of control switches);

X is the finite set of real continuous variables ($|X|$ is the dimension of H, $\dot{x}$ $(x \in X)$ is the first derivative of the continuous variable x, and x' $(x \in X)$ is the update of the continuous variable x at the discrete change from one control mode to another);

Σ is the finite set of events;

$event : E \to \Sigma$ is the function that labels arrows;

$Init$, Inv, and $Flow$ are the functions that label vertices, such that:

$Init(v)$ $(v \in V)$ is the predicate with free variables from the set X, that defines the admissible values for the continuous variables when the HA H starts in v;

Inv $(v \in V)$ is the predicate with free variables from the set X, that defines the constrains on the values of the continuous variables when the continuous evolution of the HA H is in the control mode v;

$Flow(v)$ $(v \in V)$ is the predicate with free variables from the set $X \cup \dot{X}$, that defines the admissible continuous evolutions when the HA H is in the control mode v;

$Jump$ is the function that labels arrows, such that:

$Jump(e)$ $(e \in E)$ is the predicate with free variables from the set $X \cup X'$, that defines when the discrete change modeled by the event e occurs, and what are the updates of the continuous variables, when the HA H makes this discrete change.

With any HA H the following two labeled transition systems can be associated.

To analyze the HA H on the level

the source – the duration of continuous flows – the target

the timed transition system (TTS)

$$\mathcal{S}_{\mathsf{H}}^{(t)} = (S, S_0, A, \to)$$

can be used, where:

$S = V \times \mathbb{R}^n$ is the state space (any subset of S is called a region);

S_0 $(\emptyset \neq S_0 \subseteq S)$ is the initial state space;

$A = \Sigma \cup \mathbb{R}_+$ is the set of the labels of the arcs;

$\to \subseteq S \times A \times S$ is the transition relation.

The denotation $s \to_a s'$ means that $(s, a, s') \in \to$.

It is supposed that in each control mode the continuous evolution is presented via some system of ordinary differential equations (ODEs).

Thus, to compute any transition $s \rightarrow_a s'$ ($a \in \mathbb{R}_+$) in the given control mode some numerical integration method for ODE can be applied.

The most often used is the Euler step forward in time with small steps. It consists that the ODE

$$\dot{x} = f(x) \tag{2}$$

is replaced by the finite-difference equation

$$x(t + \Delta t) = x(t) + f(x(t)) \cdot \Delta t. \tag{3}$$

The transition from the equation (2) to the equation (3) implies that for the analysis of any HA some systems intended for simulation of the discrete control systems, such as the Insertion Simulation System [11], can be used almost without any new update.

Let $w = s_0 a_1 s_1 a_2 \dots$ be any finite or infinite initialized trajectory, i.e. $s_0 \in S_0$ and $s_i \rightarrow_{a_{i+1}} s_{i+1}$ for all $i = 0, 1, \dots$. As usual, the sequence of labels $a_0 a_1 \dots$ is called a trace. The traces are usually used when the behavior of the analyzed HA is characterized in terms of the recognized language.

It is supposed that any finite initialized trajectory is a prefix of some infinite initialized trajectory. This requirement prevents deadlocks in the analyzed HA.

The duration $Drtn(w)$ of the trajectory w is the sum of all its labels $a_i \in \mathbb{R}_+$. An infinite trajectory w is nonZeno (in other words, a divergent trajectory) if $Drtn(w) = \infty$.

The single liveness condition for any HA is the requirement that the set of all infinite trajectories consists of only divergent trajectories. This liveness condition is justified by the fact that for the resolving of any real problem there can be used only those HA, that satisfy this requirement.

The TTS is usually used for the resolving of some safety problems for the investigated HA, such as checking of the emptiness of the set of divergent trajectories, checking of the liveness condition, checking of the validity of the values of the continuous variables in the process of the continuous evolution, and also for the estimation of the time of reachability of this or the other region from the initial state space.

To analyze the HA on the level

the source – the target

the time-abstract transition system (TATS)

$$\mathcal{S}_{\mathsf{H}}^{(a)} = (S, S_0, B, \rightarrow)$$

can be used, where:

S and S_0 are the same as for the TTS;

$B = \Sigma \cup \{\tau\}$ $(\tau \notin \Sigma)$ is the set of the labels of the arcs;

$\rightarrow \subseteq S \times B \times S$ is the transition relation obtained from the transition relation of the TTS $\mathcal{S}_{\mathsf{H}}^{(t)}$ by changing each transition $s \rightarrow_a s'$ $(a \in \mathbb{R}_+)$ by the transition $s \rightarrow_\tau s'$.

The TATS is usually used for the resolving those verification problems for the investigated HA, that can be reduced to the reachability of this or the other region from the initial state space.

Definition 1 provides only some conceptual model of an HA. For algorithmic analysis of the investigated HA, in particular, for the resolving verification and testing problems, the objects of an HA need to be detailed and specified essentially. For example, it is unclear what predicates can be admissible for the investigated HA. As the result, the researcher creates some model which is problematic to be provided in terms of Definition 1.

The following model is much more convenient for algorithmic analysis of any HA.

Definition 2. [12]. Any HA is a system

$$\mathsf{H} = (Q, X, I, D, f, E, G, R), \tag{4}$$

where

Q is the set of discrete states;

$X = \mathbb{R}^n$ is the set of continuous states (more precisely, it must be "$X \subset \mathbb{R}^n$ and X is at least some compact set", since just this condition meets the situations arising via resolving the problems in practice);

$Q \times X$ is the set of states;

$I \subseteq Q \times X$ is the set of initial states;

$D : Q \to \mathfrak{B}(X)$ is the domain (i.e. the set of admissible values of continuous state in the given discrete state);

$f : Q \times X \to \mathbb{R}^n$ is the vector field (more precisely, it must be "$f : \bigcup\limits_{q \in Q} (\{q\} \times D(q)) \to \mathbb{R}^n$ is the family of vector fields");

$E \subseteq Q \times Q$ is the set of arcs;

$G : E \to \mathfrak{B}(X)$ is the guard condition (it defines conditions under which the switching occurs);

$R : E \times X \to \mathfrak{B}(X)$ is the reset map (it defines the transferred values of the continuous state upon the switching).

The model (4) has the following advantages in comparison with the model (1).

All objects in the model (4) are presented in the terms of the sets and mappings. This fact gives the chance to be restricted only with predicates that are presented in terms of the graphs of the mappings. As a result, a wide number of different solvers can be used effectively in the process of resolving the problems of verification for the investigated HA.

Besides, on the basis of a hybrid time set (i.e. a sequence of intervals, such that the right end of the current interval equals to the left end of the next interval), hybrid trajectories can be formally defined and analyzed, i.e. the sequences of alternating continuous and discrete evolutions.

It is essential that the hybrid trajectories with multiple discrete transitions that occur one after another at the same time instant can be presented formally in the model (4). It is evident that on the basis of this construction the mechanisms for setting and resetting of the designed HA into this or the other discrete state can be implemented effectively.

In the process of the analysis of the hybrid trajectories, both TTS and TATS deal with each fragment of the continuous and discrete evolution as with a whole.

For the refinement analysis of the hybrid trajectories it can be applied effectively the following semigroup transition system (STS), that has been presented in [8, 13].

Let $\mathrm{T} \subseteq \mathbb{R}$ be some semigroup of time instants (i.e. if $\mathrm{t}_1, \mathrm{t}_2 \in \mathrm{T}$ then $\mathrm{t}_1 + \mathrm{t}_2 \in \mathrm{T}$), H be some semigroup of traces, S be some set of states, and $\mathrm{G} = \{\mathrm{g_t} \subseteq \mathrm{S} \times \mathrm{H} \times \mathrm{S} | \mathrm{t} \in \mathrm{T}\}$ be some parametric transition relation. We set $\mathrm{s} \xrightarrow{\mathrm{h}}_{\mathrm{t}} \mathrm{s}' \Leftrightarrow (\mathrm{s}, \mathrm{h}, \mathrm{s}') \in \mathrm{g_t}$ and $\mathrm{s} \xrightarrow{\mathrm{h}} \mathrm{s}' \Leftrightarrow (\exists \mathrm{t} \in \mathrm{T})((\mathrm{s}, \mathrm{h}, \mathrm{s}') \in \mathrm{g_t})$.

Any STS is some system

$$\mathbf{S} = (\mathrm{T}, \mathrm{S}, \mathrm{H}, \mathrm{G}),$$

such that the following axioms hold:

$$(\forall \mathrm{s} \in \mathrm{S})(\exists \mathrm{h} \in \mathrm{H})(\exists \mathrm{s}' \in \mathrm{S})(\mathrm{s} \xrightarrow{\mathrm{h}} \mathrm{s}'), \tag{5}$$

$$(\forall \mathrm{t}, \mathrm{t}' \in \mathrm{T})(\forall \mathrm{s}, \mathrm{s}', \mathrm{s}'' \in \mathrm{S})(\forall \mathrm{h}, \mathrm{h}' \in \mathrm{H})(\mathrm{s} \xrightarrow{\mathrm{h}}_{\mathrm{t}} \mathrm{s}' \xrightarrow{\mathrm{h}'}_{\mathrm{t}'} \mathrm{s}'' \Rightarrow$$

$$\Rightarrow \mathrm{s} \xrightarrow{\mathrm{hh}'}_{\mathrm{t}+\mathrm{t}'} \mathrm{s}'), \tag{6}$$

$$(\forall \mathrm{t}, \mathrm{t}' \in \mathrm{T})(\forall \mathrm{s}, \mathrm{s}' \in \mathrm{S})(\forall \mathrm{h} \in \mathrm{H})(\mathrm{s} \xrightarrow{\mathrm{h}}_{\mathrm{t}+\mathrm{t}'} \mathrm{s}' \Rightarrow$$

$$\Rightarrow (\exists \mathrm{h}', \mathrm{h}'' \in \mathrm{H})(\exists \mathrm{s}'' \in \mathrm{S})(\mathrm{s} \xrightarrow{\mathrm{h}'}_{\mathrm{t}} \mathrm{s}'' \xrightarrow{\mathrm{h}''}_{\mathrm{t}'} \mathrm{s}'). \tag{7}$$

It is worth to note that:

formula (5) is the extension axiom, and it means that there are no deadlocks in any STS;

formula (6) is the contraction axiom, and it means that any finite trajectory can be contracted to the single transition;

formula (7) is the deployment axiom, and it means that any sufficiently long transition can be deployed into some trajectory consisting of several transitions.

Evidently that any TTS and TATS can be simulated by some STS. Thus, the functioning of any model (1) or (4) also can be simulated in terms of the appropriate STS.

In the definition of the STS there has been no restrictions on the set S of the states and on the semigroup H of traces. Moreover, any computing on the base of finite-difference equations can be implemented explicitly in STS. Therefore, the use of STS for the design and analysis of the considered CPS has at least the following three advantages.

Firstly, in each discrete state, each continuous dynamics can be analyzed in detail on its successful completion and the absence of the forbidden values for the continuous variables. Besides, the possibility to activate each switching from any discrete state also can be effectively checked.

Secondly, in each discrete state, on the basis of the analysis of the continuous dynamics some external monitoring and error correction facilities for the appropriate subsystem of the considered CPS can be implemented.

Thirdly, any variable-structured CPS can be effectively designed and analyzed by using appropriate STS, perhaps, with some additional labeling of the states of the STS by these or the others attributes.

3 Problems of HA analysis

In what follows the model (4) is considered as a base.

It is not always specified in an explicit form that one of the customer requirements consists of the necessaty to design some weakly initialized HA (H, Q_{in}), where Q_{in} $(\emptyset \neq Q_{in} \subseteq Q)$ is the set of all admissible discrete initial states. This requirement imposes some specifics to the analysis of correctness of the customer requirements specifications (CRS).

Some preliminary checking of correctness for the CRS can be realized on the basis of the analysis of the transition diagram (Q, E) of the investigated HA by usual methods of Graph Theory.

Indeed, let us suppose that each of the guard conditions can be satisfied via the functioning of the investigated HA. Designing for the transition diagram (Q, E) some spanning tree with the root in the fixed initial discrete state $q \in Q_{in}$, we find the set of all discrete states that are reachable from the discrete state q. Thus, designing the set of spanning trees with the root in each fixed initial discrete state $q \in Q_{in}$ we can compute the set of all discrete states reachable from the set Q_{in} of all admissible discrete initial states.

Moreover, the analysis of the set of these rooted spanning trees gives the possibility to establish the lower estimations for the number of the discrete switchings that are necessary for reachability of this or that discrete state from the set Q_{in}.

Analysis of the transition diagram (Q, E) by usual methods of Graph Theory also gives the possibility to find bridges, cut-vertices, cycle basis and cut vectors.

It has been shown in [14] that the detailed analysis of correctness of the CRS for the designed CPS is of special importance owing to the following reasons.

The CRS while specifying the objects of the designed HA can contain both contradictory information, and information that will never be realized via correct functioning of the HA. It is clear, that all contradictions in objects of the designed HA must be removed, since they can lead to unpredictable consequences. Information that will never be realized via correct functioning of the designed HA can complicate HA analysis significantly.

Therefore, the following Problem naturally arises.

Problem 1. It is necessary to carry out the removal of the contradictions in the objects that define the designed HA, and to coordinate these objects with each other.

It is worth to note that complexity of resolving the Problem 1 increases significantly, if there are some controlled parameters in the designed HA, i.e. parameters the values of which can be varied in the process of functioning of the designed CPS. The choice of the best values for these parameters is, in its essence, the problem of multiobjective optimization, and, thus, requires the development of some ad hoc methods for its resolving.

Despite the features that can arise for the specific HA, the resolving of the Problem 1 always includes the necessity to resolve the following two sub-problems.

Problem 1.1. For each discrete state of the designed HA it is necessary to provide transformation of the set of admissible initial values of the continuous variables on the set of the values of the continuous variables defined by the guard conditions for the arcs started from this discrete state, under supposition that for the continuous evolutions in this discrete state all values of the continuous variables are admissible at each instant of time.

Problem 1.2. For each non-initial discrete state of the designed HA it is necessary to provide the mapping of the set of all values of continuous variables defined by the reset maps for the arcs terminated in this discrete state on the set of admissible initial values of continuous variables for this discrete state.

It is evident that in the process of the resolving of the Problems 1.1 and 1.2 the sets of all admissible initial values of continuous variables, the sets of the values of continuous variables defined by the guard conditions, and the sets of the values of continuous variables defined by the reset maps can be changed.

This is the main reason, why the resolving of the Problem 1 includes some iterative process determined by the repeated resolving of the Problems 1.1 and 1.2.

In [14] the Problem 1 has been investigated in detail for the class of 1-dimensional HA, such that different continuous evolutions can take place in any discrete state for different non-intersected closed intervals of the initial values of continuous state, the duration of each continuous evolution is finite (moreover, the least and the upper bounds for the duration of each continuous evolution are known for the researcher), and these durations can be different for different closed intervals of the initial values of the continuous state.

It is worth to note that the supposition that different continuous evolutions can take place in any discrete state for different non-intersected closed intervals of the initial values of continuous variables is the essential generalization of all existing models of HA, since this supposition gives the possibility to join different tightly connected with each other discrete states into the single discrete state.

The HA, designed as the result of resolving the Problem 1 represents the mathematical model intended for verification and testing of the control system of the designed CPS.

Despite the features that can arise for the specific HA, the process of verification of the designed HA is based significantly on the resolving of some versions of the following problem.

Problem 2. For the given weakly initialized HA, under the supposition that the given complex of restrictions hold, it is necessary to characterize the set of all discrete states, reachable from the initial set of discrete states.

One of the most known version of the Problem 2 is the investigation for the fixed discrete state the reachability of the given region from the region determined by some fixed subset of the set of admissible initial values of continuous variables. It is well known that this problem can be resolved only under sufficiently strong restrictions on the structures of the considered regions as well as on the continuous evolution in the given discrete state.

As it has been pointed above, the class of 1-dimensional HA, such that the duration of each continuous evolution is finite, and the least and the upper bounds for the duration of each continuous evolution are known for the researcher, has been investigated in [14].

The existence of finite the least and the upper bounds for duration of each continuous evolution gave the chance to resolve for the analyzed HA the problems of reachability of the discrete states from the set of the initial discrete states, both for the minimal number of switchings, and for the minimal time.

The method of the solution of these problems, proposed in [14], is based on the design and analysis of finite trajectories of the form

$$(q_1, [\alpha_1, \mathrm{A}_1]) \xrightarrow{[\theta_1, \Theta_1]} (q_1, [\beta_1, \mathrm{B}_1]) \xrightarrow{[0,0]} (q_2, [\alpha_2, \mathrm{A}_2]) \xrightarrow{[\theta_2, \Theta_2]} \ldots,$$

where the transition

$$(q_i, [\alpha_i, \mathrm{A}_i]) \xrightarrow{[\theta_i, \Theta_i]} (q_i, [\beta_i, \mathrm{B}_i])$$

means that the continuous evolution in the discrete state q_i with the duration $t \in [\theta_i, \Theta_i]$ transforms the interval $[\alpha_i, \mathrm{A}_i]$ onto the interval $[\beta_i, \mathrm{B}_i]$, and the transition

$$(q_i, [\beta_i, \mathrm{B}_i]) \xrightarrow{[0,0]} (q_{i+1}, [\alpha_{i+1}, \mathrm{A}_{i+1}])$$

means that the switching from the discrete state q_i to the discrete state q_{i+1} occurs and the interval $[\beta_i, \mathrm{B}_i]$ is transformed onto the interval $[\alpha_{i+1}, \mathrm{A}_{i+1}]$.

It is evident that this method can be generalized for much wider classes of HA with given finite the least and the upper bounds for duration of each continuous evolution.

It is necessary to mark especially that a great number of difficulties arise in the process of the design of compositions of HA, since there are a lot of problems connected with synchronization of both the continuous evolutions and discrete switchings of different HA. Some of these problems can be resolved by direct using of timed automata. However, in this case some problems connected with algorithmic solvability can arise.

Unfortunately, there are also some problems that require development of special mechanisms for synchronization.

4 Conclusions

In the present paper basic mathematical models intended for the design and analysis of Cyber-Physical Systems have been presented and analyzed.

In addition to the timed and time-abstract labeled transition systems associated with a hybrid automaton, a semigroup transition system (STS), presented in [8, 13], has been considered, and some its essential advantages relatively to the above pointed transition systems have been outlined.

It has been shown that the STS can be successively used for elaboration of methods intended to provide the correctness of the customer requirements specifications for the designed hybrid automaton.

Detailed analysis of the possibilities to use the STS for an effective resolving basic problems of verification and testing for a hybrid automaton determines one of the trends for future research.

It has been presented some scheme for resolving two versions of the reachability problem for the designed hybrid automaton under the supposition that finite the least and the upper bounds for the duration of each continuous evolution is known.

Detailed analysis of the possibilities to use this scheme effectively for compositions of hybrid automata determines another trend for future research.

References

[1] Leadership Under Change: Information Technology R&D in a Competitive World, 2007.
http://www.nitrd.gov/Pcast/reports/PCAST-NIT-FINAL.pdf

[2] Foundations for Innovation in Cyber-Physical Systems. Workshop Summary Report, 2013.
//www.nist.gov/sites/default/files/documents/el/CPS-WorkshopReport-1-30-13-Final.pdf

[3] Foundations for Innovation. Strategic R & D Opportunities for 21st Century Cyber-Physical Systems, 2016. http://bookprem.com/gd-ebooks/B00U37SUBG

[4] Cyber-Physical European Roadmap & Strategy, 2015. //www.cyphers.eu

[5] sCorPiuS-project.eu: European Roadmap for Cyber-Physical Systems in Manufacturing Deliverable D1.1. State of the Art on Cyber-Physical Systems, 2015. http://www.scorpius.drupal. Pulsartecnalia.com/files/documents/sCorPiuS_D1.1_SotA_v1.2.pdf

[6] V. Gunes, S. Peter, T. Givargis, and F. Vahid. *A Survey on Concepts, Applications, and Challenges in Cyber-Physical Systems.* KSII Transactions on Internet and Information Systems, Vol. 8, No. 12 (2014), pp. 4242–4268.

[7] Y. Liu, Y. Peng, B. Wang, S. Yao, and Z. Liu. *Review on Cyber-physical Systems.* IEEE/CAA Journal of Automatica Sinica, Vol. 4, No. 1, 2017, pp. 27–40.

[8] A. A. Letichevsky, O. O. Letychevskyi, V. G. Skobelev, and V. A. Volkov. *Cyber-Physical Systems.* Cybernetics and Systems Analysis, Vol. 53, Issue 6, 2017, pp. 821–834.

[9] T. A. Henzinger. *The Theory of Hybrid Automata.* Proc. of the 11th Annual IEEE Symposium on Logic in Computer Science (New Brunswick, NJ, USA, 27-30 July, 1996), Los Alamitos, California: IEEE Computer Society Press, 1996, pp. 278–292.

[10] J. F. Raskin. *An Introduction to Hybrid Automata.* In: Handbook of Networked and Embedded Control Systems, Boston, Basel, Berlin: Birkhuser, 2005, pp. 491–518.

[11] A. A. Letichevsky, O. A. Letychevskyi, V. S. Peschanenko, T. Weigert. *Insertion modeling and symbolic verification of large systems.* Lecture Notes in Computer Science, Vol. 9369, 2015, pp. 3–18.

[12] J. Lygeros. *Lecture Notes on Hybrid Systems.* https://robotics.eecs.berkeley.edu/ sastry/ee291e/lygeros.pdf

[13] A. A. Letichevsky. *Algebraic Interaction Theory and Cyber-Physical Systems.* Journal of Automation and Information Sciences, Vol. 49, Issue 9, 2017, pp. 1–19.

[14] V. V. Skobelev, V. G. Skobelev. *On some problems of analysis for hybrid automata.* Cybernetics and Systems Analysis, Vol. 54, Issue 4, 2018, pp. 517–527.

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
Phone: +38 063 431 86 05
E-mail: skobelevvg@mail.ru

Volodymyr V. Skobelev
V.M. Glushkov Institute of Cybernetics of NAS of Ukraine
40 Glushkova ave., Kyiv, Ukraine, 03187
Phone: +38 063 431 86 05
E-mail: vvskobelev@incyb.kiev.ua

The Significance of Online Monitoring Activities for the Social Media Intelligence (SOCMINT)

Elena Șușnea, Adrian Iftene

Abstract

In the last years, social networks have increased in active members who upload and share postings about daily activities (pictures, comments, news, likes etc.), but also use these networks to get informed. During a crisis situation, the occurrence of an unexpected event can generate immediate reactions from active members, which further result in a huge amount of text and other digital resources related to that event. This event can be the trigger factor for crisis escalation with unwanted consequences.

Keywords: SOCMINT, OSINT, Web mining, early warning, crisis, Twitter.

1 Introduction

A great advantage of social networks consists in the timely dissemination of information about the new event, as opposed to the classical sources of news, such as television, radio, and newspapers. From this perspective, the monitoring of social networks is a very important activity for both crisis managers and the intelligence community, because getting timely information about an event may reduce the undesirable effects, through specific actions that lead to crisis stabilization and even to its regression. For example, one unexpected event may occur during night time. In such situations, most open sources information is not updated, as in the case of TV channels which usually replay different shows until morning when new information arrives. Instead, social networks begin to “buzz” by providing information about what happened. Despite the fact that this information is not very accurate, it can be an early warning for both crisis managers and the intelligence community as it offers a snapshot of the event.

2 The Challenges of Social Media for Intelligence

Social media intelligence (SOCMINT) is “the latest member of the intelligence family” [1], joining HUMINT, SIGINT, IMINT, MASINT and OSINT. SOCMINT is recently coined term for confluence of ideas from open source intelligence (OSINT) and Web mining technique (machine learning and database methods) applied to social media data in order to identify and understand those situations from social media environment characterized by behavior of individuals that would affect national security, and accordingly try to make rational decisions to bring the situation to the desired state.

A combination of factors like increasing scale of threats of violence such as terrorism, and the economic and political instability in the Middle East and North Africa have contributed to the increasing flow of migrants and consequently, they will communicate more and more with each other using social media. Also, in the thinking of many European citizens, terrorist threats and the crisis of refugees and migrants are largely tied to each other. A series of terrorist attacks - most claimed by the Israeli Islamic and Muslim Islamic Group (ISIS), a terrorist group in which most refugees and migrants have left or will leave when in their countries of origin - have eroded European public support for refugees and migrants. The results of the survey conducted by the Pew Research Center in July 2016 [2] highlighted the decline in support from European nations of refugee and migrant flows. In eight of the ten European nations participating in the survey, more than 50% of respondents consider that the influx of refugees and migrants will increase the prospects of terrorism. The same study points out that the threat of terrorism and the crisis of refugees and migrants is largely linked to one another in the thinking of many European citizens.

The series of terrorist attacks that have taken place in Europe over the past three years and the UK’s decision to leave the European Union (EU) have diminished the public's confidence in European security. This anxiety manifests itself not only in real life but also in the virtual environment. Social media sites - blogs, micro-blogs, and social networking sites, among others - are very useful both for migrants and residents due to the facilities available for social interactions and content generating. Applications like Twitter, Facebook, YouTube, etc. abound in distributed materials on this theme. Furthermore, social media sites allow active users to immediately react to unexpected events, like the veritable human sensors. The occurrence of such events (protests and terrorist attacks) can be identified much faster by analyzing social networks rather than through classical news sources.

Therefore, the rapid development of information and communications technology and easy Internet access have not only yielded indisputable benefits but also it brings some vulnerabilities to security environment caused by lack of borders, dynamism and anonymity. Because active members are growing in numbers and they upload and share a variety of electronic resources, it is very difficult for intelligence analysts to identify and monitor all those individuals who would affect national security. It's called big data and is characterized by "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" [3]. From this perspective, applying data mining techniques on social media will conduct to extract timely and relevant information needed for intelligence process.

3 Social Media Mining: Content, Usage, Structure

In a crisis situation, social media complexity becomes the limiting factor when intelligence analysts want to extract finished intelligence product. Supplying accurate, timely and usable information to those who make national security decisions is essential.

Although there are currently many benefits to using these data mining techniques in the field of information, especially for social media analysis, there are some limitations on usability. One limitation is that, although data mining techniques can help in discovering patterns, they do not tell the user the value or significance of these models. These types of determinations must be made by the information analyst. For example, in order to assess the validity of a data mining project designed to identify a person suspected of terrorist activities, the user can test a pattern generated on data that includes information about known terrorists. But even if the pattern confirms the profile of this terrorist, it does not mean that the person has to be labeled as terrorist.

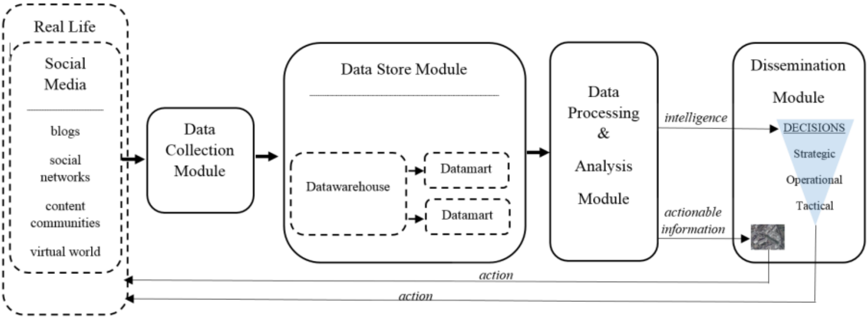


Figure 1. Information system architecture for social media monitoring

The model of the information system for monitoring the crisis of refugees and migrants based on the exploitation of open source resources comprising four individual modules (Figure 1):

(1) The data collection module that allows automatic or semi-automatic collection of social media data (social networks, blogs, microblogs, wiki) for the purpose of digital data collections. From a large data perspective, data collection is difficult to achieve. Not just the huge volume of data is a problem, but their format also generates difficulties in the collection process. Finally, digital collections must contain metadata, text, image, audio, and video files.

(2) The data storage module has the role of cleaning the data collected from multiple open sources and ensuring their efficient access by storing in a data repository. Before the data reaches the data warehouse, it goes through a specific process called extract-transform-load (ETL). Even if there is a centralized warehouse, it does not mean that the data cannot fit in and stored in specific warehouses for each open source category.

(3) The data analysis and intelligence product module convert data into intelligence information. There are three types of tasks for social

media mining. Firstly, Web content mining supports intelligence process to extract actionable information from Web page content. In contrast to one decade ago, the information in the digital universe today is predominantly unstructured (text, images, voice, video, etc.) and increasing the content of the web page was structured (data was organized in tables or relational databases) trace. The techniques used for mining Web content include classification, clustering, language processing, and decision tree. Secondly, Web usage mining focuses on the discovery of patterns from Web usage logs, which stores every click made by a user, such as IP addresses, page references, and date and time of access. In this case, there are some techniques applicable such as association rules, sequence analysis, and information extraction. Lastly, Web structure mining consists of extracting insights from the Web using social network analysis and PageRank. Of the three types of social media mining, we aim to highlight how to use language processing on web content available on the Twitter platform.

(4) The intelligence reporting and dissemination module supports the interaction between policy makers and military decision-makers and intelligence analysts, providing a collaborative virtual space. Final users will access the computer system via a Web interface. The interface will allow access to the computer system for any authorized user by using a web browser. Collaborative virtual space is very important for the creation of the finished product by information analysts.

4 Real Time Mining Twitter Content for SOCMINT

Using Twitter for social media monitoring offers some benefits, such as: The Twitter platform is used by different people to express their opinion on the various topics or events that take place, so it is a valuable source of opinion for the intelligence community and implicit for policy-makers; the Twitter app contains an enormous number of tweets and expands every day so that the volume of text you collect can be extremely high in this way the accuracy of intelligence analysis increases; the Twitter audience differs from other social network users because it is much more varied, from simple people to celebrities, representatives of military or nonmilitary organizations, politicians, heads of state; the Twitter audience is made up of users from multiple countries. As the Twitter platform publisher grows every day and the services offered expand, the data from this open source can be used in the sentiment analysis.

Although the Twitter app is somewhat new, the scientific research on its use in crisis situations has expanded rapidly. For example, the Twitter platform has been successfully used by the general public in mass protests that took place in Iran, Tunisia and Egypt, especially for the exchange of information. Thus, on June 12, 2009 in Iran, or held presidential elections whose results dissatisfied with Iranian citizens. So the next day, an estimated crowd of several million people gathered in the streets around the Azadi Square in Tehran to protest. During the weeks following the disputed elections, the government has tightened control over data networks and Internet gateways, so traffic has been greatly reduced. In this context, the use of Twitter has been particularly important. First, governments blocked access to the Twitter site, so it was accessible only through a proxy server or a text message on a mobile phone. Secondly, street protesters often needed updated real-time information to avoid confrontations with law enforcement. Thus, they used their mobile phones not only to take pictures or videos, but also to disseminate information about what was happening at that time and the specific location, such as the streets that should be avoided due to police presence. This was also the case for Tunisia and Egypt.

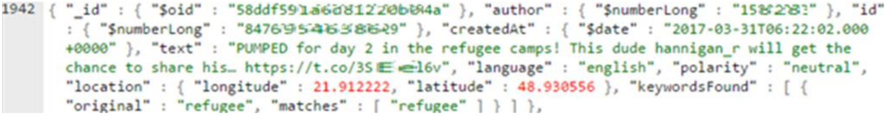
We will use Twitter to analyze two very important indicators: geolocation and sentiment analysis. Tweets feeds are important for analysis because data generation can have a very high frequency and algorithms that process them need to do so under very strict storage and time conditions. Twitter users generate about half a billion tweets a day. Some of these tweets are available to researchers and developers via the Twitter API [4]. To analyze these indicators, we will go through the following steps: first we will collect real-time tweets on the basis of which we will analyze geolocation and feelings analysis, then process data in real time with data mining techniques and save them in a database, and at the end we will use the information for the viewing component.

5 Collection data from Twitter

For collecting tweets, we used a web crawler that connects to the Twitter stream and receives, filters and sends tweets continuously for analysis. The visual representation of the geographical locations from which messages were distributed will be done by filtering the received tweets so that we only process those that have the geographic coordinates attached. In addition, we will perform a language filtering because the annotator used in this case study can only recognize emotions in tweets written in English, French, German, and Arabic. The key words used were: NATO, European Union, refugee, civilian, struggle, and peace. We will also take into account the lexical field of each keyword. Data collection took place between 18 March 2017 and 04 April 2017 and resulted in a total of 66,975 tweets. It is impossible for these data to be analyzed individually, but we will propose two examples to be discussed in the following section.

6 Real-time processing and viewing based on keyword search

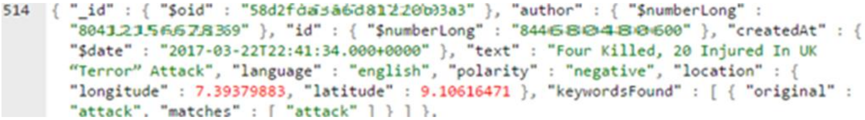
At this stage, tweets are available as text data and each row contains a tweet. We will analyze some existing tweets in the monitoring system database to exemplify the multitude and variety of information that can be extracted. The first example, a tweet received in March 2017, which contains the geographical coordinates of a reception center for asylum seekers in eastern Slovakia (Figure 2). The message was identified based on the keyword refugee. The huge amount of data we have is compelling us to a more complex approach, namely a big data analysis. In this regard, we used a Naive Bayes classifier [5] that framed this tweet, in terms of polarity, into the neutral class. Messages labeled neutral do not indicate the occurrence of events leading to escalation of the crisis.



```
1942 { "_id" : { "$oid" : "58ddf59[illegible]" }, "author" : { "$numberLong" : "15[illegible]" }, "id"
     : { "$numberLong" : "8476[illegible]" }, "createdAt" : { "$date" : "2017-03-31T06:22:02.000
     +0000" }, "text" : "PUMPED for day 2 in the refugee camps! This dude hannigan_r will get the
     chance to share his… https://t.co/3S[illegible]l6v", "language" : "english", "polarity" : "neutral",
     "location" : { "longitude" : 21.912222, "latitude" : 48.930556 }, "keywordsFound" : [ {
     "original" : "refugee", "matches" : [ "refugee" ] } ] },
```

Figure 2. Example of tweet in the iSIAD database

Another example of the tweets we propose for analysis is presented in the picture below. It contains the words terror, attack, killed and has been automatically identified as having negative polarity (Figure 3).



```
514  { "_id" : { "$oid" : "58d2fda3a6d81220b93a3" }, "author" : { "$numberLong" :
     "804121566728369" }, "id" : { "$numberLong" : "844680480600" }, "createdAt" : {
     "$date" : "2017-03-22T22:41:34.000+0000" }, "text" : "Four Killed, 20 Injured In UK
     “Terror” Attack", "language" : "english", "polarity" : "negative", "location" : {
     "longitude" : 7.39379883, "latitude" : 9.10616471 }, "keywordsFound" : [ { "original" :
     "attack". "matches" : [ "attack" ] } ] }.
```

Figure 3. An example of a tweet that refers to a terrorist attack

Indeed, the message is classified correctly from the point of view of polarity because it refers to the terrorist attack that took place on March 22, 2017 in the British Parliament, located in the vicinity of Westminster Palace in London, (Figure 4).

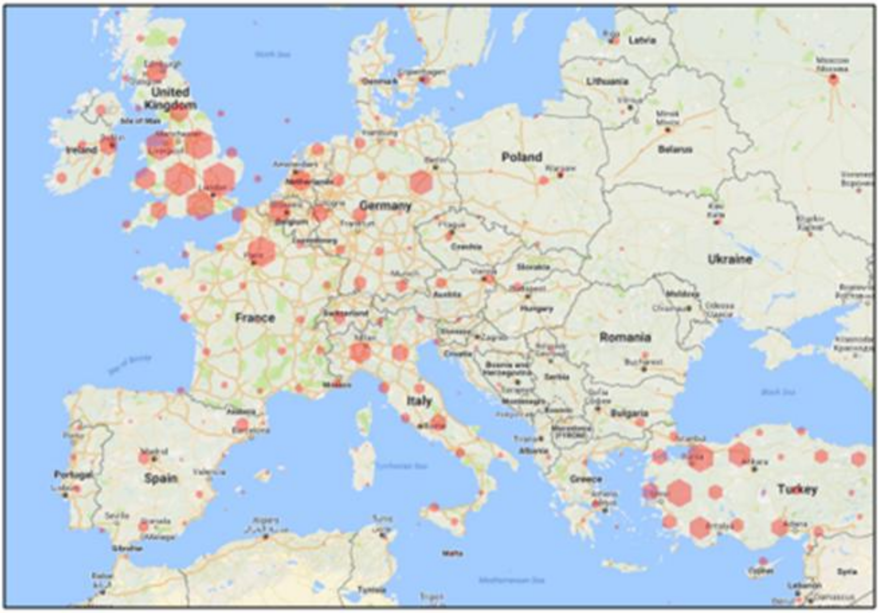


Figure 4. Visualization of results using Google Maps

7 Visualization and decision maker interaction

For visual representation of data, we first clean or remove duplicate tweets because they will induce biases in the classification process. Then we need to remove the punctuation marks and other non-useful symbols (e.g. emoticons, links), as they may reduce the efficiency and affect the accuracy of the overall process. MapReduce [6] is a new parallel programming model, so the classic Naive Bayes algorithm based on sentiment analysis is adjusted to suit the MapReduce model.

In order to classify the collected tweets by opinions we used Sentiment140 API [7], which internally uses a (semi-supervised) trained model. Unfortunately Sentiment140 API only works with Spanish and English languages. In this situation we would have been forced to discard a big part of our collected data. To overcome this situation we decided to make use of a translation API (Yandex) [8] which would translate the tweet's content from its native language into English. Once translated in English the tweet is sent to Sentiment140 API [7].

Also, in order to improve the quality of results, we chose to use the Naive Bayes classifier along with an English lexical SentiWordNet [9] to improve the accuracy of the tweet classification. The classes used to analyze feelings are the following three: positive, negative, neutral.

The map above shows custom bookmarks using hexagons (Figure 4) or circles (Figure 5). The size of the hexagons increases in proportion to the number of published tweets that contain the specified keywords in that particular location. The visualization of hot spots facilitates the understanding by political and military decision-makers of the distribution of the population interested in a particular subject/event. Instead of placing a single marker for each hot spot, we used variable-level markers to represent the data distribution.

When a cluster grows, it affects other clusters on the map, which will decrease proportionally, so the largest clusters being the most visible. To speed the display the clusters who associated a circle with a radius less than a value set by the user are displayed as dots. If the user will use the option of zooming in a given region (Figure 5 left), where he spotted a cluster higher, it will be divided into smaller clusters, until we see only clusters composed of a single tweet (Figure 5 right).

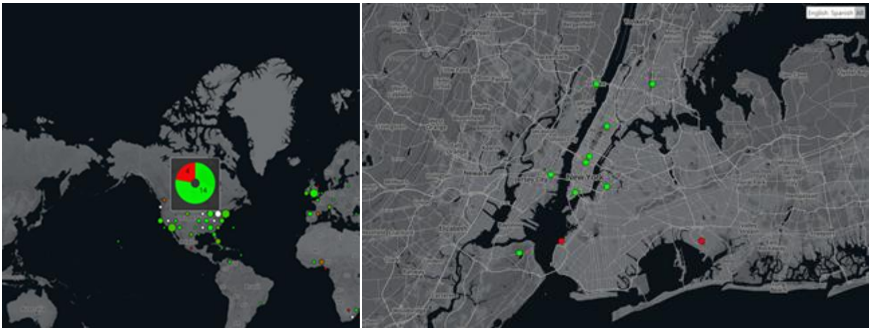


Figure 5. (a) View details of a larger cluster, (b) zoom in for a region from the map

8 Conclusion

In conclusion, real-time analysis of data available from open sources on the Web contributes significantly to the crisis management process, since both the intelligence community and policy and military decision-makers can quickly understand the context of changes that could lead to degeneration of the situation. Tweets analysis can prevent decision-makers in charge of crisis management from occurring unexpected events and develop models that can be used to make proactive decisions to mitigate unwanted consequences.

At the strategic level, SOCMINT can provide indications and warnings about both hostile intentions of crisis-engaging entities and opportunities that should be exploited by decision-makers. The analysis of various sources of information such as regional newspapers in the Middle East, comments posted on various social networks by people coming to Europe or those residing in the countries of destination of refugees and migrants, extraneous audio and video materials distributed in social networks often represent a more robust basis for estimating stability or instability than reports from clandestine sources with a limited accessibility level and a personal perspective that influences the objectively assumed character of the report.

Acknowledgments. POC-A1-A1.2.3-G-2015 program, as part of the PrivateSky project (P_40_371/13/01.09.2016) has supported part of the research for this paper.

References

[1] S. D. Omand. *Introducing Social Media Intelligence (SOCMINT)*. Intelligence and National Security, vol. 27, no. 6, (2012), pp. 801-823.

[2] Pew Research Center, Europeans Fear Wave of Refugees Will Mean More Terrorism, Fewer Jobs, July, 2016, accessed 26 April 2018 at www.pewresearch.org

[3] Gartner IT Glossary, accessed 26 April 2018 at http://www.gartner.com/it-glossary/big-data/.

[4] Twitter API, accessed 26 April 2018 at https://dev.twitter.com/rest/public

[5] S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, (1995)

[6] R. Lämmel. *Google's MapReduce programming model*. In Science of Computer Programming, vol. 70, issue 1, (2008), pp. 1-30, ISSN 0167-6423, accessed 26 April 2018 at https://doi.org/10.1016/j.scico.2007.07.001

[7] Sentiment140 API, accessed 26 April 2018 at http://help.sentiment140.com/api

[8] Yandex, accessed 26 April 2018 at https://www.yandex.com/

[9] SentiWordNet, accessed 26 April 2018 at http://sentiwordnet.isti.cnr.it/

Elena Șușnea[1], Adrian Iftene[2]

[1]"Carol I" National Defense University, Bucharest, Romania
E-mail: esusnea@yahoo.com

[2]"Alexandru Ioan Cuza" University, General Berthelot, No. 16, Iasi, Romania
E-mail: adiftene@info.uaic.ro

Multi-linear Maps in Cryptography

Ferucio Laurenţiu Ţiplea

Abstract

Bilinear maps developed from the Weil and Tate pairings over elliptic curves constitute a great achievement in cryptography. By using them, one-round three-party key exchange was possible, and practical cryptographic primitives were designed, such as identity-based encryption schemes and aggregate signatures.

This paper starts with a short overview on bilinear maps and some of its most important applications to cryptography. Then, it moves to the problem of constructing multi-linear maps as an effort to design one-round multi-party key-exchange protocols as well as more specialized cryptographic primitives such as efficient broadcast encryptions and unique digital signatures.

Keywords: bilinear map, elliptic curve, cryptography.

1 Introduction

The introduction of elliptic curves as a tool for cryptography [24, 20] has opened a large spectrum of techniques with various applications ranging from cryptanalysis to primitive cryptographic design. One of these techniques is based on modified versions of the Weill and Tate pairings, mostly known as (cryptographic) bilinear maps. Menezes et al. [23] have used bilinear maps to reduce the elliptic curve discrete logarithm problem to groups where sub-exponential methods (such as index-calculus) can be applied. Later, Joux [18] showed how bilinear maps can help to define a one-round three-party key exchange protocol as a natural generalization of the Diffie-Hellman key exchange protocol. After just one year, a seventeen years old problem had to be solved: Boneh and Franklin [5] designed the first (practical) identity-based encryption scheme. Since then, the number of applications of bilinear maps to cryptography has exploded. Although the only (cryptographic) bilinear maps we know are variations of the Weil and Tate pairings, they were continuously refined and improved (e.g., [4, 15]) and included in specialized cryptographic applications and libraries.

Generalization of bilinear maps to multi-linear maps would have even more interesting applications [6]: one-round multi-party key exchange, efficient broadcast encryption, attribute-based encryption for general Boolean circuits [30, 28, 14], more efficient searchable encryption schemes and so on. Unfortunately, multi-linear maps could not (or maybe, cannot) be obtained (at least for the time being) by simply generalizing the Weil and Tate pairings [6]. In fact, no one knows how such structures can be computed in a similar way to bilinear maps. Garg et al. [16] have proposed *graded encoding systems* (GES) as a method to "simulate" leveled multi-linear maps. Their GES scheme is based on ideal lattices. Shortly after that, an integer-based GES scheme has been proposed [11]. Unfortunately, both proposals, as well as some follow-up constructions, were proven insecure. More recently, an approach based on indistinguishability obfuscation, homomorphic encryption, and non-interactive zero-knowledge proof systems, has emerged [2, 3]. It follows the line in [6], which means that it is for clean multi-linear maps.

Our paper makes a short incursion into this fascinating field of (cryptographic) bilinear/multi-linear maps. We start in the next section with an abstract view on bilinear maps and discuss some immediate consequences with respect to the discrete logarithm and the Diffie-Hellman problems. Then, we focus on the Weil and Tate pairings as building blocks for bilinear maps, and close the section by some applications. Section 3 discusses the natural generalization to multi-linear maps. Some arguments are provided, as well as some variants of them are presented. We close the section by a brief overview of two main techniques for the simulation of multi-linear maps: GES and indistinguishability obfuscation. We conclude in the last section.

2 Bilinear maps

From a mathematical point of view, a *bilinear map* is defined on a Cartesian product domain and is linear in each of the two arguments. Matrix multiplication and bilinear forms on vector spaces over some field are examples of bilinear maps. However, in cryptography we are interested in bilinear maps that are efficiently computable and also offer some security properties (sometimes called cryptographic bilinear maps). We begin our exposition with an abstract view on such maps and then we show how they can be effectively obtained.

2.1 Bilinear maps from an abstract point of view

Throughout this paper we assume the reader familiar with concepts on group and field theory. We shall use $G = \langle g \rangle$ to denote a finite cyclic group generated by g, and 1_G for the unity of G. The notation $a \leftarrow A$ specifies that a is drawn uniformly at random from A.

Definition 1. *A map $e : G_1 \times G_2 \to G_t$, where G_1, G_2, and G_t are finite cyclic groups of the same order, is called* bilinear *if:*

1. (bilinearity) $e(x^a, y^b) = e(x, y)^{ab}$*, for all* $x \in G_1$*,* $y \in G_2$*, and* $a, b \in \mathbb{Z}$*;*
2. (non-degeneracy) $G_t = \langle e(x, y) \rangle$*, for some* $x \in G_1$ *and* $y \in G_2$*;*
3. (computability) e *is efficiently computable.*

As G_1, G_2, and G_t are cyclic groups of the same order n, they are isomorphic to the additive cyclic group $\mathbb{Z}_n$. Some authors use additive notation for G_1 and G_2, and multiplicative notation for G_t (the target group). We prefer to use multiplicative notation for all groups.

The non-degeneracy property can equivalently be replaced by "$G_t = \langle e(g_1, g_2) \rangle$, for any generator g_1 of G_1 and g_2 of G_2", and when n is a prime, this property is equivalent to "$e(x, y) \neq 1_{G_t}$, for some $x \in G_1$ and $y \in G_2$" (because any $z \in G_t$ with $z \neq 1_{G_t}$ is a generator for G_t).

One may easily remark that for composite orders $n = km$, where $2 \leq k, m < n$, it follows $e(x, y) = 1_{G_t}$ for any x in the subgroup of order k of G_1 and any y in the subgroup of order m of G_2.

In many practical applications, bilinear maps are defined on $G \times G$. If we extend this to $G_t = G$ as well, then we arrive at the concept of *self-bilinear map*.

The existence of bilinear maps has very interesting consequences with respect to some hard problems in the cyclic groups on which they act. Recall first that, given a cyclic group $G = \langle g \rangle$ of prime order p:

- The *discrete logarithm* (DL) problem in G is to compute a, given (g, g^a);
- The *decisional Diffie-Hellman* (DDH) problem in G is to distinguish between the distributions (g, g^a, g^b, g^{ab}) and (g, g^a, g^b, g^z), where $a, b, z \leftarrow \mathbb{Z}_p$;
- The *computational Diffie-Hellman* (CDH) problem in G is to compute g^{ab}, given (g, g^a, g^b).

Some consequences of the existence of bilinear maps are in order.

Theorem 1. *If $e : G \times G \rightarrow G_t$ is a bilinear map, then the DL problem in G is no harder than the DL problem in G_t.*

Theorem 1 has been used in [23] to reduce the DL problem in some *elliptic curve groups* to the DL problem in finite groups where subexponential attacks are known (more details in Section 2.3).

Theorem 2. *If $e : G \times G \rightarrow G_t$ is a bilinear map, then the DDH problem in G is easy.*

Although the DDH problem is easy in G when bilinear maps $e : G \times G \rightarrow G_t$ are known, the CDH problem could still be hard in G (it is not known how bilinear maps may help solving the CDH problem). Therefore, the existence of *gap Diffie-Hellman* (GDH) *groups* (i.e., groups where the DDH problem is easy but the CDH problem is hard) might not be affected by the existence of bilinear maps on them.

However, if we ask for self-bilinear maps on G, their existence excludes hardness of the CDH problem in G.

Theorem 3 ([7]). *Let G be a cyclic group of prime order p. If there are self-bilinear maps $e : G \times G \to G$, then the CDH problem in G can be solved by $\mathcal{O}(\log p)$ evaluations of e.*

As one can see, the existence of bilinear maps has a very serious impact on the hardness status of the DL, DDH, and CDH problems. It is natural to expect that some other problems become hard in the context of bilinear maps. Two such problems are obtained as "bilinear variations" of the DH problems. Let $e : G \times G \to G_t$ be a bilinear map, where $G = \langle g \rangle$ and G_t are cyclic groups of prime order p. Then:

- The *decisional bilinear Diffie-Hellman* (DBDH) problem in G is to distinguish between the distributions $(g, g^a, g^b, g^c, e(g,g)^{abc})$ and $(g, g^a, g^b, g^c, e(g,g)^z)$, where $a, b, c, z \leftarrow \mathbb{Z}_p$;
- The *computational bilinear Diffie-Hellman* (CBDH) problem in G is to compute $e(g,g)^{abc}$, given (g, g^a, g^b, g^c).

We refer the reader to [19, 7] for a diagram with relations between complexity assumptions in pairing-based cryptography.

2.2 Constructions of bilinear maps

More or less, the only cryptographic bilinear maps we know are variations of the Weil and Tate pairings on *elliptic curve groups.* An *elliptic curve* (EC) over a field F is defined by a non-singular generalized Weierstrass equation

$$E: \quad y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where $a_1, a_2, a_3, a_4, a_6 \in F$. If $char(F) > 3$, a linear change of variables leads to a simpler form of the curve, namely

$$E: \quad y^2 = x^3 + ax + b,$$

where $a, b \in F$ and $4a^3 + 27b^2 \neq 0$.

Given a field extension $\overline{F} \supseteq F$, define the set $E(\overline{F})$ of $\overline{F}$*-rational points* as $E(\overline{F}) = \{(x,y) \in \overline{F}^2 \mid (x,y) \text{ satisfy } (E)\} \cup \{\mathcal{O}\}$, where $\mathcal{O}$ is the *point at infinity* [1]. The set $E(\overline{F})$ can be structured as an abelian group with identity $\mathcal{O}$ [29, 26]. The point addition follows the chord-and-tangent rule and the negative of a point (x_0, y_0) is simply obtained by taking the other solution for y when x is fixed to x_0. When the field F is finite, the group $E(F)$ is either a cyclic group or the product of two cyclic groups.

From now on, assume that $\overline{F}$ is the algebraic closure of F. For a positive integer n, define $E[n] = \{P \in E(\overline{F}) \mid nP = \mathcal{O}\}$ the set of *n-torsion points* and $\mu_n = \{x \in \overline{F} \mid x^n = 1\}$ the set of n-roots of unity. If $char(F) \nmid n$, then $E[n] \subseteq E(F)$ and $\mu_n \subseteq F$.

Any function f that does not vanish identically on E has *zeros* and *poles*. The *divisor of f* is a formal sum with integer coefficients

$$div(f) = \sum_P n_P(P),$$

where $n_p > 0$ indicates that P is a zero of f with multiplicity n_p, and $n_p < 0$ indicates that P is a pole of f with multiplicity $|n_p|$.

We define now the *Weil function* $e_n : E[n] \times E[n] \to \mu_n$, by

$$e_n(P, Q) = \frac{f_P(Q+S)}{f_P(S)} \cdot \frac{f_Q(R)}{f_Q(P+R)},$$

where R and S are two arbitrary points in $E(F)$ and f_P and f_Q are such that $div(f_P) = n(P+R) - nR$ and $div(f_Q) = n(Q+S) - nS$.

Theorem 4 ([29, 26]). *Let E be an elliptic curve over a field F and let n be a positive integer not divisible by $char(F)$. Then, the Weil function e_n satisfies the following properties:*

1. *e_n is bilinear in each variable;*
2. *e_n is non-degenerate in each variable: if $e_n(P, Q) = 1$ for all $Q \in E[n]$, then $P = \mathcal{O}$, and similar for the other variable;*
3. *$e_n(P, P) = 1$, for all $P \in E[n]$;*

[1] For a clean treatment of the point at infinity one needs projective spaces, which is far beyond of the scope of this paper.

4. $e_n(P,Q) = e_n(Q,P)^{-1}$*, for all* $P,Q \in E[n]$*.*

To be useful in cryptography, the Weil function (pairing) has to be adapted. For instance, if $P \neq \mathcal{O}$ then $e_n(P,Q) \neq 1$ for at least one Q (Theorem 4(2)). However, we would like to have $e_n(P,P) \neq 1$ for all $P \neq \mathcal{O}$. But this contradicts Theorem 4(3). Another issue with the current form of e_n is that $e_n(P,Q) = 1$ whenever P and Q are linearly dependent. There is one more undesirable property of e_n: the group $E[n]$ is the product of two cyclic groups while we would like to have just one cyclic group for the domain of each argument of e_n. The obvious solution in this case would be to work with cyclic subgroups of $E[n]$. Moreover, this solution also solves the non-degeneracy problem mentioned above.

Theorem 5 ([29, 26])**.** *Let* E *be an EC over a field* F_q*,* n *be an integer such that* $n \mid q-1$*,* $E(F_q)[n]$ *be the set of elements in* $E(F_q)$ *whose order divides* n*, and* $\mu_n = \{x \in F_q \mid x^n = 1\}$*. Then, there are non-degenerate bilinear maps*

$$\langle \cdot, \cdot \rangle_n : E(F_q)[n] \times E(F_q)/nE(F_q) \to F_q^\times/(F_q^\times)^n$$

and

$$\tau_n : E(F_q)[n] \times E(F_q)/nE(F_q) \to \mu_n.$$

The first part in Theorem 5 gives rise to the *Tate-Lichtenbaum pairing*, while the second one leads to the *modified Tate-Lichtenbaum pairing*. The second pairing is more suited for computation because it gives a definite answer instead of a coset in $F_q^\times$ mod n.

A choice for the Tate-Lichtenbaum pairing is as follows. Given $P \in E(F_q)[n]$, we find f_P with $div(f_P) = n(P) - n(\mathcal{O})$, and then choose any $S \in E(F_q)$. Then,

$$\langle P, Q \rangle_n = \frac{f_P(Q+S)}{f_P(S)}.$$

As $F_q^\times$ is a cyclic group of order $q-1$, we may easily get an isomorphism from $E(F_q)/nE(F_q)$ to μ_n by simply raising each element in domain to the power $(q-1)/n$. Therefore, we get

$$\tau_n(P,Q) = \langle P,Q\rangle_n^{\frac{q-1}{n}}.$$

To obtain a symmetric bilinear map from τ_n, we may assume that $E(F_q)[n]$ does not contain elements of order n^2. Then, $E(F_q)[n]$ can be identified with $E(F_q)/nE(F_q)$.

The Weil/Tate pairing can efficiently be calculated by Miller's algorithm [4, 15].

2.3 Applications to cryptography

The development of (cryptographic) bilinear maps has a tremendous impact on cryptography. We will discuss below a few crucial applications that have allowed very important achievements in cryptography.

MOV attack. The DL problem can be formulated for $E(F_q)$ as well, although this group is not generally cyclic. Namely, given two points P and Q such that $Q = \langle P\rangle$, the *elliptic curve discrete logarithm* (ECDL) problem is to compute m such that $Q = mP$. Unlike the DL problem, there is no known sub-exponential algorithm to solve the ECDL problem. This gives an important advantage in practice because we may use smaller groups that offer the same security level like large number theoretic groups. However, some elliptic curves, like the super-singular ones, have to be avoided to benefit of this advantage. This is because the ECDL problem can be reduced to the DL problem [23].

Assume that E is a super-singular EC over a field F_q, and let P a point of order n. Consider $\overline{e} : \langle P\rangle \times \langle P\rangle \rightarrow F_{q^k}^{\times}$ the restricted Weil pairing with k the embedding degree. Given $Q \in \langle P\rangle$, we may obtain m with $Q = mP$ by computing first $\overline{e}(P,Q)$ that must be of the form $\overline{e}(P,P)^m$ and then calculating the DL in $F_{q^k}^{\times}$ w.r.t. $\overline{e}(P,P)$ (see also Theorem 1). As the embedding degree for super-singular curves is no larger than 6, the reduction above shows that the ECDL problem on such curves is vulnerable to this attack, known as the *MOV attack* [23].

One-round three-party key exchange. The Diffie-Hellman (DH) key exchange protocol proposed in 1976 [13] is a method of securely exchanging cryptographic keys between two parties A and B over a public channel. Assuming a cyclic group $G = \langle g \rangle$ of prime order p, A sends to B a value g^a and keeps secret a, while B sends to A a value g^b and keeps secret b. Both parties are able to compute g^{ab}, while the computation of this value from (g, g^a, g^b) is intractable if the group G is suitable chosen.

A natural and important question is whether this one-round two-party protocol can be extended to three parties (in one round). Bilinear maps make this possible as it was shown in [18]. Assume that A, B, and C are the three parties involved in the protocol and $e : G_1 \times G_1 \to G_2$ is a bilinear map. The party A sends g^a to both B and C and keeps a secret, B sends g^b to both A and B and keeps b secret, and C sends g^c to both A and B and keeps c secret. All parties are then able to compute $e(g, g)^{abc}$, while computing this value from (e, g, g^a, g^b, g^c) is intractable if e is suitable chosen.

Identity-based encryption. *Identity-based cryptography* was proposed in 1984 by Adi Shamir [27] who formulated its basic principles but he was unable to provide a solution to it, except for an identity-based signature scheme. A standard scenario on using identity-based encryption (IBE) is as follows. Whenever Alice wants to send a message m to Bob, she encrypts m by using Bob's identity $ID(B)$. In order to decrypt the message received from Alice, Bob asks the private key generator to deliver him the private key associated to $ID(B)$.

In 2000, Sakai et al. [25] have proposed an identity-based key agreement scheme, and one year later, Cocks [10] and Boneh and Franklin [5] have proposed the first IBE schemes. Cocks' solution is based on quadratic residues. It encrypts a message bit by bit and requires $2 \log n$ bits of cipher-text per bit of plain-text. The scheme is quite fast but its main disadvantage is the ciphertext expansion. Boneh and Franklin's solution is based on bilinear maps. Moreover, Boneh and Franklin also proposed a formal security model for IBE, and proved that their scheme is secure under the DBDH assumption.

Boneh and Franklin's scheme uses a bilinear map $e : G_1 \times G_1 \to G_2$, where $G_1 = \langle g \rangle$ and G_2 are cyclic groups of prime order p. The public key of the encryption is simply obtained by applying a hash function $h_1 : \{0,1\}^* \to G_1$ to the identity ID (viewed as a binary string). A security parameter y is kept secret by the private key generator, but g^y is made public.

Whenever the encryptor wants to encrypt a message m, he generates a random s, and computes the ciphertext

$$c = (g^s, m \oplus h_2(e(a, g^y)^s)),$$

where $a = h_1(ID)$ and $h_2 : G_2 \to \{0,1\}^{|m|}$ simply adjusts the length of $e(a, g^y)^s$ to make $\oplus$ possible.

The decryption is simply performed by using the private key a^y (delivered by the private key generator) and noticing that

$$e(a, g^y)^s = e(a^y, g^s)$$

in the virtue of the bilinearity property of e.

3 Multi-linear maps

3.1 Why multi-linear maps

A natural generalization of bilinear maps, for which already many interesting applications have been pointed out [6, 30, 14], would be to consider multi-linear maps.

Definition 2. *Let $G_1, \ldots, G_m, G_t$ be finite cyclic groups of the same order, where $m \geq 2$. A map $e : G_1 \times \cdots \times G_m \to G_t$ is n-*linear *if:*

1. (n-linearity) *$e(x_1^{a_1}, \ldots, x_m^{a_m}) = e(x_1, \ldots, x_m)^{a_1 \cdots a_m}$, for all $x_i \in G_i$ and $a_i \in \mathbb{Z}$, $1 \leq i \leq m$;*
2. (non-degeneracy) *$G_t = \langle e(x_1, \ldots, x_m) \rangle$, for some $x_i \in G_i$, $1 \leq i \leq m$;*

3. (computability) *e is efficiently computable.*

The existence of multi-linear maps would easily lead to a generalization of the Joux protocol to a one-round multi-party key exchange protocol, design of unique signature schemes (where every message has a unique digital signature), broadcast encryption with short keys and transmissions [6]. The recent development of attribute-based encryption (ABE) for general Boolean circuits makes use of "special forms" of multi-linear maps whose existence is closely related to the existence of multi-linear maps. Two such special forms are in order.

Definition 3 ([30])**.** *A* leveled multi-linear map *is a set of bilinear maps $\mathbf{e} = \{e_{i,j} : G_i \times G_j \to G_{i+j} | i, j \geq 1, i + j \leq k\}$ that satisfy $e_{i,j}(g_i^a, g_j^b) = g_{i+j}^{ab}$, for all $i, j \geq 1$ with $i + j \leq k$ and all $a, b \in \mathbb{Z}_p$, where $G_i = \langle g_i \rangle$ are cyclic groups of prime order p, for all i.*

Definition 4 ([14])**.** *A* chained multi-linear map *is a sequence of bilinear maps $(e_i : G_i \times G_1 \to G_{i+1} | 1 \leq i \leq k)$, where $G_1, \ldots, G_{k+1}$ are multiplicative groups of the same prime order.*

If $(e_i | 1 \leq i \leq k)$ is a chained multi-linear map (as above) and $G_1 = \langle g_1 \rangle$, then $g_{i+1} = e_i(g_i, g_1)$ is a generator of G_{i+1}, for all $1 \leq i \leq k$ (because e_i is a bilinear map). Therefore, $(e_i | 1 \leq i \leq k)$ can be regarded as a particular case of a leveled multi-linear map.

Neither chained nor leveled multi-linear maps have constructions similar to bilinear maps. The authors of [30, 14] have used such maps arguing that they exist due to the existence of secure *graded encoding systems* (GES). Unfortunately, the latest results have shown that all GES schemes have security problems (more details are provided in the next section).

3.2 In search for multi-linear maps

As natural generalizations of bilinear maps, one may look for multilinear extensions of the Weil and Tate pairings in order to obtain multilinear maps. Boneh and Silverberg gave evidence that this might not be possible (even within the realm of algebraic geometry), suggesting that genuinely new techniques would be necessary to construct multi-linear maps [6]. A natural approach then would be to look for cryptographic systems capable to "simulate" multi-linear maps. There are mainly two approaches along this line: *graded encoding systems* and *constructions based on indistinguishability obfuscation*. The first approach focuses on leveled multi-linear maps, while the second is for (plain) multi-linear maps. The connection between multi-linear maps and indistinguishability obfuscation is even more intricate: all existing indistinguishability obfuscation constructions rely on multi-linear maps and graded encodings. Moreover, it was recently shown [22] that trilinear maps and block-wise local pseudo-random generators suffice to construct indistinguishability obfuscation.

Graded encoding systems. The approach proposed in [16] for leveled multi-linear maps comes with an encoding system for the exponent x of g_i^x. When g_i^x and g_i^y are multiplied, encodings of x and y should lead to an encoding of $x+y$. Similarly, as $e_{i,j}(g_i^x, g_j^y) = g_{i+j}^{xy}$, encodings of x and y should lead to encodings of xy. Moreover, if the encoding of x is on some level i and the encoding of y is on j, then the encoding of xy must be on the level $i+j$. More precisely, we have the following definition.

Definition 5 ([16])**.** *A* k-graded encoding system *(k-GES) consists of a ring R and a family of sets $S = (S_i^{(\alpha)} \mid 0 \leq i \leq k, \alpha \in R)$ with all $S_i^{(\alpha)} \subseteq \{0,1\}^*$ and such that the following properties hold:*

1. *For all i, the sets $(S_i^{(\alpha)} \mid \alpha \in R)$ are pairwise disjoint;*
2. *There exist an associative binary operation "$+$" and a unary operation "$-$" such that $u_1 + u_2 \in S_i^{(\alpha+\beta)}$ and $-u_1 \in S_i^{(-\alpha)}$, for all i, all $\alpha, \beta \in R$, $u_1 \in S_i^{(\alpha)}$, and $u_2 \in S_i^{(\beta)}$;*
3. *There exists an associative binary operation "$\times$" such that $u_1 \times u_2 \in S_{i+j}^{(\alpha\cdot\beta)}$, for all i and j, all $\alpha, \beta \in R$, $u_1 \in S_i^{(\alpha)}$, and $u_2 \in S_i^{(\beta)}$.*

The meaning of $S_i^{(\alpha)}$ is that it is a set of encodings of α on level i. In fact, all these encodings stand for g_i^α.

Any GES scheme should provide not only algorithms for the three operations in Definition 5 (addition, negation, multiplication), but also algorithms for encoding, zero-test (to decide whether a string u is in some set $S_i^{(\alpha)}$), and extraction (of elements from sets $S_i^{(\alpha)}$).

The security of the lattice-based GES scheme in [30] relies on seemingly hard problems in ideal lattices. Unfortunately, the attack proposed in [8] applies not only to this scheme but also to the follow-up constructions [21, 1]. Other GES schemes was proposed in [11, 12], using integers instead of ideal lattices. Unfortunately, these constructions were proven insecure as well [9].

Constructions based on indistinguishability obfuscation. In order to see how indistinguishability obfuscation is used to construct multi-linear maps, we need first a few concepts from cryptography.

A *circuit family* is a family $\mathcal{C} = (\mathcal{C})_{\lambda\in\mathbb{N}}$ of sets of (deterministic or randomized) circuits, where each circuit $C \in \mathcal{C}_{\lambda\in\mathbb{N}}$ has a polynomial size $poly(\lambda)$, an arity $a(\lambda)$, an input domain $(\{0,1\}^\lambda)^{a(\lambda)}$, and a co-domain $\{0,1\}^\lambda$ ($poly$ and a are functions given for the entire circuit family $\mathcal{C}$).

A *homomorphic public-key encryption* (HPKE) scheme for $\mathcal{C}$ is a PKE scheme $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ together with an *Eval* function that on input a *pk* key, a circuit $C \in \mathcal{C}_\lambda$, and ciphertexts $c_1, \ldots, c_{a(\lambda)}$, outputs a ciphertext c such that if $c_i \leftarrow \mathcal{E}(m_i, pk)$ for all i and $m = C(m_1, \ldots, m_{a(\lambda)})$, then $c \leftarrow \mathcal{E}(m, pk)$.

An *obfuscator* for $\mathcal{C}$ is a PPT algorithm that on input 1^λ and $C \in \mathcal{C}_\lambda$ outputs a circuit $\bar{C}$ such that no distinguisher can find, with non-negligible probability, inputs x with $C(x) \neq \bar{C}(x)$. An obfuscator as above is an *indistinguishability obfuscator* (IO) if no PPT algorithm can distinguish with non-negligible probability from which of two circuits C_0 and C_1 comes an obfuscated circuit $\bar{C}$.

Now, we are ready to briefly describe the approach recently taken in [2]. We want to construct a multi-linear map $e : G_1^m \to G_t$. This is done as follows:

1. Choose a group $G_0 = \langle g_0 \rangle$, where it is hard to distinguish $g_0^{\omega^m}$ from a random element in G_0, given $(g_0, g_0^{\omega}, \ldots, g_0^{\omega^{m-1}})$ (ω being a secret value). Then, G_t may simply be chosen as G_0;

2. The elements of G_1 are vectors (h, c_1, c_2, π), where $h = g_0^{x_1} g_0^{\omega x_2} \in G_0$, c_1 is a homomorphic encryptions of $x = (x_1, x_2)$ under a public keys pk_1, c_2 is a homomorphic encryption of another vector y representing h under a public key pk_2, and π is a non-interactive zero-knowledge (NIZK) proof that x and y both represent the same element h.

   Addition in G_1 is carried out by an obfuscation of a circuit that has the secret keys sk_1 and sk_2 hard-coded in (recall that pk_1 and pk_2 are used to define the elements of G_1);

3. The multi-linear map on $((h_i, c_{i,1}, c_{i,2}, \pi_i) \mid 1 \leq i \leq m)$ is computed by the obfuscation of another circuit that has sk_1 and ω hard-coded in. This allows to extract $(x_{i,1} + \omega x_{i,2})$ and compute $g_0^{\prod(x_{i,1}+\omega x_{i,2})}$.

For a multi-linear map $e : G_1 \times \cdots \times G_m \rightarrow G_t$, m values ω_i are needed; the construction is similar to the one above.

To be useful in cryptography, a multi-linear map as above should give evidence that some problems are hard with respect to them. One of the most natural problem is the *decisional multi-linear Diffie-Hellman* (DMDH) problem which asks to distinguish from an element $e(g_1, \ldots, g_m)^{a_1 \cdots a_m}$ and a random element in G_t, given g_i and $g_i^{a_i}$, for all $1 \leq i \leq m$.

In [2], the authors argued that the DMDH problem is hard in their construction. However, several technical problems were encountered later, that led to a new version of the construction [3]. It remains to see if this new construction is indeed secure.

Unlike lattice-based GES schemes, the construction based on indistinguishability obfuscation is noiseless and is much closer to multi-linear maps as defined in [6].

4 Conclusion

The introduction of elliptic curves to cryptography have had a great impact. On the one side it was possible to design cryptographic primitives with smaller security parameters without compromising security, and on the other side it has opened the door to the Weil and Tate pairings. From these, cryptographic bilinear maps became reality and many applications of them have evolved. But then, multi-linear maps come as a natural generalization that bring even more very important applications.

Although some good steps were made in order to clarify the mathematics behind bilinear maps, many crucial problems still remain open:

1. Is it possible to construct (more efficient) bilinear maps from other mathematical structures ? As far as we know, the only bilinear maps we use so far are variations of the Weil and Tate pairings. Bilinear maps are the most costly operations in cryptography;
2. Is it possible to generalize the existing constructions of bilinear maps in order to get multi-linear maps ? If this is not possible, what other mathematical structures can be used to define cryptographic multi-linear maps ?
3. Graded encoding systems were, at least for a moment, a hope for multi-linear maps. Recent results have shown that this hope is very fragile and more work is needed to understand well these kinds of structures. The connection with indistinguishability obfuscation is important but, unfortunately, nothing is known on the existence of such constructions.

References

[1] M. Albrecht, C. Cocis, F. Laguillaumie, A. Langlois. *Implementing candidate graded encoding schemes from ideal lattices.* ASIACRYPT 2015, LNCS 9453, pp. 752–775, 2015.

[2] M. Albrecht, P. Farshim, D. Hofheinz, E. Larraia, K.G. Paterson. *Multilinear maps from obfuscation.* TCC 2016, LNCS 9562, pp. 446–473, 2016.

[3] M. Albrecht, P. Farshim, S. Han, D. Hofheinz, E. Larraia, K.G. Paterson. *Multilinear maps from obfuscation.* Cryptology ePrint Archive, Report 2015/780, last revised 18 Dec 2017.

[4] P. Barreto, B. Lynn, M. Scott. *Efficient implementations of pairing-based cryptosystems.* Journal of Cryptology, vol. 17, pp. 321–334, 2004.

[5] D. Boneh, M.K. Franklin. *Identity-based encryption from the Weil pairing.* CRYPTO 2001, LNCS 2139, pp. 213–229, 2001.

[6] D. Boneh, A. Silverman. *Applications of multilinear forms to cryptography.* Contemporary Mathematics, Vol. 324, American Mathematical Society, pp. 71–90, 2003.

[7] J.H. Cheon, D.H. Lee. *A note on self-bilinear maps.* Bulletin of the Korean Mathematical Society 46(2), pp. 303–309, 2009.

[8] J.H. Cheon, J. Jeong, C. Lee. *An algorithm for NTRU problems and cryptanalysis of the GGH multilinear map without a low level encoding of zero.* LMS Journal of Computation and Mathematics, 19(A), pp. 255–266, 2016.

[9] J.H. Cheon, P.-A. Fougue, C. Lee, B. Minaud, H Ryu. *Cryptanalysis of the new CLT multilinear map over the integers.* EUROCRYPT 2016, LNCS 9665, pp. 509–536, 2016.

[10] C. Cocks. *An Identity based encryption scheme based on quadratic residues.* Proceedings of the 8th IMA International Conference on Cryptography and Coding, pp. 360–363, 2001.

[11] J.-S. Coron, T. Lepoint, M. Tibouchi. *Practical multilinear maps over the integers.* CRYPTO 2013, LNCS 8042, pp. 476–493, 2013.

[12] J.-S. Coron, T. Lepoint, M. Tibouchi. *New multilinear maps over the integers.* CRYPTO 2015, LNCS 9215, pp. 267–286, 2015.

[13] W. Diffie, M. Hellman. *New directions in cryptography.* IEEE Transactions on Information Theory, Vol. 22, pp. 644–654, 1976.

[14] C. Drăgan, F.L. Ţiplea. *Key-policy attribute-based encryption for general Boolean circuits from secret sharing and multi-linear maps.* BalkanCryptSec 2015, LNCS 9540, pp. 112–133, 2015.

[15] S. Galbraith, K. Harrison, D. Soldera. *Implementing the Tate pairing.* Algorithmic Number Theory Symposium ANTS 2002, LNCS 2369, pp. 324–337, 2002.

[16] S. Garg, C. Gentry, S. Halevi. *Candidate multilinear maps from ideal lattices.* EUROCRYPT 2013, LNCS 7881, pp. 1–17, 2013.

[17] C. Gentry, S. Gorbunov, S. Halevi. *Graph-induced multilinear maps from lattices.* TCC 2015, LNCS 9015, pp. 498–527, 2015.

[18] A. Joux. *A one round protocol for tripartite Diffie-Hellman.* Algorithmic Number Theory Symposium ANTS 2000, LNCS 1838, pp. 385–393, 2000.

[19] A. Joux. *The Weil and Tate pairings as building blocks for public key cryptosystems.* Algorithmic Number Theory Symposium ANTS 2002, LNCS 2369, pp. 20–32, 2002.

[20] N. Koblitz. *Elliptic curve cryptosystems.* Mathematics of Computation, vol 48 (177), pp. 203–209, 1987.

[21] A. Langlois, D. Stehlé, R. Steinfeld. *GGHLite: More efficient multilinear maps from ideal lattices.* EUROCRYPT 2014, LNCS 8441, pp. 239–256, 2014.

[22] H. Lin, S. Tessaro. *Indistinguishability Obfuscation from Trilinear Maps and Block-Wise Local PRGs*, Cryptology ePrint Archive, Report 2017/250, 2017.

[23] A.J. Menezes, T. Okamoto, S.A. Vanstone. *Reducing elliptic curve logarithms to logarithms in a finite field.* IEEE Transactions on Information Theory, vol. 39, no. 5, 1993.

[24] V. Miller. *Use of elliptic curves in cryptography.* CRYPTO 1985, LNCS 85, pp. 417–426, 1985.

[25] R. Sakai, K. Ohgishi, M. Kasahara. *Cryptosystems based on pairings.* Proceedings of Symposium on Cryptography and Information Security, 26–28, Japan, 2000.

[26] J.H. Silverman. *The Arithmetic of Elliptic Curves (2nd Edition).* Springer-Verlag Graduate Texts in Mathematics 106, 2009.

[27] A. Shamir. *Identity-based cryptosystems and signature schemes.* CRYPTO 1984, LNCS 196, pp. 47–53, 1985.

[28] F.L. Ţiplea, C. Drăgan. *Key-policy attribute-based encryption for Boolean circuits from bilinear maps.* BalkanCryptSec 2014, LNCS 9024, pp. 175–193, 2014.

[29] L. Washington. *Elliptic Curves: Number Theory and Cryptography (2nd Edition).* Chapman and Hall/CRC, 2008.

[30] B. Waters, S. Garg, S., C. Gentry, S. Halevi, A., Sahai. *Attribute-based encryption for circuits from multilinear maps.* CRYPTO 2013, LNCS 8043, pp. 479–499, 2013.

Ferucio Laurenţiu Ţiplea

Department of Computer Science, “Alexandru Ioan Cuza” University, Iaşi, Romania
Email: `ferucio.tiplea@uaic.ro`

Modeling of the movement of human flows in the process of evacuation from multi-storey buildings by Hierarhical Petri Nets

Inga Titchiev

Abstract

The aim of this article is to model the movement of different kinds of human flows in the process of evacuation from multi-storey buildings, taking into account the specificity of building construction. Being in multi-storey buildings, people exposure their life to various types of disasters and in this case the safety issue is one of the major importance.

Keywords: human flows, kinds of human flows, Petri nets, building construction.

1. Introduction

Exposing their life to the dangerous impact of the environment and the economic factors daily, the problem of the successfully evacuation from multi-storey buildings becomes an actual one. In the Republic of Moldova The Normative Supervision Section of Buildings and Fire Department exist, which performs the activity in the field of normative supervision in construction [5], taken into account the specificity of building construction. In order to evaluate the evacuation parameters we will propose to use the formalism of High Level Petri Nets (HLPN).

Two issues: Modeling and analysis can be solved by High Level Petri Nets. Modeling concerns the abstracting and representing the systems under consideration using HLPNs, and analysis deals with effective ways to study the behaviors and properties of the resulting HLPN models.

2. High-Level Petri Nets

High Level Petri [1, 2] nets provide a good framework for the design, specification, validation, and verification of evacuation system. They support data and functionality definitions, such as using complex structured data as tokens and algebraic expressions as transition formulas. Modeling and simulation can be used in study of the mitigation of consequences of disaster, because they are the tools and methods that are effective and efficient.

Animation and gaming are the other two rapidly growing fields associated with HLPNs used in order to monitor and study these processes in a realistic and interactive environment.

Definition 1 [2]. A High-level Petri Nets is a structure $HLPN = (P; T; D; Type; Pre; Post; M_0)$:

- P is a finite set of elements called *places,*
- T is a finite set of elements called *transitions* $(P \cap T = \emptyset;)$.
- D is a non-empty finite set of non-empty domains where each element of D is called a *type*.
- $Type: P \cup T \rightarrow D$ is a function used to assign types to places and to determine transition modes.
- $Pre; Post: TRANS \rightarrow \mu PLACE$ are the *pre* and *post* mappings with

$$TRANS = \{(t; m) \mid t \in T; m \in Type\ (t)\}$$
$$PLACE = \{(p; g) \mid p \in P; g \in Type\ (g)\}$$

- $M_0 \in \mu PLACE$ is a multiset called *initial marking* of the net.

In [6] we analyzed and modeled the evacuation scenarios for people in case of disaster, and in Figure 1 there is the transition from the evacuation plan to the Petri net representation.

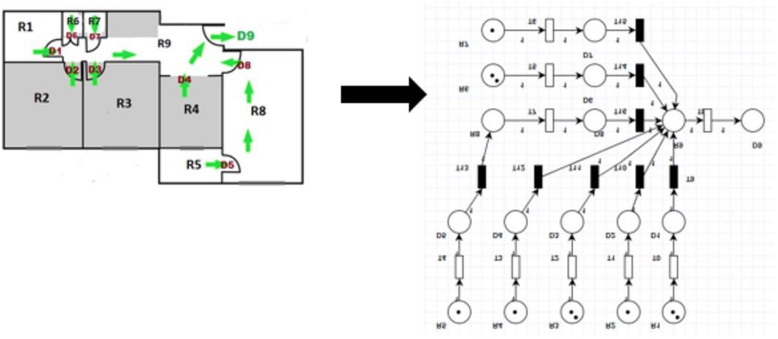


Figure 1. Transition from the evacuation plan to the Petri net representation.

3. Movement of human flows

The Normative Supervision Section of Buildings and Fire Department performs the activity in the field of normative supervision in construction in the Republic of Moldova. Main cases [7] of the movement of the human flows are:

3.1. Crossing the border of an adjacent room of the road

The *boundary* of the adjacent section of the road is the section of the path where its width or appearance changes, as well as the number of people, with equal width and length of the section.

Traffic intensity in rooms - $q_{i+1} = \frac{q_i b_i}{b_{i+1}}$, b_i – rooms length.

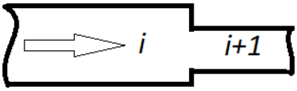


Figure 2. Crossing the border of an adjacent room

3.2. Fusion of human flows

The *fusion of human flows* is the process of forming a flow with combined parameters when different human flows are connecting.

Traffic intensity in rooms - $q_{i+1} = \frac{\sum q_i * b_i}{b_{i+1}}$

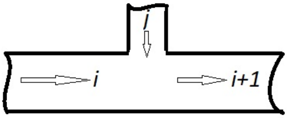


Figure 3. Fusion of human flows.

3.3. Formation of clusters of people (crush) and crush in movement

A *cluster* of people is formed if more people are approaching the unit of time *i* in a unit of time than can miss the next *i +1*. Then, before the boundary of section *i + 1*, a part of people is delayed, which at subsequent times increases. At the border of adjacent areas a congestion of people, the so-called *crush*, is formed.

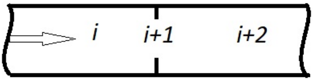


Figure 4. Formation of cluster of people.

3.4. Diffusion of human flow

The phenomenon of *diffusion* consists in the fact that when a cluster with a maximum density is formed before the boundary of the adjacent section of the path, the density in the subsequent section is much less. When the flow of people through the border between the sites instantaneously restructures its structure, and its head, getting to the area *i + 1* occupies a large area along the path, then the actual speed is increasing and the previous intensity is retaining.

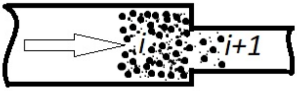


Figure 5. Diffusion of human flows.

4.5. Reforming of the human flow

When the human flows move along the sections of the path, it is very likely that the combined human flux has several zones with different density. *Reforming* the human flow is the process of equalizing the motion parameters in different parts of the flow. As a result, regardless of the initial parameters, each part of the stream acquires the parameters of the leading part. The speed of the transformation *V'* - the velocity of the boundary of the increase in the forward-moving part - is determined by the speed of the displacement of the boundary between parts of the flow with different densities.

The rate of reformation is determined from the relation: $V' = \frac{q_1 - q2}{D2 - D1}$.

Flow reform time is $t = \frac{\Delta l(D_1 - D2)}{D1(V2 - V1)}$, where *Di* – *i* is flow density, *Vi* – *i* is flow speed.

Increment of the length of the leading part of the flow is $\Delta l = \frac{N2}{b * D1}$.

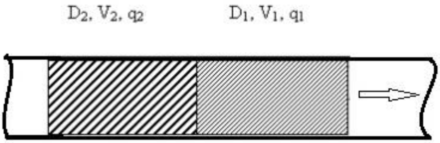


Figure 6. Reforming of the human flow.

Based on the above approach, we can simulate the movement of people modeling a steady flow by means of HLPNs and indicating the distance between floors, the number of people on each of the floors, the time of arrival of the first people from the floors to the staircase. We can obtain full evacuation time.

5. Conclusion

An approach to ensure the safe evacuation of people from multi-storey building in case of disaster modeled by Hierarhical Petri Nets has been proposed. It takes into account different kinds of movement of the human flows. This allows us to verify the safe and successful evacuation of people from multi-storey buldings.

Acknowledgments. 17.80013.5007.01/U, bilateral project Development of a toolkit for modeling strategies to mitigate social disasters caused by catastrophes and terrorism has supported part of the research for this paper.

References

[1] X. He, T. Murata. *High-Level Petri Nets: Extensions, Analysis, and Applications, Electrical Engineering Handbook* (ed. Wai-Kai Chen), Elsevier Academic Press, pp. 459-476, 2005.

[2] K. Jensen, G. Rozenberg. *High-level Petri Nets: Theory and Applications.* SpringerVerlag Eds., p.724, London, UK, 1991.

[3] V. Komyak, A. Danilin. *Approaches to the simulation of the motion of human flows in the building and their comparison.* Proceedings of the Problems of fire safety. Edition 35, pp. 110-115, 2014.http://nuczu.edu.ua/sciencearchive/ProblemsOfFireSafety/vol35

[4] S. Cojocaru, M. Petic, I. Titchiev. *Adapting Tools for Text Monitoring and for Scenario Analysis Related to the Field of Social Disasters,* In the proceedings of The 18th International Conference on Computer Science and Electrical Engineering (ICCSEE 2016), October 6-7, Prague, Czech Republic, pp. 886-892, 2016.

[5] Gh. Cojusneanu. *Normativ in constructii*. Siguranta la incendii, NCM E.03.02 - 2001, p. 55, Chisinau, 2001.

[6] I. Titchiev. *Modelling and verification of evacuation system using Time Petri nets in case of disaster*, Proceedings of the 18-th International Conference System Analysis and Information Technology (SAIT 2016), May 30 June 2, Kyiv, Ukraine, pp. 46-47, 2016.

[7] Fire risk, http://fireevacuation.ru/raschet-potoki-lydskie.php

Titchiev Inga[1,2]

[1]Institute of Mathematics and Computer Science, Chisinau, Republic of Moldova

E-mail: inga.titchiev@gmail.com

[2]"Dimitrie Cantemir" State University, Chisinau, Republic of Moldova

Supporting Democracy Through Cryptographic e-Services

Denis Trček[*]

Abstract

Electronic services (e-Services) have a notable potential to foster democratic values in emerging countries, while in developed countries they can further promote and protect them. In both cases democracy is particularly endangered through irregular settings and unacceptable practices like corruption. And cryptography based electronic services can play a significant role here. By deploying the infrastructure that is already in place, the proposed solution in this paper is using cryptographic protocol to ensure fair treatment in formal settings. In particular, the paper focuses on fair treatment in courts, where uncovered relationships may severely degrade the judicial process. However, the proposed solution can be adapted to support various other public services. Further, the solution can be fully automated and enables such implementations that are aligned with legislative requirements enforced nowadays by many countries (digital-signature and privacy related laws).

Keywords: information technologies, cryptographic protocols, e-services, democracy.

1. Introduction

Electronic services are penetrating all areas of our lives since the inception of e-Business era in the mid-nineties of the former century. Today they are playing an increasingly significant role not just in business, but also in our social and private settings. Their main advantages are not only lower costs, increased accessibility and availability compared to traditional services, but also (so far often unobserved) potential to support democracy. One example of such kind is fairness in every judicial system. In particular, one basic requirement in case of courts is that no personal relationship (be it intimate, professional, etc.) between the involved parties interferes with judicial treatment and procedures.

And this is where the main contribution of our short research paper comes in. It uses advanced information technology with cryptographic services and focuses on courts to prevent corruption-like scenarios. But it should be noted that the solution can be easily generalized and adapted to many other areas where legally and ethically questionable practices have to be prevented.

The paper is structured as follows. There is a detailed description of the problem and the proposed solution in the second section. In the third section there is a short discussion, followed by conclusions in the fourth section. The paper ends with acknowledgement and references.

2. The Problem Statement and e-Service Solution Description

The core contribution of this paper and its proposed solution is inspired by one similar privacy preserving solution that we developed for medical environments a few years ago [1]. At the core of our solution there are cryptographically strong one-way hash functions (one such widely known and used function is Secure Hash Algorithm version 3, or SHA-3 for short). As a result of using such one way hash functions, the proposed solution is computationally non-intensive, while providing additional benefits like privacy preservation.

The main idea behind the solution goes as follows. When in a certain formal setting a problematic relationship exists, a kind of a “generalized alarm” is raised, but it is raised in such a way where it is not possible to infer which particular relationship is problematic. Consequently, a new setting of involved parties (a new “team”) can be formed that is ethically and / or legally not problematic.

The more detailed description follows. Suppose there is a plaintiff who is claiming a financial compensation from a physician because of her medical mistake. However, there exists a problematic relationship that may remain completely uncovered. The judge is a patient of the accused physician, and nor the judge nor the physician want to disclose this relationship (which, in principle, may remain hidden as it is a matter of privacy protection and special treatment of relationships between physicians and their patients).

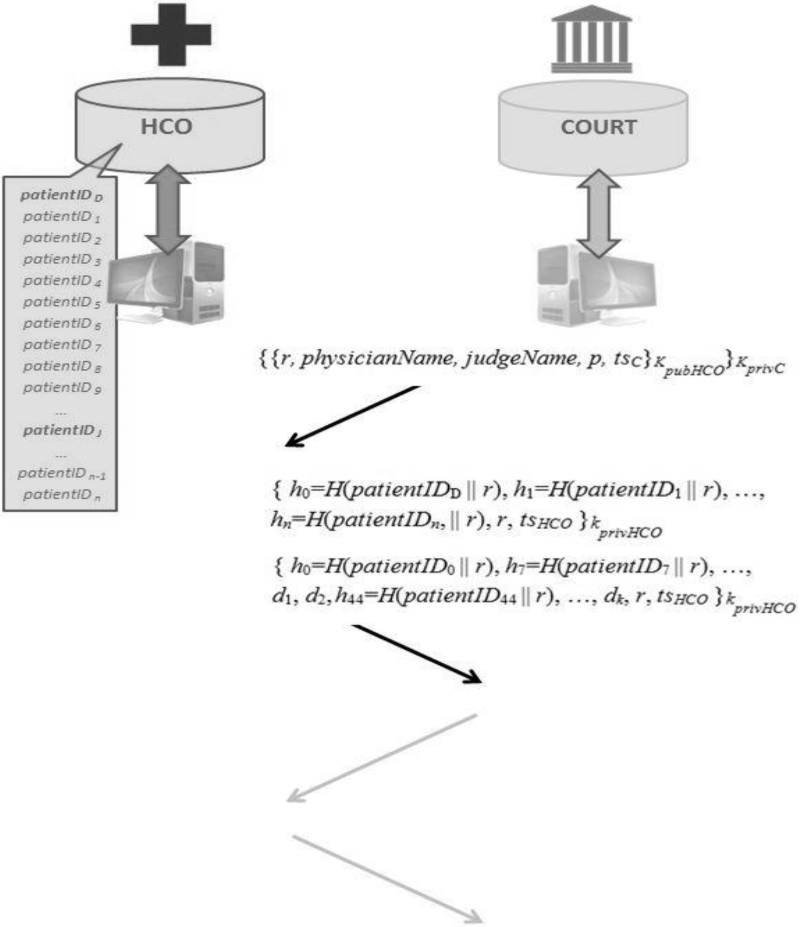

HCO
COURT
$patientID_D$
$patientID_1$
$patientID_2$
$patientID_3$
$patientID_4$
$patientID_5$
$patientID_6$
$patientID_7$
$patientID_8$
$patientID_9$
...
$patientID_J$
...
$patientID_{n-1}$
$patientID_n$
$\{\{r, physicianName, judgeName, p, ts_C\}_{K_{pubHCO}}\}_{K_{privC}}$
$\{ h_0=H(patientID_D \| r), h_1=H(patientID_1 \| r), \ldots, h_n=H(patientID_{n}, \| r), r, ts_{HCO} \}_{k_{privHCO}}$
$\{ h_0=H(patientID_0 \| r), h_7=H(patientID_7 \| r), \ldots, d_1, d_2, h_{44}=H(patientID_{44} \| r), \ldots, d_k, r, ts_{HCO} \}_{k_{privHCO}}$

Figure 1: The example case presented with the involved parties' IDs (dummy traffic that further prevents attacks is denoted by grey arrows).

How to tackle the above described problem? Each physician has a corresponding population of patients - in the primary care segment their number may be well over one thousand. No doubt that among such patients' population of a certain physician there may likely be a few judges. Therefore let's assume that the defendant is a personal physician of the judge. Due to assurance of privacy in medical relationships the involved physician (denoted as D, the defendant) remains silent, while the judge (denoted as J) does the same (the plaintiff is denoted as P, the attorney of P is denoted as AP, and attorney of D is denoted as AD). Jury, in this case, does not need to be considered, which is the most common case in continental law jurisdictions.

Further, in medical information systems each user is linked to a kind of a unique identifier (ID). For example, this can be an electronic health-care record ID called Master Patient Index, which is based on the HL7 data [2, 3] and which is associated with patient's physician ID. And this presents the basis for the protocol that automates prevention of unacceptable settings, which goes as follows:

1. A court C produces a random value *r*, which has to be always generated afresh for every run of this protocol. Next, knowing that in the trial the involved person will be a physician D, the court uses the name of this physician and the name of the judge J to form a message for a health care organization (HCO), which includes also the random value *r*. Finally, the court includes a parameter *p*, which serves to protect privacy as it will be explained below, and a timestamp ts_C. The court C encrypts all these values with the public key of HCO K_{pubHCO} and signs the message with its private key K_{privC}:

   $$\{\{r, physicianName, judgeName, p, ts_C\}_{K_{pubHCO}}\}_{K_{privC}}$$

   Then C sends this message to the health care organization, which takes care of medical documentation and corresponding medical IDs.

2. Next, after checking the digital signature and decrypting the received message (which contains the random value *r* and names of the judge and the doctor, and the parameter *p*) with its private key the HCO retrieves the requested IDs from its database (those that are associated with the doctor's ID denoted as $patientID_{\mathrm{D}}$). It appends each of these IDs with the received *r*, hashes the obtained string with a hash function *H*, digitally signs the whole list of hashed IDs with its private key $k_{privHCO}$ (together with a time-stamp ts_{HCO} and *r*):

   $$\{\ h_0{=}H(patientID_{\mathrm{D}} \parallel r),\ h_1{=}H(patientID_1 \parallel r),\ \ldots,\ h_k{=}H(patientID_{\mathrm{J}}, \parallel r, \ldots,\ h_n{=}H(patientID_n, \parallel r),\ r,\ ts_{HCO}\ \}_{k_{privHCO}}$$

   This is the first half of the message; the next one is formed as follows. HCO forms a randomly ordered sequence with *p* values. In a case where the judge is among the physician's patients, one value has to be the hashed ID of the physician. When the judge is not among the physician's patients, all values are randomly chosen dummy values *d*:

   $$\{\ d_1,\ d_2,\ \ldots,\ h_0{=}H(patientID_{\mathrm{D}} \parallel r),\ \ldots,\ d_{p\text{-}1},\ r,\ ts_{HCO}\ \}_{k_{privHCO}}$$

   Now both halves are sent to the court.

3. Upon receipt of these two messages, C sorts all the values from the first halve of the message to obtain a sorted list.

4. Next, C takes the first element from the second halve of the received message and looks for a match with elements from the sorted list. It repeats this procedure as long as there are elements in the second halve of the received message (i.e., for *p* steps). When there is a match, a general alarm is raised, and a new court setting has to be chosen.

Clearly, if the checking procedure like the one above would be performed in a traditional way, this would mean significant costs in terms of time and personnel needed, not to mention the paperwork required to preserve the necessary proofs. Without computerized lightweight protocol like the one above such procedure is close to impossible. But current technology changes the situation significantly. Databases with the required data are already available, the same holds true for public key infrastructure, while the computational power for such a procedure like the one above is abundant. By linking accordingly the given determinants, the above e-Service brings an important advantage in terms of fairness of formal settings, privacy protection, and provides verifiable proofs needed in case of disputes. Moreover, as the whole service causes only negligible costs, tens of thousands, millions, and even more such verification procedures can be done simultaneously to easily meet everyday's needs in any democratic state.

3. Brief Analysis

Let us now make a brief and informal analysis of privacy provisioning of the above crypto protocol, as privacy nowadays is generally required by legislation [4]. Actually, none of the IDs or names is disclosed, because they are hashed and these hashed values are updated with a fresh random value *r* in every round of the protocol. If an attacker finds matches, he or she cannot be sure which match belongs to a particular physician and which to a particular judge.

Further, the produced messages are encrypted (hashed) and digitally signed accordingly by deploying PKI [5]. Consequently, formal investigations are enabled in cases, where, for example, HCO is cheating and not providing the right inputs. Last but not least, the whole solution can be easily programmatically implemented by using the e-documents and e-data infrastructures that are already widely available in an increasing number of states.

Further, let's assume that a physician has 10^3 patients. Further, let the value of *p* in the above protocol be 10^3 as well. Then the whole procedure approximately requires 10^3 reading operations from a database, 10^3 concatenations with a random value *r*, and 10^3 hashes. Further, $p = 10^3$ random values have to be generated. When the message from the second step is received, the first half has to be sorted, and using an algorithm like Quick sort with $O(n) = n * \log n$ time complexity, one would need approximately $3*10^3$ operations, and in the worst case $3*10^3$ comparisons to check for a match between the elements in the first half of the second message with the elements from the second half of the second message (assuming that a binary search is used for each element in the second half, for which the $O(n) = \log n$). If we neglect digital signing costs and assume IDs to be less than 512 bits long, the most demanding operation is hashing (which in principle depends on the length of the input, clock speed and the processor type). But based on credible sources like [6] it can be concluded that nowadays even smart phones are powerful enough to easily handle such tasks as the one required by the above introduced e-Service.[1]

And finally, the above protocol can be notably improved with enabling dummy traffic. Actually, an additional field in the first message, when set to one, would indicate to the receiving party (i.e., HCO) that the message has to be processed accordingly as it is about a real request to check a certain setting. If this additional field in the first message is set to zero, this would indicate to a HCO that it can reply with a bogus, completely random reply message, which will serve for traffic padding. In this case entropy of the service is increased, because an attacker now does not know when a setting is really being checked and when not. Consequently, this would also increase the computational burden for the attacker.

4. Conclusions

This paper presents an e-service solution, which ensures fairness of relationships in formal settings - in our particular case this is in courts. The solution enables to uncover existing problematic relationships, because such relationships may seriously affect the fairness of judicial process and erode trust in one key pillar of democratic societies. But the presented s-Service solution is not limited to court cases. It can be easily adapted in a rather straightforward way to other situations where preventing corruption practices are necessary to ensure well-functioning democratic societies.

[1] A good estimate of hashes/s with today's techonolgy give Bitcoin mining data – see, e.g. https://en.bitcoin.it/wiki/Non-specialized_hardware_comparison.

From the technological point of view the solution presented in this paper is a modest step forward. However, from the society benefits point of view it presents a notable advancement, because it can be easily implemented with the modern information technology and can directly use the existing information technology based infrastructure that is already in place in majority of states. In addition, its cost is almost negligible as the required computational power for its implementation can be nowadays met by an ordinary smart phone. Last but not least, the solution is aligned with existing legislation in many countries, in particular in the EU. This alignment spans from privacy protection of the involved entities to deployment of formally valid exchange of digitally signed documents by using PKI. And finally, the solution is aligned with the general directions of many governments related to digital societies, in particular the one accepted by the EU Commission [7].

Acknowledgment. The author would like to thank the Slovene research agency ARRS for supporting this research through the research program Pervasive computing.

References

[1] D. Trček and A. Brodnik. *Hard and soft security provisioning for computationally weak pervasive computing systems in E-health.* IEEE Wireless Communications, vol. 20, no. 4 (2013), pp. 22-29.

[2] P. Schloeffel, B. Thomas, G. Hayworth, H. Sam, L. Heather. *The Relationship between CEN 13606, HL7, and OpenEHR*. Proc. of the HIC 2006 and HINZ ’06, Health Informatics Society of Australia, 2006, pp. 24-28.

[3] HL7 Deutschland. *Segment PID (Patient Identification Segment)*. http://wiki.hl7.de/, last accessed on May 6, 2018.

[4] EU Commission. *Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation).* Official Journal of the European Union, L119, 4/5/2016, pp. 1–88, 27.th April, 2016.

[5] EU Commission. *Community framework for electronic signatures*. Directive 1999/93/EC of the European Parliament and of the Council, Official Journal of the European Union OJ L 1, 19th January, 2000, pp. 12-20.

[6] C. Shu-jen et al. *Third-Round Report of the SHA-3 Cryptographic Hash Algorithm Competition*. NISTIR 7896, NIST, 2012.

[7] EU Commission. *Commission and Its Priorities*. Creating a Digital Society, https://ec.europa.eu/digital-single-market/en/policies/creating-digital-society, last accessed on May 6, 2018.

* Faculty of Computer and Information Science
University of Ljubljana
Večna pot 113, 1000 Ljubljana
Slovenia / EU
E-mail: denis.trcek(at-sign)fri.uni-lj.si

Table of contents