I T H E A

International Journal

INFORMATION THEORIES & APPLICATIONS

2014 Volume 21 Number 1

International Journal
INFORMATION THEORIES & APPLICATIONS
Volume 21 / 2014, Number 1

International Journal "INFORMATION THEORIES & APPLICATIONS" (IJ ITA)
is official publisher of the scientific papers of the members of
the ITHEA International Scientific Society

IJ ITA welcomes scientific papers connected with any information theory or its application.
IJ ITA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITA are given on *www.ithea.org*.

Responsibility for papers published in IJ ITA belongs to authors.
General Sponsor of IJ ITA is the **Consortium FOI Bulgaria** (*www.foibg.com*).

International Journal "INFORMATION THEORIES & APPLICATIONS" Vol. 21, Number 1, 2014

Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with:
Institute of Mathematics and Informatics, BAS, Bulgaria,
V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,
Universidad Politécnica de Madrid, Spain,
Hasselt University, Belgium,
St. Petersburg Institute of Informatics, RAS, Russia,
Institute for Informatics and Automation Problems, NAS of the Republic of Armenia

Printed in Bulgaria
Publisher ITHEA®
Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com
Technical editor: **Ina Markova**

ISSN 1310-0513 (printed) **ISSN 1313-0463 (online)** **ISSN 1313-0498 (CD/DVD)**

ANALYZING THE COLLECTIVE INTELLIGENCE APPLICATION SOFTWARE "WISDOM PROFESSIONAL" FOR ADVERTISING IN (SOCIAL) MEDIA, CASE STUDY: COCA-COLA

Elham Fayezioghani, Koen Vanhoof

Abstract*: The main goal of this article is set to demonstrate how collective intelligence application software works and can be used for advertising purposes. Following this objective, it is aimed to optimize advertising in (social) media for a company or enterprise by using collective intelligence software. To this end, the Wisdom Professional software from MicroStrategy Company is utilized to derive required information. These data are kind of raw data that are derived from the information of Facebook users that are collected in Wisdom Professional. Wisdom Professional includes Facebook information of millions of people such as demographic, geographic, psychographic information and so on. The information of Facebook users in Wisdom Professional are considered to be representative information of the whole population in a sense that what is observed in this sample (i.e. Facebook users in Wisdom Professional) can be extended and generalized for the whole customer population. The data derived from Wisdom Professional are further processed by means of different methods to demonstrate the best possible advertisement options for a specific enterprise. The two utilized methods are Popularity oriented and lift factor methods. The results of analyses reveal that both methods provide with reliable and consistent outcome. As for the application part of this article, Coca-Cola Company is considered to be the chosen case study. Coca Cola fans' information are analyzed in Wisdom Professional to provide best possible advertising channels that this company could benefit by advertising its product in those channels.*

Keywords*: collective intelligence, Wisdom Professional, advertising, social media*

ACM Classification Keywords*: H.4.2 Information Systems Applications - Types of Systems - Decision support*

Introduction

In this article it is aimed to analyze collective intelligence application software. This software is called Wisdom Professional and is a product of MicroStrategy. In this article the objective is to evaluate the situation of a company (here Coca Cola) by means of Wisdom Professional and provide the company with some helpful information that they can use to improve their business. The results of this type of analysis will empower organizations to make better business decisions, realize their status among competitors and ultimately improve their organization's efficiency through analyzing interests of millions of people on Facebook. It is very important for an organization, enterprise or a brand to know everything about its customers or target groups. The results of this article can help them to get a better insight into the position of their competitors and their status in market, products and partners for gaining competitive advantage. Therefore, the main motivation on carrying out this research is to show that adopting such business intelligence software could result in better decision making for a company and furthermore acting in a collective manner is more intelligent than deciding individually.

In this article, we aim at optimizing advertisements in Social media for a company or enterprise by means of collective intelligence software. Figure (1) shows the three major contribution factors of this article. Adopted software is Wisdom professional from MicroStrategy Company. Target Company is Coca-Cola and for the

comparison purposes some of its competitors and other Coca-Cola's brands (e.g. Fanta, Sprite and etc.). The study area is also selected to be Europe.

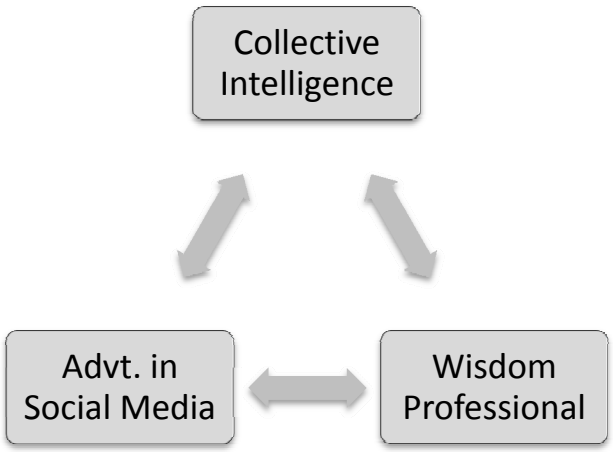


Figure (1): Relation between 3 key components in this article

Literature review

Collective intelligence

Collective intelligence means using more than an individual intelligence for achieving complex goals. The basis of this approach relies on the fact that a group of people is more intelligent than an individual [Surowiecki, 2005]. The crowd can collectively do something such as solving problems or recognizing patterns better than machines [Leimeister, 2010].

In general collective intelligence helps organizations to improve business outcomes through accessing to untapped knowledge and experience of their networks [IBM].

Collective intelligence could be categorized and defined as follows:

- The collective intelligence resulted from interactions among different people with diverse knowledge working together;
- The collective intelligence created by independent customers in a market;
- The collective intelligence of global information systems that can be achieved by means of computers (The Co-Intelligence Institute).

Social theories

The classical social theory can help to have a better recognition of social media, collective intelligence and the relation between them. Merton (1967) declares in his social theories that humans need to act reciprocally, to learn from each another, share and exchange ideas to enhance level of knowledge and working together as a group to make better and more effective decisions [Merton, 1967]. In social capital framework, Bourdieu (1986) declares that social networks are an origin of capital that made up of social obligation and "social capital is actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition or in other words, to membership in a group" [Bourdieu, 1986]. The social value of a group improves when they think and act collectively. Field's (2003) social capital theory position is defined in a way that relationships matter and social networks are valuable assets where people develop communities and commit themselves to each other. The human experience of trust and tolerance bring benefits to people in the network including mutual understanding [Nickel, 2013].

Link between social media and collective intelligence

Social media and collective intelligence have close relation to each other. Using internet and online collective communication vehicles such as web sites and web pages, people can interact together and create, share and exchange their information and wisdom in virtual spaces. In other words, people can understand what other persons are thinking and feeling. This means using collective intelligence in social media. Two key examples are Wikipedia and Facebook.

In Facebook there is much information about users. This information includes demographic information such as age, gender, marital status, level of knowledge and so on and interests, idea or opinion of users. For example a media planner wants to plan for Coca-Cola Company. The goals he follow them are: which vehicles are more appropriate (TV, magazine, internet, …) to advertise? And when and where is better for advertising? He can uses software like Wisdom Professional that it contains many information about users of Facebook. Now he should use needed Facebook information with regards to Coca-Cola plan to make decisions.

MicroStrategy and Wisdom Professional

MicroStrategy is founded in 1989. It is a global provider of enterprise software for business intelligence (BI), mobile intelligence and social intelligence (i.e. Wisdom Professional software) applications. It provides reporting, analyzing and monitoring that enable organizations make business decisions better than before [Wisdom].

Wisdom Professional provides analytical application of existing data of Facebook users. It has many capabilities such as Dashboard, Demographics, Interest analysis, Place analysis, Psychographics, Scores and Comparison (compares several pages simultaneously). In this article main analysis is done based on Interest analysis capability (Figure (2)).

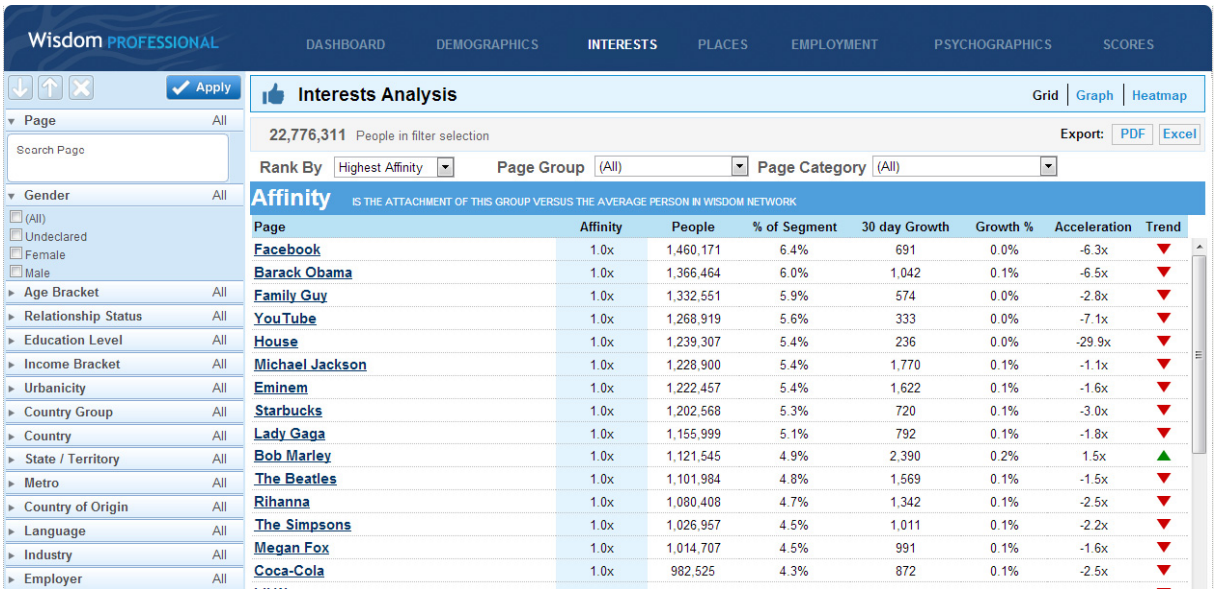


Figure (2): A snapshot of Interests tab in Wisdom Professional

Interest analysis

In interest analysis, Interests are divided into 14 groups such as companies/products, sports, music, movies, people, TV, books, games/Apps, art, general entertainment, going out, news/media, travel and other and related page categories. In this article, seven interests are selected namely TV, news/media, companies/products, music, movies, books and games/Apps. These interest categories are selected because they are the top interest categories with highest number of people in Wisdom Professional.

There are 5 main metrics that can be referred to for interest analysis, namely Affinity, People, % of segment, Growth and Acceleration [Wisdom]. Here we worked with two metrics, People and % of segment.

The People represent the number of people who are fans of a particular page. The % of Segment shows the percentage of fans of a specific page in the selected area.

Wisdom data of Coca-Cola fans in Europe

In Wisdom database 160369 people who live in Europe like Coca-Cola. Demographic information of Coca-Cola fans in Europe is as follows: average age is 26, average yearly income is $41k, in gender 52% are male and 48% are female, in marital status 47% are single, in education level 50% are holding at least a college degree, 76% live in urban areas [Wisdom].

Methodology

The problem we aim to solve in this article is optimizing media advertisement for companies. Due to the limited budget and time, companies cannot address their advertisements in all media. They need to advertise in most important and effective media that gives them the most benefit. As a solution for this problem, one could think of employing collective intelligence. Collective intelligence is a powerful method that gathers useful information by collectively gathers revealed preferences of a large scale population. A large scale population's decision is always more reliable than individual's decisions. For this purpose we aimed to analyze and use collective intelligence software namely "Wisdom Professional". This is the tool that systematically gathers Facebook users' opinions. In fact Wisdom Professional could be considered as a small community which appropriately resembles the real world and, therefore, its information can be generalized to the global population.

In order to analyze the information revealed from Wisdom Professional two major types of data are being collected. These data are basically population-based or penetration-based measures. Population-based measures are the ones where we exclusively consider the number of fans of a product but penetration-based measures deal with the relevant figures. In Wisdom Professional there are many different channels such as TV, News/Media, Companies/Products, Movies, Music, Books and etc. where people's interests are categorized in those channels. Each channel includes many cases which are basically Facebook pages. By having all those channels and their embraced cases, one could easily analyze a specific product (here in this article Coca-Cola) and different interests of this products fans. In this article and due to self-imposed constraint we focus on the top 10 cases of interest in any kind of analysis. In this way, we make sure to focus on the most popular and influential cases that gives us the most efficient advertisement opportunities.

The first method (i.e. Method A) deals with different approaches through which different raw numbers from Wisdom Professional are collected and further processed in order to achieve the goals of this article. These numbers are calculated to show which advertising channel is the most suitable one and via which channel Coca-Cola Company can target its customers more effectively. Method A mainly considers the number of fans of each case; however, Method B deals more with the penetration rate of different cases. In what follows a brief description of both two methods is illustrated.

Method A; Popularity oriented method

To achieve the best possible advertising method, there are 3 different approaches via which the best advertisement place can be identified. These approaches are the following:

1. Single product with single channel;
2. Single product with multiple channels;
3. Multiple products with single channel.

Here product means a type of drink (e.g. Coca-Cola), channel refers to the category of interests (e.g. Media, TV, Companies, etc.) and case or page refers to cases included in each channel (e.g. The Simpsons, House, etc. in TV channel).

Single product with single channel: in simple words, we should correlate only one product like Coca-Cola with only one channel and then find the top ten cases of this specific channel based on the products preferences from Wisdom Professional. These cases will be analyzed in terms of number of fans and the results will be further processed and reported accumulatively and in percentage. To limit our choices and to emphasize on the most beneficial cases for doing advertisements we highlight the cases by which 80% of top ten cases fans are already covered. The aim is to express which pages have more visitors in each channel and for each product. Details of this procedure are mentioned as follows.

Single product with multiple channels: This approach is the same as the first approach except of the fact that there are multiple channels such as TV, News/Media, companies/products or any other possible channels all together. When considering all possible channels together, the two considerably most favorite pages of Coca-Cola fans are YouTube and Facebook; these two pages cover the most number of Coca-Cola fans in comparison to any other pages.

Multiple products with single channel: Here the combination of two products is analyzed in correspondence with a single channel. There are 2 different methods applicable for this approach. The first method is appropriate for channels that have many similarities in their programs when combining Coca-Cola and the comparison product. In this method we can select top 20, 15 or 10 of Coca-Cola in a specific channel like TV. Then we should rank the other products based on Coca-Cola priorities.

The second method is suitable for channels in which Coca-Cola fans have several dissimilar favorite pages in comparison to their competitor product fans. In this method, products are mutually selected where always one party is Coca-Cola (e.g. Coca-Cola with Heineken or Coca-Cola with RedBull). Then top 10 of Coca-Cola and top 10 of the other product favorite channels are selected. At this stage, all repeated channels form this 20 channels list are kept once. At this stage the same procedure will be employed to distinguish the number of channels by which 80% of Coca-Cola fans will be covered.

Method B; Lift factor method

In data mining, lift is a quantity of the performance of a targeting model at predicting cases as having an improved respond (with respect to the total population as a whole), measured against a random choice targeting model [SQL Server Microsoft]. A targeting model is doing a good job if the response within the target is much better than the average for the population. In other words, lift factor is simply the representation of this ratio: target response divided by average respond. Lift factor shows how companies can prepare for their advertising plans by considering four main elements; 1) whole population of the study area (e.g. Europe zone), 2) number of specific product's fans in the study area (e.g. number of Coca-Cola fans in Europe), 3) total number of fans of a target channel that the company wants to advertise in it (e.g. "MTV" from TV cases) and 4) the percentage of a channel fans who like a specific product (e.g. percentage of "MTV" fans who like Coca-Cola).

$$lift\ factor\ = \frac{\frac{\text{Coca} - \text{Cola fans of MTV in Europe in Wisdom}}{\text{total fans of MTV in Europe in Wisdom}}}{\frac{\text{total Coca} - \text{Cola fans in Europe in Wisdom}}{\text{total Wisdom population in Europe}}}$$

For example, imagine that a population has an average respond rate of 6.6% and a specific model has recognized a segment with a respond rate of 53%. Then lift of this segment will be 8. Above is an example of lift factor for MTV which is among the top 10 Coca-Cola cases in TV.

This should be noted that these lift factors are different from the affinity measures available in Wisdom Professional since for the calculation of affinity the geographical choice (i.e. here Europe as the case study area) is not considered and the calculation is made based on the whole population of Wisdom.

Results

In this section the results of the two employed approaches will be explained. For a better representation, results will be mainly expressed in tables and graphs. Using visual aids such as graphs and tables is an effective way for an easy understanding of the results. First approach contains three parts, namely single product - single channel, single product - multiple channels and multiple products - single channel. The objective of this part is to demonstrate which channels in each category cover 80% of fans within top 10 favorite channels. Furthermore, the results of the second method (i.e. Lift factor method) are demonstrated.

Single product - single channel

Single product - single channel means that the analysis considers one product like Coca-Cola, Red Bull, Pepsi and etc. and one channel like TV, News/Media and etc . Cases presented in Table (1) are top 10 Coca-Cola fan's TV category priorities and are ranked base on their popularity. The numbers of fans for each product are mentioned accumulatively. All numbers are derived from Wisdom Professional. The 80% threshold are calculated by multiplying 0.8 by the last number (i.e. 10th number which represents the number of Coca-Cola fans who like at least one of those 10 cases) of each product. For instance, 80% of top Coca-Cola fans are covered by any advertisement prepared for the first three TV programs, namely "The Simpsons", "House" and "South Park". For a better understanding of these results, Table (1) is followed by its corresponding graph.

Table (1): Single product - single channel=TV in Europe

		Coca-Cola	Red Bull	Pepsi	Fanta	Sprite	Heineken
1	Simpsons	63100	46664	5069	10050	6920	11094
2	House	87222	60961	6440	11451	8484	17620
3	South park	94667	66975	6909	11895	9100	19946
4	MTV	101018	71295	7354	12489	9568	21244
5	Family guy	104300	74045	7566	12767	9721	22378
6	How I met your mother	107816	77485	7758	13002	9869	23902
7	Two and half man	109800	79709	7857	13116	9931	24964
8	Spongebob Squ.	111199	80450	7935	13228	10063	25189
9	Futurama	111802	80842	7968	13280	10091	25396
10	The big bang theory	112746	81504	8032	13346	10120	25856
80% of top 10 favorite cases		90196.8	65203.2	6425.6	10676.8	8096	20684.8

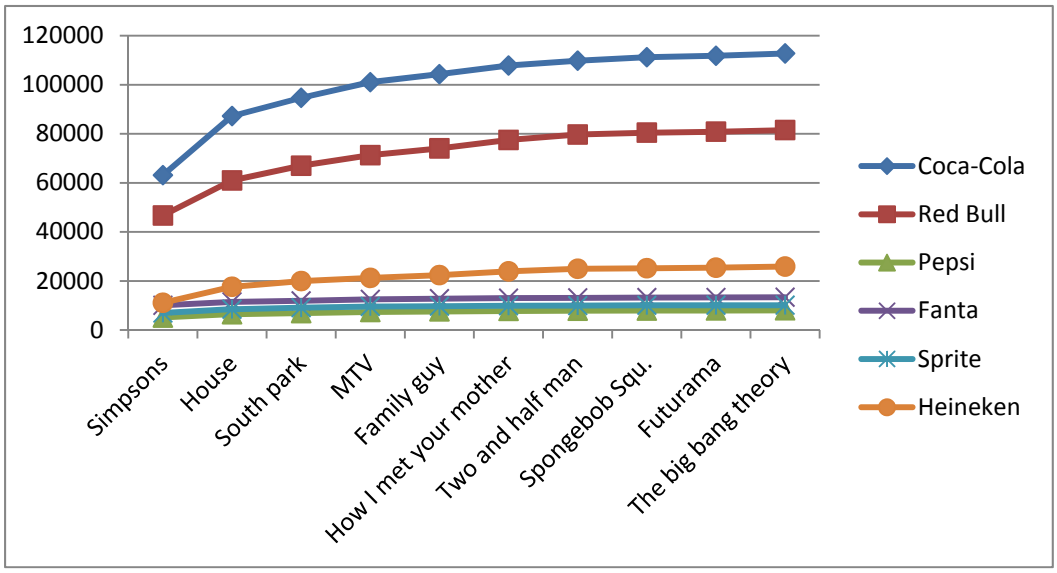


Figure (3): Single product - single channel=TV in Europe

Table (2) shows the absolute numbers of Table (1) in percentage. In Table (2) the numbers of each column in Table (1) are divided by total fans of each product. The aim is to show the percentage of each product's fans who like that specific product in comparison to its total fans. When only the absolute measures are considered, Coca-Cola is always on top because it has more fans in comparison with its competitors but when the rates are considered the orders will be different. This implies the fact that by advertising on a specific TV case, more fans (in terms of absolute numbers) but a smaller share of total fans (percentage of fans) might be exposed to that advertisement. The last percentage in each product's column shows the percentage of each specific product fans who like at least one of the top 10 TV cases. For instance, 72.50% of Coca-Cola fans like at least one of those top 10 TV cases. In fact what is reported as 80% of top 10 TV cases fans in Table (1), is 80% of this 72.50%. Table (2) is followed by its corresponding graph.

Table (2): Single product - single channel=TV in Europe, in percentage

		Coca-Cola	Red Bull	Pepsi	Fanta	Sprite	Heineken
1	Simpsons	40.57%	41.05%	49.24%	59.73%	59.37%	22.46%
2	House	56.08%	53.63%	62.56%	68.05%	72.79%	35.66%
3	South park	60.87%	58.92%	67.12%	70.69%	78.07%	40.37%
4	MTV	64.96%	62.72%	71.44%	74.22%	82.09%	43.00%
5	Family guy	67.07%	65.14%	73.50%	75.87%	83.40%	45.30%
6	How I met your mother	69.33%	68.16%	75.36%	77.27%	84.67%	48.38%
7	Two and half man	70.60%	70.12%	76.33%	77.95%	85.20%	50.53%
8	Spongebob Squ.	71.50%	70.77%	77.08%	78.61%	86.33%	50.98%
9	Futurama	71.89%	71.12%	77.40%	78.92%	86.57%	51.40%
10	The big bang theory	72.50%	71.70%	78.03%	79.31%	86.82%	52.33%
	Total fans	155519	113676	10294	16827	11656	49405

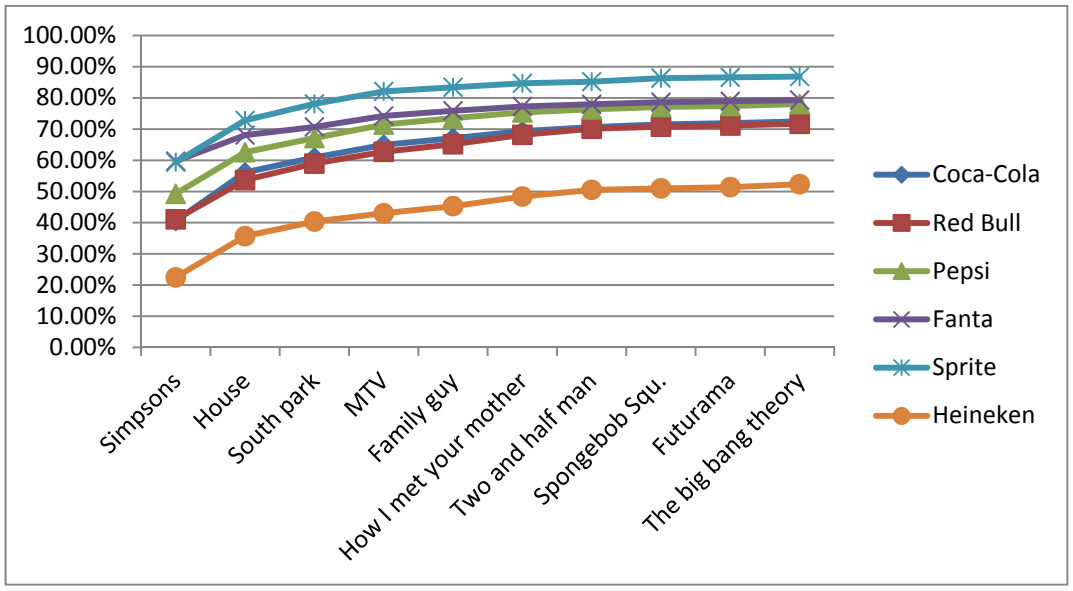


Figure (4): Single product - single channel=TV in Europe, in percentage

From the business point of view, when we cover fans of first three TV cases (80% of top 10 cases) in advertising for Coca-Cola, in fact we also cover fans of Red Bull, Pepsi, Fanta and Sprite which are Coca-Cola's competitors. This is due to the fact that fans of these products share a number of similar favorite cases. When considering the percentage of fans of competitor products the analysis reveals that for a couple of products, even a greater share of fans are exposed to Coca-Cola advertisement.

The aforementioned conclusion could lead to another business strategy which is usually referred to as cooperative advertising. Cooperative advertising is a cost-effective way for manufacturers and retailers to reach their target markets. If Coca-Cola strategy planners are well informed of the common favorite pages of their product and their competitors' products, they can easily set up a co-op advertisement where they can share both costs and benefits. This could sound more beneficial if Coca-Cola set up co-op advertisement with other Coca-Cola brands (e.g. Coca-Cola zero, Fanta and Sprite) or Red Bull since they do not exactly belong to the same category of drinks. On the contrary, it is almost impossible if they try to follow this co-op advertising strategy with Pepsi because they are exactly in the same category of drinks and are known to be the two opposing poles in the market.

Single product - Multiple channels

Single product - multiple channels approach means that one product like Coca-Cola, Red Bull, Heineken and etc. will be simultaneously analyzed with several channels like TV and News/Media and etc. The results of this analysis could give a great view of the fans of all possible pages together.

Cases in Table (3) are ranked base on top 10 Coca-Cola priorities. Number of fans of each product are noted accumulatively. As can be seen in Table (3), 80% of top 10 cases fans who like Coca-Cola are covered by the first 3 pages, namely "YouTube", "Facebook" and "The Simpsons". For a better understanding of this result, Table (3) is followed by its corresponding graph.

Table (3): Single product - multiple channels=all channels in Europe

		Coca-Cola	Red Bull	Pepsi	Fanta	Sprite	Heineken
1	YouTube	86154	57942	6760	11172	9106	14434
2	Facebook	103897	69729	7759	12749	9973	19378
3	The Simpsons	113984	78215	8230	13807	10464	22705
4	Rihanna	119707	82412	8512	14238	10724	24561
5	Michael Jackson	124830	85684	8714	14450	10883	26456
6	House	130524	89198	8920	14716	11046	28772
7	Red Bull	134013	116995	9146	15025	11181	30647
8	David Guetta	135756	116995	9208	15153	11243	31642
9	Eminem	136775	116995	9266	15242	11296	32039
10	Disney	138497	116995	9322	15369	11349	32516
80% of top 10 favorite cases		110797.6	93596	7457.6	12295.2	9079.2	26012.8

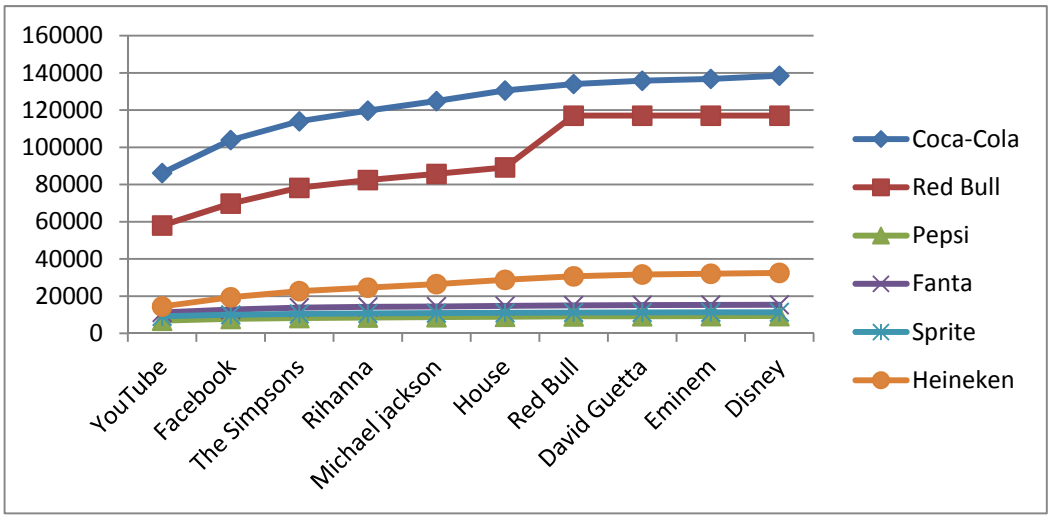


Figure (5): Single product - multiple channels=all channels in Europe

Table (4) corresponds to the numbers of Table (3) but in percentage. In Table (4) the numbers of each column in Table (3) are divided by total fans of each product. The aim is showing percentage of each product's fans that like specific case in comparison to its total fans. Table (4) is followed by its corresponding graph.

Table (4): Single product - multiple channels=all channels in Europe, in percentage

		Coca-Cola	Red Bull	Pepsi	Fanta	Sprite	Heineken
1	YouTube	54.00%	49.53%	64.73%	64.37%	76.33%	28.47%
2	Facebook	65.13%	59.60%	74.29%	73.45%	83.60%	38.22%
3	The Simpsons	71.45%	66.85%	78.80%	79.55%	87.72%	44.78%
4	Rihanna	75.04%	70.44%	81.50%	82.03%	89.90%	48.44%

5	Michael Jackson	78.25%	73.24%	83.44%	83.25%	91.23%	52.18%
6	House	81.82%	76.24%	85.41%	84.78%	92.60%	56.75%
7	Red Bull	84.00%	100.00%	87.57%	86.56%	93.73%	60.44%
8	David Guetta	85.10%	100.00%	88.17%	87.30%	94.25%	62.41%
9	Eminem	85.74%	100.00%	88.72%	87.81%	94.69%	63.19%
10	Disney	86.81%	100.00%	89.26%	88.55%	95.14%	64.13%
	Total fans	159532	116995	10444	17357	11929	50704

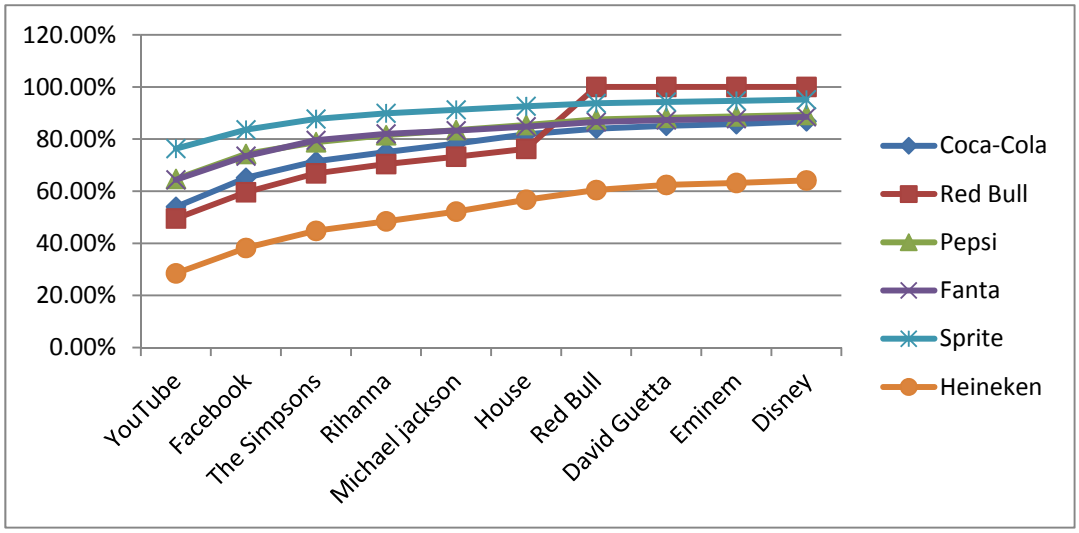


Figure (6): Single product - multiple channels=all channels in Europe, in percentage

Multiple products - single channel

Multiple products - single channel approach means that two products like Coca-Cola and Red Bull or Coca-Cola and Heineken are analyzed in correspondence with a single channel like TV, News/Media and etc. For a better understanding of differences between Coca-Cola and its competitor cases they are mutually analyzed. It is worth mentioning that favorite cases are ranked based on Coca-Cola priorities.

The first method is more appropriate for channels that have many similarities in their favorite cases comparing with Coca-Cola. Following this method, top 20 or 15 or 10 of Coca-Cola fans favorite channels are selected. Furthermore, competitor's cases are illustrated based on Coca-Cola's ranking. In this method, from the business point of view if the ranking line of a competitor product (e.g. Pepsi) for a specific case is on top of the baseline (i.e. Coca-Cola's line) it means that specific case is less favorable for that competitor product in comparison with Coca-Cola. Therefore, it can be concluded that top favorite Coca-Cola cases that their ranking line falls below the baseline are better potential candidates for advertising purposes. It is due to the fact that by targeting those cases, not only top favorite Coca-Cola fans are covered, but also higher number of competitor fans are exposed to the advertisement. These results can be considered as helpful tools for media planners to have a better understanding of each product/channel position; however, the choice of which case to do the advertisement in remains dependent of many other factors like company's advertisement policies, advertisement costs and judgment of decision makers and media planning experts.

In TV category, top 20 TV programs are very similar to each other; therefore, the first method is used for TV. More details are followed by Table (5) and one of the corresponding graphs (Red Bull).

Table (5): Multiple products - single channel=TV in Europe (first method)

		Coca-Cola	Red Bull	Heineken	Pepsi	Fanta	Sprite
1	Simpson	1	1	2	1	1	1
2	House	2	4	1	5	4	4
3	South Park	3	2	3	3	5	3
4	Family guy	4	3	4	2	2	5
5	MTV	5	5	7	4	3	2
6	How I met your mother	6	7	6	9	10	9
7	Two and half men	7	6	5	7	12	7
8	Spongebob sq.	8	9	18	6	6	6
9	Futurama	9	8	9	8	8	8
10	The big bang theory	10	10	8	10	14	12
11	Grey's anatomy	11	15	12	18	13	18
12	Friends (TV show)	12	14	10	11	17	11
13	CSI: Miami	13	13	15	12	9	10
14	American dad	14	11	17	13	7	13
15	Sex and the city	15	16	14	17	18	16
16	Scrubs	16	12	16	15	31	22
17	Gossip girl	17	17	22	21	21	17
18	Glee	18	19	34	14	11	14
19	NCIS	19	20	21	19	20	20
20	Desperate house.	20	21	24	24	27	28

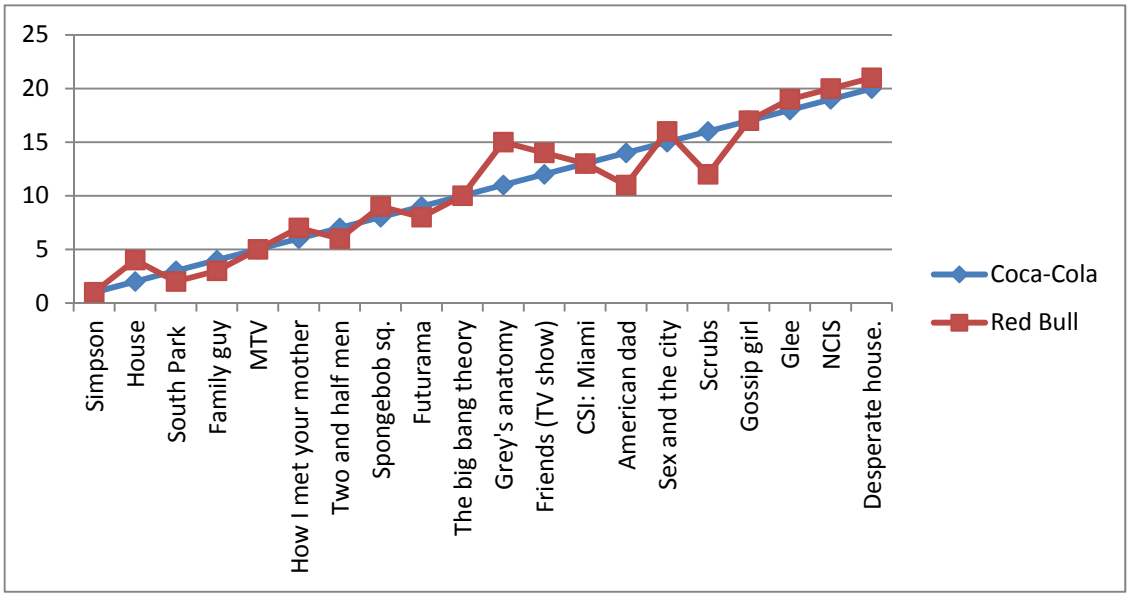


Figure (7): Multiple products - single channel=TV in Europe (first method), Coca-Cola with RedBull

The second method is suitable for products which have differences in their favorite cases compared with Coca-Cola (e.g. Heineken as an alcoholic drink). In the second method, top 10 favorite cases of any couple of products (i.e. Coca-Cola and other products) are selected and combined in one single table. Expectedly many of these top 10 favorite cases are not similar, however, there will be a number of cases which are favorable for Coca-Cola and the comparison product. In this step, all repeated cases are refined and only kept in the table once.

For example top 10 cases in News/Media channel for Heineken are not completely the same as the ones for Coca-Cola. First we select top 10 favorite cases of Coca-Cola and top 10 favorite cases of Heineken in News/Media category. Then similar cases should be separated and only counted once. As can be seen in Table (6), there are 15 cases left in the News/Media table. According to the measures in Table (6), 80% of top 15 cases fans of Coca-Cola are covered by first eight cases of News/Media. Table (6) is followed by its corresponding graphs. The first graph corresponds to columns 2 and 3 and second graph corresponds to columns 4 and 5.

Table (6): Multiple products-single channel=News/Media in Europe (second method)

		Coca-Cola	Heineken	Coca-Cola	Heineken
1	National geographic	23692	8030	15.12%	16.12%
2	Playboy	31555	11039	20.13%	22.15%
3	Ta' bonito	37313	11793	23.81%	23.67%
4	Patatine fri.	44530	12551	28.41%	25.19%
5	WWE	48626	12972	31.02%	26.03%
6	Publico	49857	13337	31.81%	26.77%
7	Sports on FB	52533	14196	33.52%	28.49%
8	The New York times	54368	14910	34.69%	29.92%
9	Celebs on FB	57485	15183	36.68%	30.47%
10	Fanpage.it	59885	15634	38.21%	31.38%
11	Mashable	62545	17341	39.90%	34.80%
12	Eurosport	63997	18002	40.83%	36.13%
13	VICE	65029	18821	41.49%	37.77%
14	The Economist	65710	19164	41.92%	38.46%
15	The Cool hunter	66269	19588	42.28%	39.31%
80% of top 15 favorite cases		53015.2	15670.4		
Total fans		156737	49827		

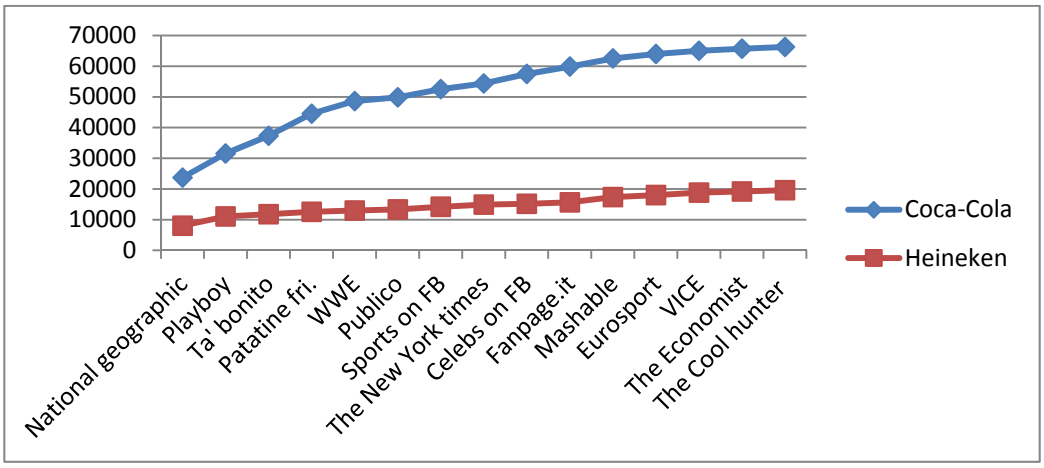


Figure (8a): Multiple products-single channel=News/Media in Europe (second method)

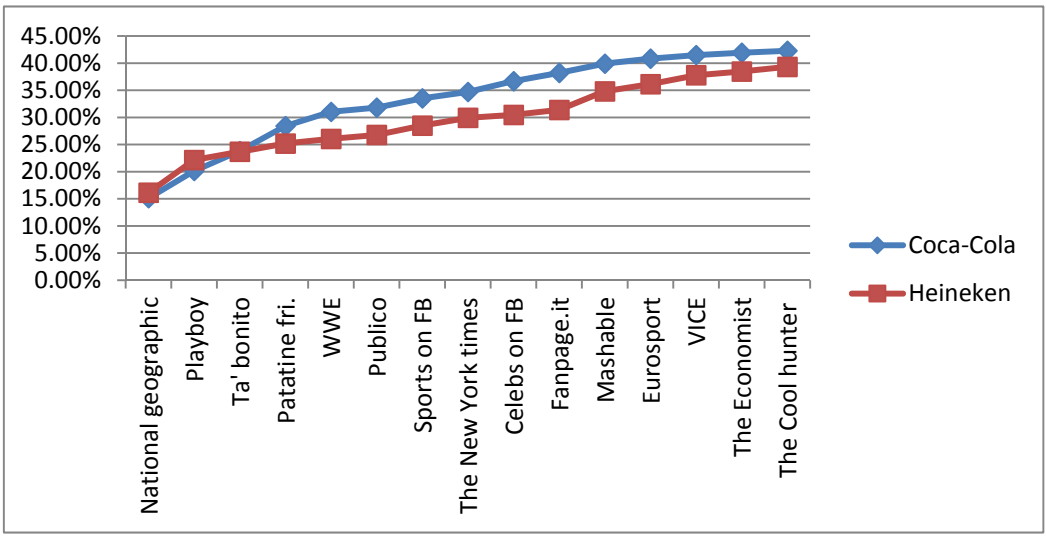


Figure (8b): Multiple products-single channel=News/Media in Europe (second method), in percentage

What was reported in Table (6) was an example of situations where two products have so many dissimilar favorites. As can be seen, the 80% constraint will be met when advertising in several cases and not a few cases like TV channel. This implies that advertising in News/Media channel might not result in favorable business outcome since the number and percentage of expose fans are significantly less than other channels.

Method B, Lift Factor

The meaning of lift factor is mentioned in methodology section in detail. Here the interpretation of the results will be explained by reviewing the values mentioned in different tables. There are 3 tables and 3 corresponding graphs for each channel such as TV, Companies/Products, News/Media, Music, Movie, Books, Gams/Apps and All channels together.

In Table (7) TV programs are mentioned and ranked based on the number of Coca-Cola fans' preferences (column 1). The numbers of each top ten program are expressed individually and independent to each other. These numbers include fans of each TV program in Europe in Wisdom network (column 2), Coca-Cola fans of each TV program (column 3) and ratio that expresses percentage of Coca-Cola fans in comparison to entire related TV program fans in Europe. These ratios (column 4) are calculated by dividing column 3 by column 2 and

the lift factors that are proportions of ratios divided by the fix figure of 6.6 % (column 5). Table (7) is followed by its corresponding graph.

Table (7): Lift factors, TV channel in Europe, individual measures

Total fans of Coca-Cola in Europe / wisdom population in Europe is 6.6%

	TV program	Fans in Europe	Coca-Cola fans	Ratio (%)	Lift factor
1	The Simpsons	181566	64846	35.71%	5.409
2	House	194147	55240	28.45%	4.309
3	South park	138979	49614	35.70%	5.407
4	Family guy	155668	49473	31.78%	4.813
5	MTV	79676	42216	52.98%	8.025
6	How I met your mother	140307	31745	22.63%	3.427
7	Two and a half men	116320	31207	26.83%	4.063
8	SpongeBob Squ.	56456	29930	53.01%	8.029
9	Futurama	73755	28697	38.91%	5.893
10	The Big bang theory	103892	25113	24.17%	3.661

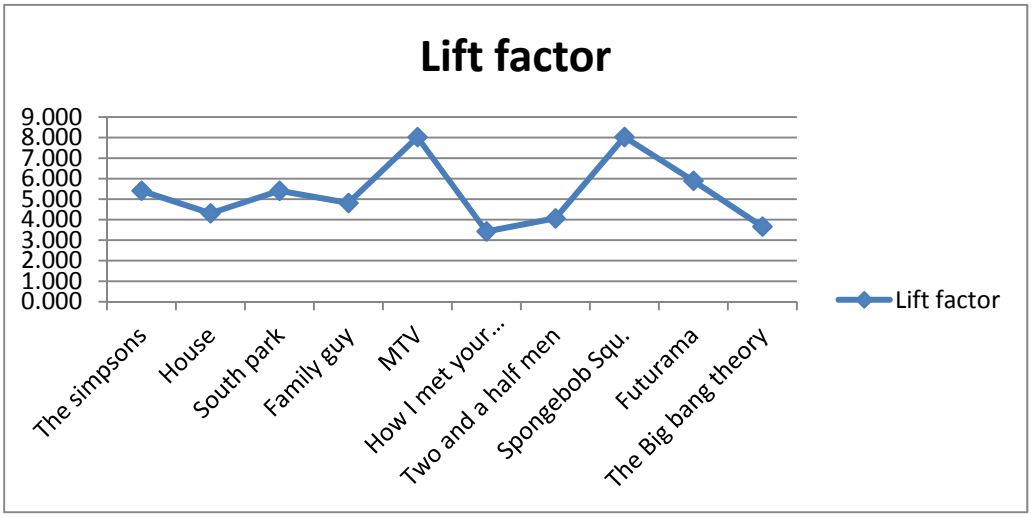


Figure (9): Lift factors, TV channel in Europe, individual measures

In Table (8), TV programs are ranked based on highest to lowest lift factor of table (7) and measures in columns 2 and 3 are accumulatively calculated. Table (8) is chosen to be the best and most useful Table for advertising purposes because programs are ranked according to individual lift factors in Table (7) in descending order and also accounts for accumulative representation of case fans. When the lift factor is high the ratio is high as well and this means more numbers of specific case fans like Coca-Cola rather than others. In other words, higher lift factor for an individual program or a higher accumulative lift factor for a combination of programs represents the most favorable programs that should be considered for marketing and advertising purposes. For TV programs most favorable programs would be SpongeBob, MTV, Futurama and etc. respectively (the number of chosen

programs depends on the budget that is dedicated for advertisement). For better representation of this result, Table (8) is followed by its corresponding graph.

Table (8): Lift factors, TV channel in Europe, accumulative measures, ranked by highest to lowest lift factors

	TV program	Fans in Europe	Coca-Cola fans	Ratio (%)	Lift factor
	SpongeBob Squ.	56456	29930	53.01%	8.029
or	MTV	112577	54441	48.36%	7.324
or	Futurama	159946	65191	40.76%	6.173
or	The Simpsons	257983	85458	33.13%	5.017
or	South park	306992	92516	30.14%	4.564
or	Family guy	349097	96159	27.55%	4.172
or	House	451278	109485	24.26%	3.674
or	Two and a half men	492643	112363	22.81%	3.454
or	The Big bang theory	520712	113922	21.88%	3.314
or	How I met your mother	552846	115695	20.93%	3.170

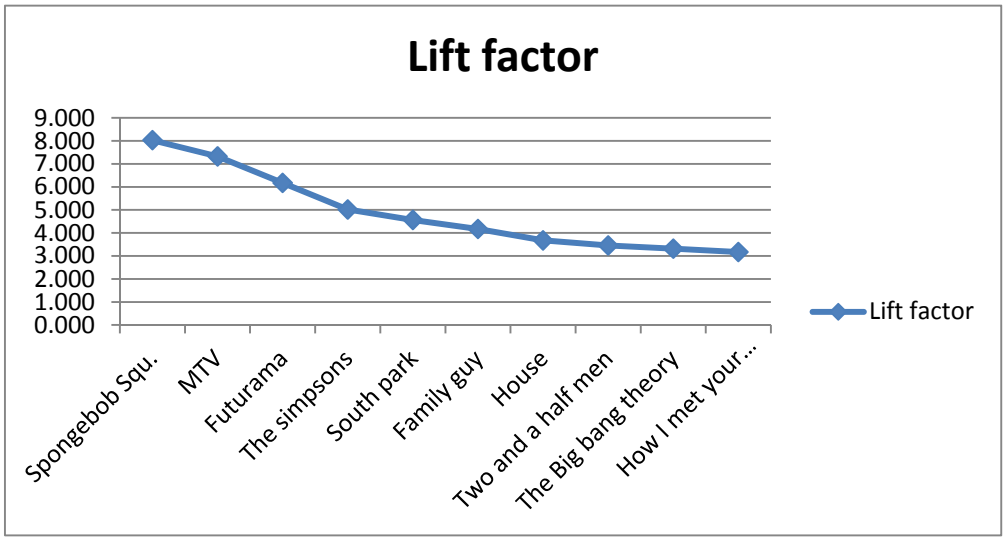


Figure (10): Lift factors, TV channel in Europe, accumulative measures, ranked by highest to lowest lift factors

To have a better overview of all possibilities together, lift factors of top 10 cases for all studies channels are drawn in a single graph. Figure (11) depicts the relationships between lift factors and the number of Coca-Cola fans of top 10 cases in each channel. If the position of a channel is more to the right it means that by advertising in that channel you cover the most number of fans, while the more you go to the top advertising will be more influential (i.e. the greater the impact factor will be). As can be seen in Figure (11) "Company" and "All" are the

most advantageous channels to advertise in. Figure (11) deals with the number of fans of cases who are also Coca-Cola fans. To have a broader impression on the total number of people that will be imposed to a possible advertisement, Figure (12) correlates the lift factors with the total number of fans of each case (i.e. regardless of whether they are Coca-Cola fans or not). The results of Figure (12) further confirms previous findings that the most favorable channels are "Company" and "All" channels.

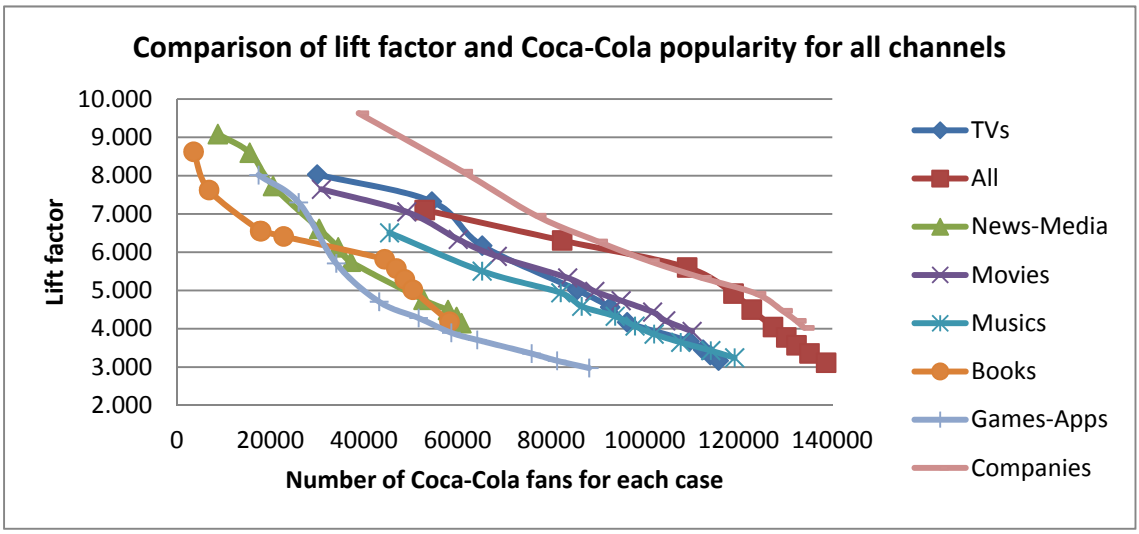


Figure (11): Relationships between lift factors and number of Coca-Cola fans for all channels

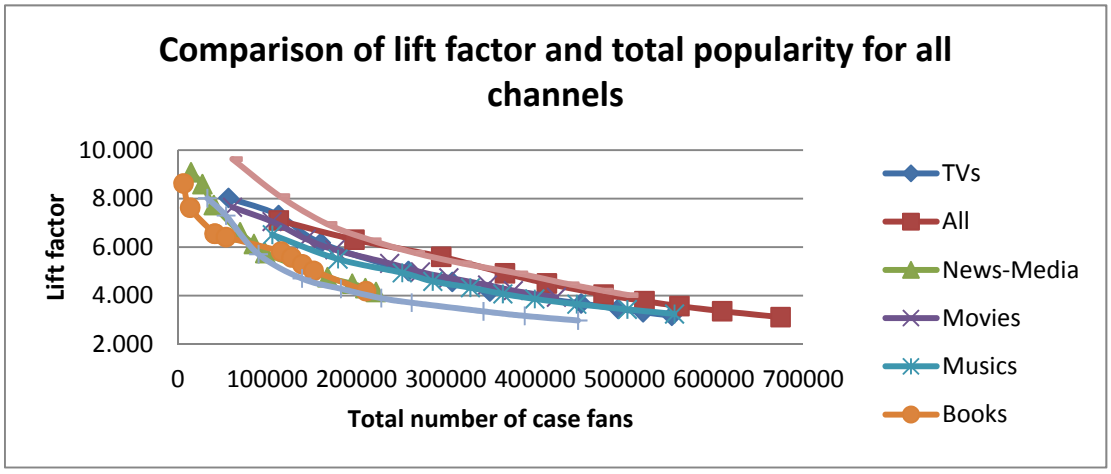


Figure (12): Relationships between lift factors and total number of fans for all channels

Conclusion and discussion

As it was mentioned earlier in the introduction, the main objective of this article was set to analyzing the collective intelligence application software "Wisdom Professional" with "Coca-Cola" as the case study. The study area is also considered to be Europe. To this end, it was tried to introduce Wisdom Professional, explain application of collective intelligence in this software, and use collective intelligence information from Wisdom Professional for improving advertisement in (social) media for Coca-Cola case. Wisdom Professional software was utilized as collective intelligence software that uses Facebook information of users to make better decisions in advertising.

In order to select the best advertising case from any possible channel, two methods were used. First method was named as "popularity-oriented method". The basis of this method relies on the number of fans of Coca-Cola who like a specific Facebook page. Based on different possibilities in categorizing Facebook fan pages, this method is followed in three different directions. These three directions are called "single product – single channel", "multiple products – single channel" and "single product – multiple channels".

By following the first method, the results of single product – single channel in TV category showed that 60.87% of Coca-Cola fans are also fans of top three TV programs that are "The Simpsons", "House" and "South Park". This implies the fact that for advertising purposes we should pay more attention to these three programs because 80% of top 10 TV program's fans like these three programs.

By following the first method, the results of multiple products – single channel based on first way (similarities) for TV channel showed that top 20 TV Programs for six products namely Coca-Cola, Pepsi, Red Bull, Fanta, Sprite and Heineken were very similar to each other. It means Coca-Cola competitors fans also like these top 20 programs just with a little difference in the order of these TV Programs. It shows that when we advertise in each of these top 20 TV programs, competitor fans are also watching Coca-Cola advertisement because these top 20 TV programs are the same favorite programs for them as well.

By following the first method, the results of multiple products – single channel based on first way (dissimilarities) for news/media channel showed that 80% of top 15 cases contain first eight news/media cases for Coca-Cola. These news/media cases are the same for Heineken plus 3 other cases. Therefore when Coca-Cola advertise in first eight cases, Heineken fans also watch the advertisement because these 8 cases are similar but when Coca-Cola wants to attract more customers of competitors like Heineken should pay attention to 3 other exclusive Heineken favorite cases as well.

By following the first method, the results of single product – multiple channels revealed that 71.45% of Coca-Cola fans are also fans of top three cases that are "YouTube", "Facebook" and "The Simpsons". This implies that for advertising purposes we should pay more attention to these three cases because 80% of top 10 all cases fans like these three cases.

In the second method and as for the TV channel different cases are ranked based on highest to lowest lift factor. The first five programs are "Sponge Bob squ.", "MTV", "Futurama", "The Simpsons" and "South park". This means that for instance, the proportion of "Sponge Bob's" fans who like Coca-Cola in comparison to total fans of "Sponge Bob" are more than other programs. Therefore, media planners are advised to advertise in these channels since a bigger fraction of fans will be covered by advertising in these channels.

Now the question is which method is more appropriate for advertising in (social) media, popularity-oriented method or lift factor method? The answer to this question would be that we need to use both methods. The reason is that popularity method covers more people and on the other hand lift factor method highlights the penetration degree on specific market. That is why media planners need combination of high popularity and penetration for better and more effective advertising. However and in the case of having contradictive results, the priority should be given to popularity-oriented method. However, the final choice of channels or pages to advertise in, depends on companies' objectives and their social media planning strategies. Companies might have different advertising strategies for short-term as well as for long-term. Achieving highest penetration rate might be more suitable for long-term while attracting more customers in short-term might be more beneficial.

Currently we have no information about the costs of advertising in social media that is why we cannot offer any recommendation in this area that which social media is better for advertising in terms of expenses. In this article expressed opinions are based on the results from the methodological analysis and not the economical assessments.

Acknowledgement

We would like to acknowledge Mr. Eddy Conjaerts for providing us the access to Wisdom Professional and also for his constant support.

Bibliography

[Bourdieu, 1986] P. Bourdieu, The forms of capital, In J. Richardson (Ed.) Handbook of Theory and Research for the Sociology of Education, New York, and Greenwood press, 241-258, 1986.

[IBM] IBM, http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-collective-intelligence.html, last access 5/8/2013.

[Leimeister, 2010] Jan Marco, Leimeister,"collective intelligence", Research Center for IS Design (ITeG), Universität Kassel, 2010.

[Merton, 1967] R.K. Merton, Social theory and social structure, New York: Free Press, 1967.

[Nickel, 2013] Eric D. Nickel, "Collective Intelligence in Emergency Management: Social Media's Emerging Role in the Emergency Operations Center", it is derived from http://www.usfa.fema.gov/pdf/efop/efo45101.pdf, last access 24/7/2013.

[SQL Server Microsoft] SQL Server Microsoft, Lift Chart (Analysis Services- Data Mining), http://technet.microsoft.com/en-us/library/ms175428.aspx, last access 23/10/2013.

[Surowiecki, 2005] James, Surowiecki,"The Wisdom of Crowds", first edition, Anchor books, 2005.

[Wisdom] Wisdom, https://apps.facebook.com/wisdomapp/?fb_source=bookmark_apps&ref=bookmarks&count=0&fb_bmpos=3_0, last access 20/6/2013.

Authors' Information

Elham Fayezioghani *– master in Social Planning, University of Tehran, Iran and master in MIS, Hasselt University, Belgium; e-mail: ertebatat_2005@yahoo.com*

Major Fields of Scientific Research: Collective intelligence applications, marketing, social planning

Vanhoof Koen *- Professor Dr., Hasselt University; Campus Diepenbeek; Department of Applied Economic Sciences. Agoralaan; BE-3590 Diepenbeek; Belgium;*

e-mail: koen.vanhoof@uhasselt.be

Major Fields of Scientific Research: data mining, knowledge discovery, process modeling and minig.

A NEW METHOD FOR THE BINARY ENCODING AND HARDWARE IMPLEMENTATION OF METABOLIC PAHTWAYS

Carlos Recio Rincon, Paula Cordero, Juan Castellanos, Rafael Lahoz-Beltra

***Abstract**: In this paper we introduce a new method for the binary encoding of metabolic pathways. Our method assigns a 5-bit word to the functional groups of the molecules or metabolic intermediates, sorting the functional groups by its redox potential. We illustrate our approach modelling two very well known metabolic pathways, glycolysis and the Krebs cycle, showing how sugars and other glycolytic molecules could be modeled as binary matrices as well as LED dot matrices. The method enables the design of 'metabolic hardware' which may be useful in the study of the optimization of metabolic pathways as well as in the area of molecular and natural computing.*

***Keywords**: molecular topology representation, computational chemistry, biosinpired architectures, molecular computing.*

***ACM Classification Keywords**: F.1. Computation by abstract devices; F1.1. Models of computation*

Introduction

Metabolic pathways are series of biochemical reactions occurring within a cell. In each pathway an enzyme catalyzes a reaction transforming a molecule or substrate to a new molecule or product. Modelling and simulation of self-organization in metabolic pathways (Fig. 1) has been addressed under different approaches. In fact the simulation of biochemical reactions is related to the history of computers and biomathematics [Mendes and Kell, 1996]. For instance, in the context of game theory [Melendez-Hevia, 1990; Melendez-Hevia et al., 1994] introduced 'the game of the pentose phosphate cycle', a mathematical game that gives a simple explanation of how the metabolic reactions of the pentose cycle find an optimal configuration. The authors found an optimal solution to the problem of transforming six pentoses in five hexoses applying the principle of Darwinian natural selection [Lahoz-Beltra and Perales-Gravan, 2010] and the simplicity theorem. The study of metabolic pathways is also required for the design of virtual cellular systems [Sipper, 1990; Takahashi et al., 2002] e.g. MCell, VCell and E-Cell, as well as in the field of molecular and natural computation [Stefanovic, 2008].

In all these studies, there are two main 'ingredients', metabolites or metabolic molecules and enzymes (Fig. 1). However, taken together these two components enzymes have received much attention since they are very important in the metabolic pathways by enabling the biochemical reactions take place to a reasonable speed. From a historical perspective, the application of the theory of finite automata has enabled the modelling of important biological molecules such as proteins (enzymes are a particular type of proteins). During the decade of the 90s several researchers in the area of molecular computing considered the possibility that future computers arise based upon an architecture composed of proteins [Hameroff et al., 1992]. However, and although there is currently a lack of studies on protein-inspired computers we believe that this is a promising field that will give interesting results in the future. To date most of the proposals has been based on studies with enzymes, proteins with catalytic function responsible for the thousands of chemical reactions that sustain life on Earth. In the scientific literature are described theoretical models [Birge, 1995; Bray, 1995] with no practical implementations as well as experimental devices using real enzymes [Hiratsuka et al., 1999].

Theoretical models of proteins and enzymes can be classified into two main groups. On the one hand, some models assume an analogy between an enzyme and a transistor, due to the computational and electronic characteristics of enzymatic processing. For instance, [Di Paola et al., 2004] proposed a model wherein the bacterial chemotaxis proteins are implemented as hardware using an operational amplifier (Fig. 2). On the other hand, there are a number of models that assume that an enzyme is a finite automaton with two or more states which correspond to conformational states of the enzyme. For example, [Marijuan, 1991] introduced a probabilistic model of an enzyme with its state table and transition probabilities. Within this group there are also models in which an enzyme is considered as a McCulloch-Pitts neuron [Okamoto et al., 1999; Di Paola et al., 2004]. In other instances it is possible to find models of protein assemblies. For instance, [Lahoz-Beltra et al., 1993] introduced a model bio-inspired in the microtubules of cellular cytoskeleton showing the possibility of molecular computation via Boolean operations in microtubules. In microtubules protein subunits are assembled and behaving according the theory of coherent excitations introduced by [Frohlich, 2012] and in consequence like automata which conformational changes occurring in an orchestrated fashion. Moreover, in 2008 [Lin and Chen, 2008] developed evolvable hardware bio-inspired in cytoskeleton. Therefore, in the models of this second group a network of proteins or enzymes is a network of finite automata capable of performing Boolean operations. Based on this approach [Lahoz-Beltra, 2001] introduced a model of electronic enzyme (Fig. 3) which is under Spanish patent [Lahoz-Beltra, 2003].

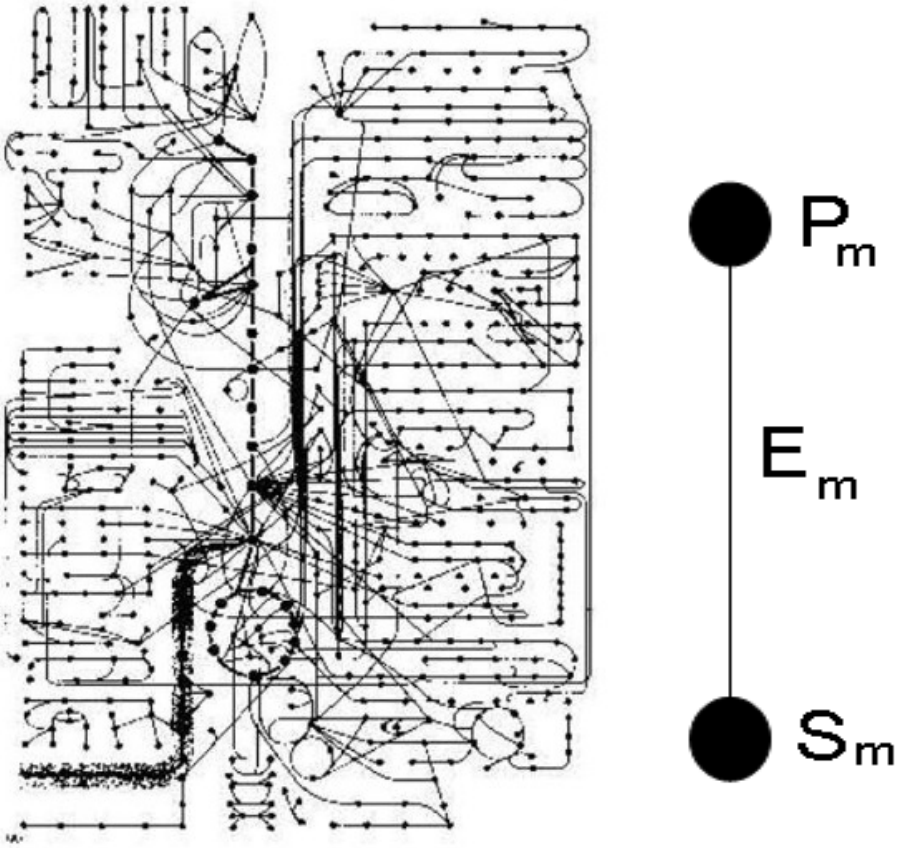


Figure 1. (Left) Metabolic pathways are series of biochemical reactions occurring within a cell. In each pathway an enzyme E_m (lines or edges) catalyzes a reaction transforming a molecule or substrate S_m (vertices or nodes) to a new molecule or product P_m (vertices or nodes). A set of metabolic pathways is called a metabolic network, e.g. glycolysis and the Krebs cycle [Source of the biochemical circuit diagram: Molecular Biology of the Cell. Alberts et al. Fourth edition]. (Right) Biochemical reaction: $S_m + E_m \rightarrow P_m$.

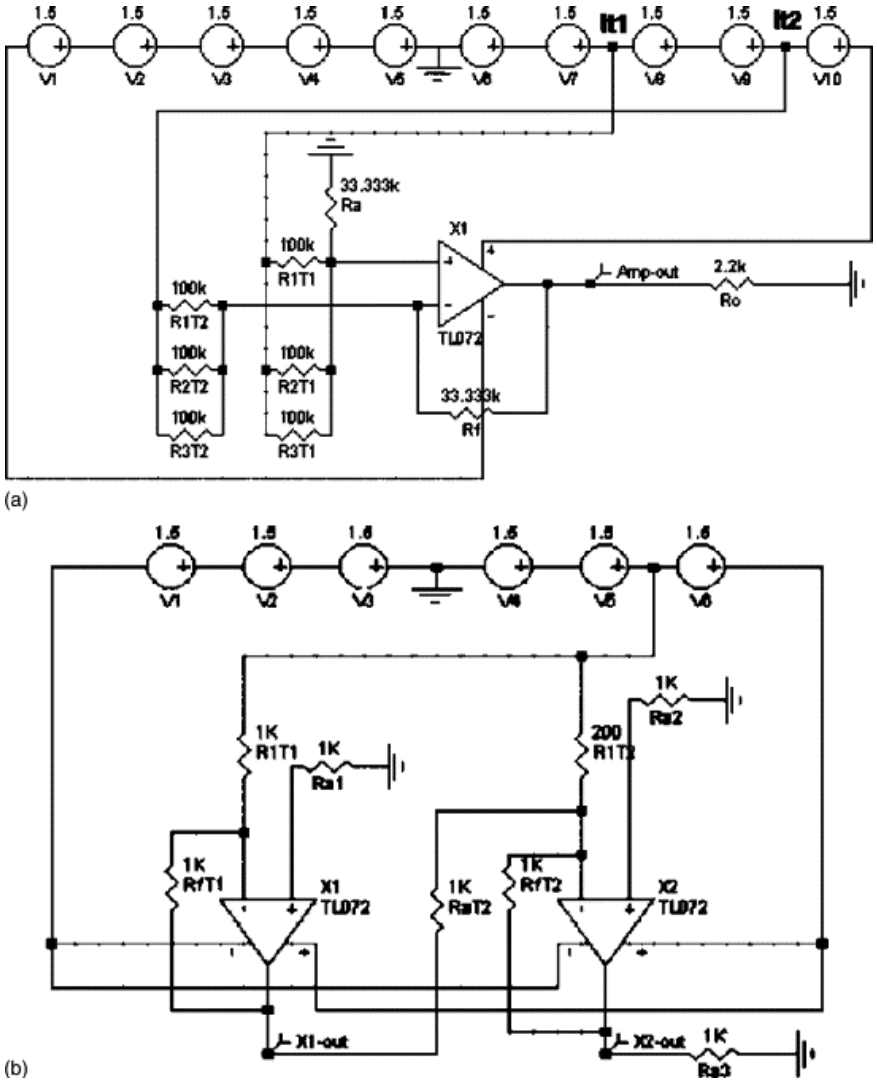


Figure 2. Modelling and simulation of bacterial taxis using operational amplifiers [Di Paola et al., 2004]

In this paper we did not study the computational role of enzymes in metabolic pathways, but we explored the possibility of using metabolic networks as hardware in the field of molecular and natural computing. We call to these biosinpired architectures as *metabolic hardware*. In particular, adopting as an example the intermediate molecules of two very well known metabolic pathways, glycolysis and the Krebs cycle, we introduce the methodology to translate the molecular structure or topology of their metabolic intermediates to a binary matrix.

Methods

From a historical perspective one of the first procedures to translate the molecular topology to a matrix was introduced by [Randic, 1974], taking an element a_{ij} the value 1 when the vertices are adjacent or 0 otherwise. Figure 4 illustrates an example of this method for vitamin A or retinol [Lahoz-Beltra, 2012].

Our method assigns a 5-bit word to the functional groups of the molecule. For that purpose we define a table or *Rosetta stone* (Table I) that includes the most frequent functional groups in metabolic intermediates, which were ordered by its redox potential (tendency of a functional group to acquire electrons).

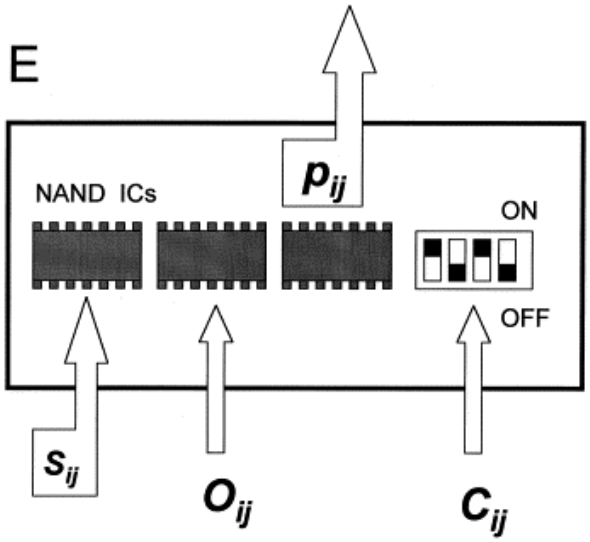
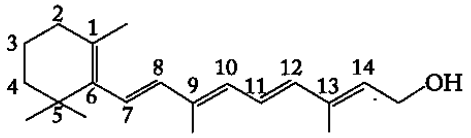


Figure 3. An enzyme $E_m(c_{1j}, c_{2j},..., c_{nm}, o_{1j}, o_{2j},..., o_{nm})$ has been defined as an automaton with a finite number of internal 'conformational' states represented by an *n*-bit word $c_{1j}, c_{2j},..., c_{nm}$ and a set of operations or instructions modelling the 'active groups' of the active site and given by $o_{1j}, o_{2j},..., o_{nm}$ Boolean operators (e.g. AND, XOR). We define an enzymatic reaction as $S_m + E_m \rightarrow P_m$ where $s_{1j}, s_{2j},..., s_{nm}$ and $p_{1j}, p_{2j},..., p_{nm}$ are the *n*-bit words representing the substrate S_m and product P_m respectively of the enzymatic reaction performed by enzyme E_m. Based on above definitions the electronic enzyme 'catalyzes' a biochemical reaction conducting the Boolean operations given by: $p_{1j} = s_{1j}\, o_{1j}\, c_{1j}$, $p_{2j} = s_{2j}\, o_{2j}\, c_{2j},..., p_{nm} = s_{nm}\, o_{nm}\, p_{nm}$ [Lahoz-Beltra, 2001].

steczki witaminy A:



to cząsteczka ta może być wyrażona za pomocą następującej macierzy:

$$\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}$$

Przyjmujemy dla x_{ij} wartość 1, jeżeli atomy i, j są między sobą związane, i wartość 0 w przypadku przeciwnym, to znaczy gdy wiązanie między nimi nie istnieje.

87

Figure 4. The molecule of vitamin A or retinol represented as a binary matrix [Lahoz-Beltra, 2012] (Transl: Polish).

Let S and P be two binary matrices which represent respectively the substrate S_m and product P_m of a biochemical reaction catalyzed by an enzyme E_m. Since that glycolysis and Krebs cycle all metabolites or metabolic intermediates are molecules of 3, 4, 5 or 6 carbon atoms, we will define the (1), (2), (3) and (4) matrices respectively:

$$C_3=\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{pmatrix} \quad (1) \qquad C_4=\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{pmatrix} \quad (2)$$

$$C_5=\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix} \quad (3) \qquad C_6=\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} \end{pmatrix} \quad (4)$$

Note that given a value i, $\begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{i5} \end{pmatrix}$ is a row vector representing the functional group of the substrate s_{ij} or product p_{ij} molecules. Thus, each row in the matrices C_3, C_4, C_5 and C_6 represents a carbon atom in the molecule, having a total of 32 possible binary vectors from 00000 to 11111 (Table I). Using as a criterion the redox potential vectors were classified from its most reduced (addition of hydrogen or the removal of oxygen) form or alkyl group to the most oxidized (addition of oxygen or the removal of hydrogen) or CO_2. However, since the metabolites of glycolysis and the Krebs cycle are the result of assembling functional groups among a total of 22 combinations of carbon, then 10 binary vectors are without chemical meaning. In order to perform future simulation experiments, molecules of CO_2 and acetyl-CoA were represented as a row vector (5) and 2x5 matrix (6) shown below:

$$CO_2=\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad (5) \qquad \text{acetyl-CoA}=\begin{pmatrix} 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad (6)$$

Results

Applying the technique described above sugars and other glycolytic molecules were modeled as binary matrices as well as hardware (Fig. 5). The hardware representation was conducted implementing molecules as LED dot matrices using CEDAR Logic Simulator program [Sprague, 2007]. The route of glycolysis was modeled as shown below:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \rightarrow$$

$$\begin{pmatrix}1&1&0&0&0\\1&0&0&0&1\\1&0&1&0&0\end{pmatrix}\rightarrow\begin{pmatrix}1&1&1&1&0\\1&0&0&0&1\\1&0&1&0&0\end{pmatrix}\rightarrow\begin{pmatrix}1&1&1&0&0\\1&0&0&0&1\\1&0&1&0&0\end{pmatrix}\rightarrow\begin{pmatrix}1&1&1&0&0\\1&0&1&0&1\\1&0&0&0&0\end{pmatrix}$$

$$\rightarrow\begin{pmatrix}1&1&1&0&0\\1&0&1&1&0\\0&1&1&0&0\end{pmatrix}\rightarrow\begin{pmatrix}1&1&1&0&0\\1&1&0&1&1\\0&1&0&0&0\end{pmatrix}$$

representing each matrix the following metabolites of glycolysis:

Glucose $\rightarrow$ Glucose-6-phosphate $\rightarrow$ Fructose-6-phosphate $\rightarrow$ Fructose-1,6-biphosphate $\rightarrow$

Glyceraldehyde-3-phosphate $\rightarrow$ 1,3-biphosphateglycerate $\rightarrow$ 3-phosphoglycerate $\rightarrow$ 2-phosphoglycerate

$\rightarrow$ phosphoenolpyruvate $\rightarrow$ pyruvate

Using the same method the Krebs cycle was modeled as follows:

$$\begin{pmatrix}1&1&1&0&0\\0&1&0&0&1\\1&0&0&1&1\\0&1&0&0&1\\1&1&1&0&0\end{pmatrix}\begin{pmatrix}1&1&1&0&0\\0&1&0&0&1\\0&1&0&1&0\\1&0&0&1&0\\1&1&1&0&0\end{pmatrix}\begin{pmatrix}1&1&1&0&0\\0&1&0&0&1\\0&1&0&0&1\\1&1&0&1&1\\1&1&1&0&0\end{pmatrix}\begin{pmatrix}1&1&1&0&0\\0&1&0&0&1\\0&1&0&0&1\\1&1&1&0&1\end{pmatrix}$$

$$\leftarrow(1\ 1\ 1\ 0\ 0)\rightarrow\leftarrow(1\ 1\ 1\ 0\ 0)\rightarrow\qquad\rightarrow\qquad\downarrow$$

$$\begin{pmatrix}1&1&1&0&0\\1&1&0&1&1\\0&1&0&0&1\\1&1&1&0&0\end{pmatrix}\leftarrow\begin{pmatrix}1&1&1&0&0\\0&1&0&0&1\\1&0&0&1&0\\1&1&1&0&0\end{pmatrix}\leftarrow\begin{pmatrix}1&1&1&0&0\\0&1&1&0&1\\0&1&1&0&1\\1&1&1&0&0\end{pmatrix}\leftarrow\begin{pmatrix}1&1&1&0&0\\0&1&0&0&1\\0&1&0&0&1\\1&1&1&0&0\end{pmatrix}$$

where each matrix stands for one of the following metabolites:

Citrate $\rightarrow$ Iso-citrate $\rightarrow$ α -Ketoglutarate $\rightarrow$ Succinyl-CoA $\rightarrow$

Oxalacetate $\leftarrow$ Malate $\leftarrow$ Fumarate $\leftarrow$ Succinate

Note that we have used special notation for citrate and iso-citrate matrices, because the third carbon atom is bonded to three others (Fig. 6).

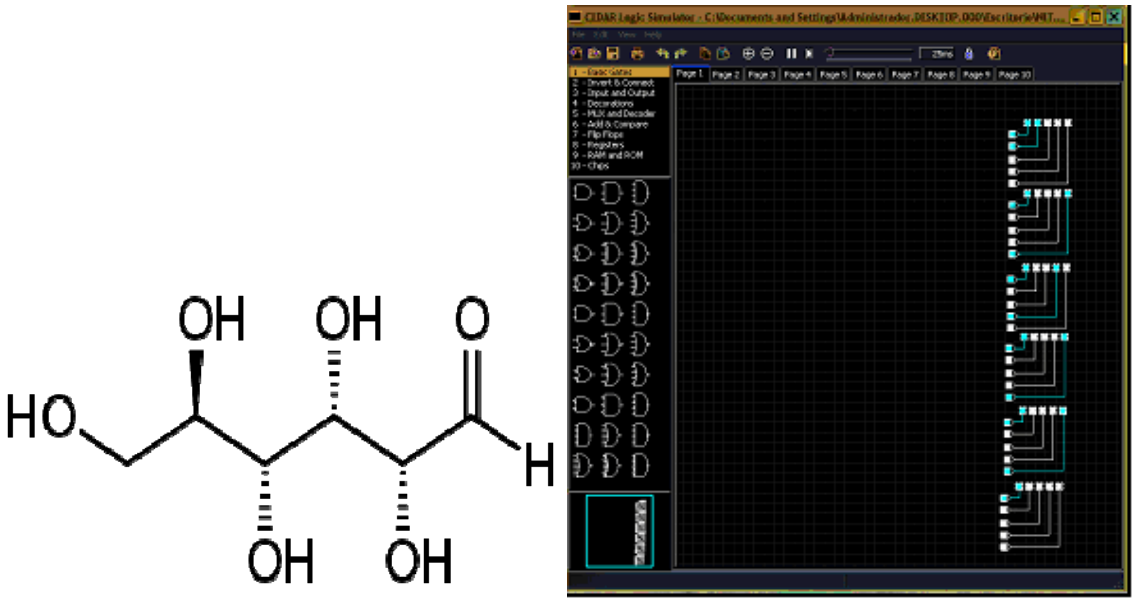


Figure 5. Glucose molecule (Left) and its hardware version as a matrix of LEDs (Right) simulated with CEDAR Logic Simulator.

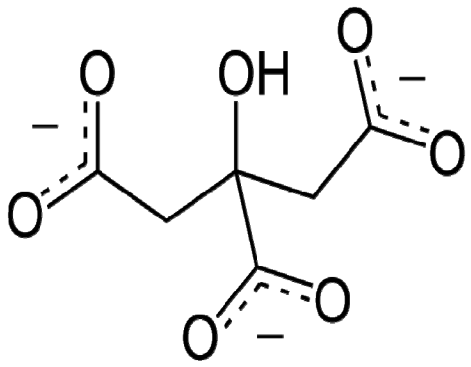
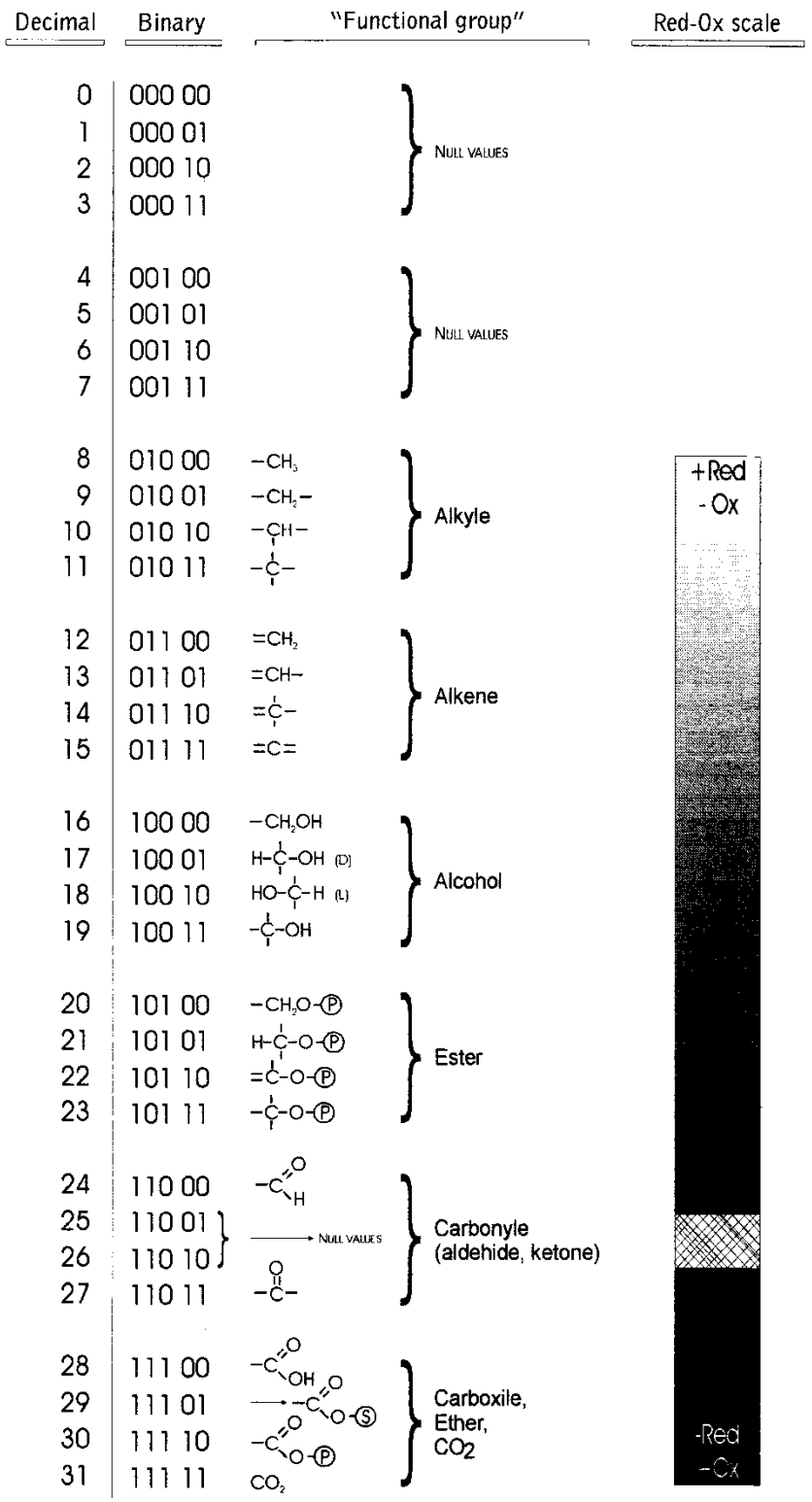


Figure 6. Citrate molecule

Conclusion

This paper presents a novel method to represent the topology of a molecule as a binary matrix. The method enables the design of 'metabolic hardware', developing an example with two well-known metabolic pathways, glycolysis and the Krebs cycle. In our opinion the binary representation of molecules or metabolites is a first step that will lead in a future to study the metabolic pathways in search of bioinspired architectures with special interest in the field of molecular and natural computing.

TABLE I. *Rosetta stone* for the hardware implementation of metabolic pathways



Bibliography

[Birge, 1995] R.R. Birge. 1995. Protein-based computers. Scientific American 272: 66-71.

[Bray, 1995]. D. Bary. 1995. Protein molecules as computational elements in living cells. Nature 376: 307-312.

[Di Paola et al., 2004] V. Di Paola, P.C. Marijuan, R. Lahoz-Beltra. 2004. Learning and evolution in bacterial taxis: an operational amplifier circuit modelling the computational dynamics of the prokaryotic 'two component system' protein network. Biosystems 74: 29-49.

[Frohlich, 2012] H. Frohlich. 2012. Coherent Excitations in Biological Systems. Proceedings in Life Sciences. Springer-Verlag Berlin and Heidelberg GmbH & Co. KG: 1-240.

[Hameroff et al., 1992] S.R. Hameroff, J.E. Dayhoff, R. Lahoz-Beltra, A.V. Samsonovich, S. Rasmussen. 1992. Models for molecular computation: conformational automata in the cytoskeleton. IEEE Computer 25: 30-39.

[Hiratsuka et al., 1999] M. Hiratsuka, T. Aoki, T. Higuchi. 1999. Enzyme transistor circuits for reaction-diffusion computing. IEEE Transactions on Circuits and Systems – I 46: 294-303.

[Lahoz-Beltra and Perales-Gravan, 2010] R. Lahoz-Beltra, C. Perales-Gravan. 2010. A survey of nonparametric tests for the statistical analisis of evolutionary computational experiments. International Journal Information Theories and Applications 17: 49-61.

[Lahoz-Beltra et al., 1993] R. Lahoz-Beltra, S.R. Hameroff, j.E. Dayhoff. 1993. Cytoskeletal logic: a model for molecular computation via Boolean operations in microtubules and microtubule-associated proteins. BioSystems 29: 1-23.

[Lahoz-Beltra, 2001] R. Lahoz-Beltra. 2001. Evolving hardware as model of enzyme evolution. BioSystems 61: 15-25.

[Lahoz-Beltra, 2003] R. Lahoz-Beltra. 2003. Enzima electronico, con sitio activo reconfigurable. Patente No. P200000174 (27.01.2000). Oficina Española de Patentes y Marcas. http://invenes.oepm.es/InvenesWeb/detalle?referencia=P200000174.

[Lahoz-Beltra, 2012] R. Lahoz-Beltra. 2012. Matematica zycia. Modele matematyczne w biologii i ekologii. RBA (Transl: Polish): 1-159.

[Lin and Chen, 2008] Y-H. Lin, J-C. Chen. 2008. Neuromolecularware – A bio-inspired evolvable hardware and its application to medical diagnosis. In: ARC 2008 (R. Woods et al., Eds.), LNCS 4943: 324-329.

[Marijuan, 1991] P.C. Marijuan. Enzymes and theoretical biology: sketch of an informational perspective of the cell. BioSystems 25: 259-273.

[Melendez-Hevia et al., 1994] E. Melendez-Hevia, T.G. Waddell, F. Montero. 1994. Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle. Journal of Theoretical Biology 166: 201-220.

[Melendez-Hevia, 1990] E. Melendez-Hevia. The game of the pentose phosphate cycle: a mathematical approach to study the optimization in design of metabolic pathways during evolution. Biomed. Biochem. Acta 49: 903-916.

[Mendes and Kell, 1996] P. Mendes, D.B. Kell. 1996. Computer simulation of biochemical kinetics. In: BioThermoKinetics of the living cell (H.V. Westerhoff et al., Eds.), BioThermoKinetic Press: 254-257.

[Okamoto et al., 1999] M. Okamoto, T. Sekiguchi, K. Tanaka, Y. Maki, S. Yoshida. 1999. Biochemical neuron: harwdare implementation of functional devices by mimicking the natural intelligence such as metabolic control systems. Computers and Electrical Engineering 25: 421-438.

[Randic, 1974] M. Randic. 1974. On the recognition of identical graphs representing molecular topology. The Journal of Chemical Physics 60: 3920-3927.

[Sipper, 1990] M. Sipper. 1990. The emergence of cellular computing. IEEE Computer 32:18-26.

[Sprague, 2007] B. Sprague, M. Lewellyn, D. Knierim, J. Lansford, N. Harro. 2007. CEDAR Logic Simulator. Cedarville University. http://sourceforge.net/projects/cedarlogic/

[Stefanovic, 2008] D. Stefanovic. 2008. Emerging models of computation: directions in molecular computing. In: Software-Intensive Systems (M. Wirsing et al., Eds.) LNCS 5380: 255-265.

[Takahashi et al., 2002] K. Takahashi, K. Yugi, K. Hashimoto, Y. Yamada, C.J.F. Pickett, M. Tomita. 2002. Computational challenges in cell simulation: a software engineering approach. IEEE Intelligent Systems 17: 64-71.

Authors' Information

Carlos Recio Rincon - *B*iologist, co-*author of the course* on CD-*ROM* entitled 'Histologia animal y vegetal' (2007, Editorial Complutense). His main interest is agriculture and ecology, *currently* works in '*rural* tourism'*;* e-mail: carlosrecio@gmail.com

Paula Cordero – *Natural Computing Group, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s.n., 28660 Boadilla del Monte, Madrid, Spain; e-mail:* paula.cormo@gmail.com

Major Fields of Scientific Research: complex systems, natural computing, evolutionary computation and the design of bioinspired algorithms.

Juan Castellanos – *Head of Natural Computing Group, Facultad de Informática,Universidad Politécnica de Madrid, Campus de Montegancedo s.n., 28660 Boadilla del Monte, Madrid, Spain; e-mail:* jcastellanos@fi.upm.es

Major Fields of Scientific Research: natural computing, formal language and automata theory.

Rafael Lahoz-Beltra – *Department of Applied Mathematics, Faculty of Biological Sciences, Complutense University of Madrid, 28040 Madrid, Spain; e-mail:* lahozraf@ucm.es

Major Fields of Scientific Research: evolutionary computation, embryo development modeling and the design of bioinspired algorithms.

ADVENT OF CLOUD COMPUTING TECHNOLOGIES IN HEALTH INFORMATICS

Omer K. Jasim, Safia Abbas, El-Sayed M. El-Horbaty, Abdel-Badeeh M. Salem

Abstract: *Cloud computing is internet based computing that allows client computers to access shared resources, software, and information from servers on the web/cloud. Health informatics technology describes the use of computer information systems to manage the patients' electronic health records based on electronic health record. The advent of cloud computing technology provides effective and dependable results to support healthcare services. The cloud technology reduces these costs for consumers and IT by improving clinical and quality outcomes for patients. This paper discusses the potential rule of the cloud technology in healthcare informatics. In addition, it presents the global challenges and technical difficulties which are facing this new technology.*

Keywords: *Cloud Computing, Healthcare Informatics, Cloud Challenges, Health Cloud Technology.*

Introduction

Cloud technology is a new way of delivering computing resources and services. This technology is defined as a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services) [Alex R. et al, 2011]. In essence, cloud computing is the legal transfer of computing as a service rather than as a product where the approach of these shared resources or services is furnished as a utility over a net. Over the cloud computing, customers can utilize network-based tools or applications through a web browser just as if they were programs installed locally on their own computer [Srinivasa R. at al, 2009; Omer K. et al, 2013; NSA, 2013].

On the other side, health informatics technology (HIT) describes the use of computer information systems to manage the patients' electronic health records based on electronic health record (EHR). Precisely, EHR system allows users in healthcare governance, such as hospitals, clinic, or a doctor's to enter, store, process, access, and manage patient healthcare data [Sanjay P. et al, 2012; Ruoyu Wu. et al, 2010]. As shown in Figure 1, typical data in EHRs include hospital's information, doctor's order entries and comments, patient's identification, laboratory test results for trainer, and others. EHRs can support clinicians towards providing better healthcare by granting access to comprehensive patient data, help to reduce medical prescription errors with various alerting functions, and can help patients and doctors to oversee their treatment and charge books for insurance payments [King M. et al, 2012; Samuel, O.W. et al, 2013; Ruoyu Wu. et al, 2012].

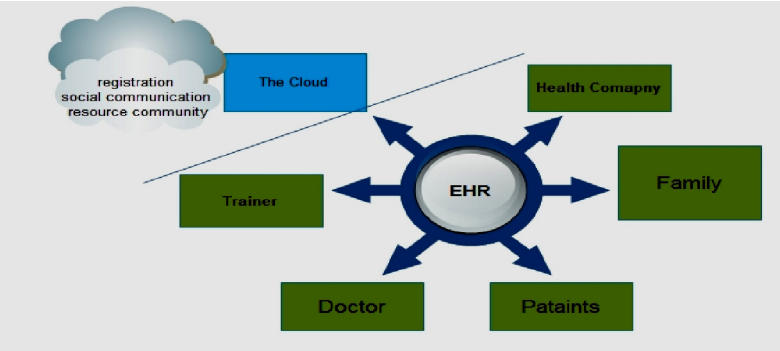


Figure1. Schematic of EHRs

All EHRs components are mentioned above, can easily relate to the cloud environment based on an accumulation of "www" based application or numbers of services tools [Ruoyu Wu. et al, 2012]. The cloud has an infrastructure built so consumers can deploy and run their applications [Eman AbuKhousa et al, 2012]. They also have different platforms with multiple operating systems so consumers can build, test, and deploy their applications on virtual servers. Thus, the cloud provides highly scalable environment to effectively manage the load, it provides many benefits for HIT by integration with healthcare providers, and these integration mechanisms help the HIT provider to share the data between many organizations. Data sharing serves various purposes that helping to improve the healthcare services. Moreover, many advantages are gone from this integration such dynamically updating, file storages, manageability, and fasting for health care operations [Thomas Trojer et al, 2012; Sunyaev A. et al, 2010; Buyya, Jwt al, 2011; H. Liohr et al, 2009].

This paper explains the impact of cloud technology on HIT rely on studies and discuss the healthcare systems and models in the context of this technology. Moreover, the paper discusses the proposed technical solutions for the HIT challenges. The remainder of this paper is structured as follows. Section II reviews the existing methods for healthcare in cloud computing technology, health cloud architecture and components are given in Section III. Section IV explains the global challenges and solutions for health cloud technology (HCT). Section V presents the information are enriched by this article. Section VI shows the conclusion and future works.

Existing Methods

Cloud computing can take on a vital part in containing healthcare integration costs, optimizing resources and ushering in a new era of inventions. Current trends aim towards accessing information anytime, anywhere, which can be achieved when moving healthcare information to the cloud. This new delivery model can make healthcare more efficient and effective, and at a lower cost to technology budgets. There are several articles that introduce contributions to building the environment for HCT.

Ortho [2014], plans to implement a cloud-based practice management technology solution through a company called Care Cloud (CeC). The CeC of Soma predicts that the use of the system will eventually evolve into a more advanced form of data sharing among Soma network of clinics as well as third party institutions such as insurance companies. The platform is designed to offer greater care efficiency to both the practice and its patients by minimizing redundancy inpatient procedures, therefore, the costs associated with them are minimized.

Yu [2011] investigates utilizing a service modeling approach to model the requirements and design of different Service-Oriented Architecture (SOA) based services by using Service Oriented Modeling and Architecture (SOMA) and employing Service Oriented Modeling Framework (SOMF) modeling styles and assets. It shows how to rapidly implement and evaluate e-health applications using this approach. Generally SOA can provide a full solution for facing some of the development and performance challenges facing the HCT.

Teng et al. [2010] provided a long term off-site medical image archive solution for Digital Imaging and Communication in medicine (DICOM). One of the biggest challenges which the healthcare industry struggles with is the growing cost of managing long-term on-site medical imaging archives. The continually increasing need for high volumes of medical images is resulting in scalability and maintenance issues with picture archiving and communication systems (PACS).

Guo et al. [2010] proposed a loud-based intelligent Hospital File Management System (HFMS) that aims to improve some of the restrictions (storage capacity, low performance) which characterize the traditional hospital management systems (HMS).

Fan et al. [2011] presented the Data Capture and Auto-Identification Reference (DACAR). DACAR aims to develop, implement and disseminate a novel secure platform in the Cloud for capturing, storing and consuming

data within a healthcare domain. By using a single point of contact, the DACAR platform promises to provide solutions for the challenges of HCT services.

From above attempts and methods , the cloud computing has given opportunities for clinics, hospitals, insurance companies, pharmacies, and other healthcare companies to agree in collaborating between them and share healthcare information to offer better quality of service and reduce costs.

Health Cloud Architecture

Recently, many medical organizations install software on their office or interconnection system. It shares with potential interruptions such as power outages, software upgrades, hardware failures and human mistake. When the software migrates to the cloud, upgrade the software to the open environment without breaking up your practice. The cloud base is built in redundancy, meaning that your system is perpetually usable, even if there is an outage on our goal. The services are designed so that outages remain transparent to the users and all services available.

Cloud Computing Architecture

A cloud computing architecture can be basically divided into three layers the characteristics layer, the model layer, and the deployment layer [Guo L. et al, 2010]. The characteristic layer contains four phases (on demand service, broad network, resource pooling & Rapid elasticity, measured services), it aims to (i) develop and adopt the rapidly evolving of cloud technology, (ii) abstract the details of inner implementations, and (iii) facilitate the information retrieving service anywhere, anytime [Fan L. et al, 2011].

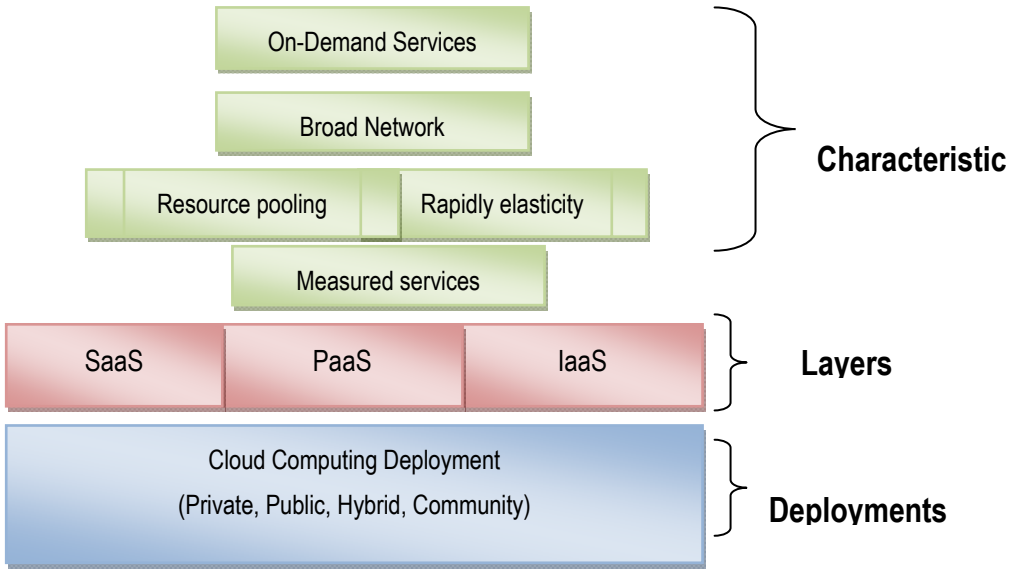


Figure 2. Cloud Environment Architecture

Generally, the model layer consists of three models arranged as follows (see Figure 2):

1. **Infrastructure as a Service** (IaaS): this is a providing service in which the provider is responsible for providing housing, running and maintaining the equipment used to support operations including storage, hardware, servers and networking components. The Amazon web service (S3) [Ruoy Wu. et al, 2012] is an example for IaaS.

2. **Platform as a Service** (PaaS): this service enables the users to use virtualizes servers and associated services for running existing applications or developing and testing new ones, Google Apps are an example for the PaaS [Matt Matlock , 2013; Eman AbuKhousa et al, 2012].
3. **Software as a Service** (SaaS): this service aims to run software on the provider's infrastructure and provide licensed applications to enable users to use the services. Moreover, SaaS offers more transparent to the end user. An example of SaaS is the Salesforce.com CRM application [Pearson S. et al, 2009; Hosseini, 2012].

Health Cloud Technology (HCT)

Over the HIT system will grow largely due to the increasing amount of patient data and additional improvements in the application software that may require more computing power. This will require additional computing resources in order to keep performing efficiently. On the cloud can add more servers with the push of a button and will be transparent to HIT providers. This eliminates the need to buy additional hardware and perform ground-up configuration and disruption that would be required with an in-office solution to keep your system running [Rosado D. et al, 2012].

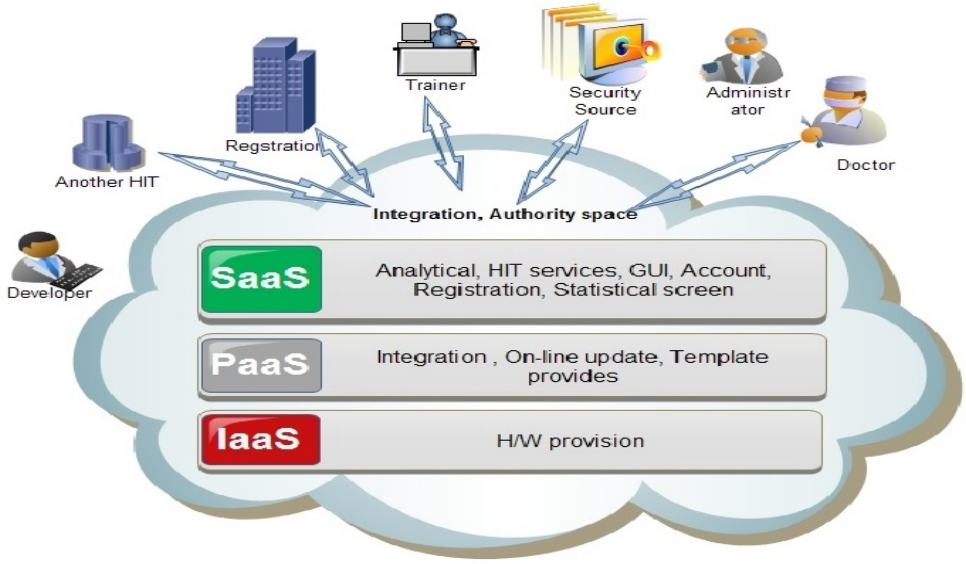


Figure 3. HCT architecture (adopted from [Eman AbuKhousa et al, 2012])

As shown in Figure 3, upgrade the standard cloud architecture to HCT architecture, the HCT consists of an integration area and the three basic layers of cloud. The integration and authority area contribute to achieve following objectives:

I. Accomplish the cloud authority access such as (registration, portal web, connects to another communication system);
II. EHR verification: checking the authority of EHR after migration to the cloud;
III. Integration between traditional EHR files with the Cloudy EHR for HCT;
IV. Determined the type of cloud deployment is used to share the health company files.

Consequently, the other components are classified into three layers same as the basic layers in cloud architecture, firstly, SaaS responsible for the clinical system services, healthcare provider, and system login page, secondly, PaaS utilizes the report testing, updating system, and integration with the other environment, finally, IaaS provides a physical processing and storage requirements.

HCT Challenges and Solutions

The slow adoption of the cloud computing model in the health informatics field is mostly due to two important concerns can be summarized into (i) security and data privacy, (ii) Data probability and Integrity. Those issues need to be fixed in order to overcome obstacles when moving to the cloud environment.

Privacy Concerns

Originally, the exchange of the files and data between traditional HIT systems is not an easy task from the security point view, due to unsecure communication between these companies. Therefore, the Migration of data or file storages to a third party organization is more complex to do, especially when moving sensitive information such as healthcare data. Hence, more robust security should be assured to avoid all concerns to adopt HCT such as access controls, audit controls, authentication, authorization, transmission security and storage security in order to avoid exposing the information to unauthorized entities [He C. et al, 2010; Grobauer, B. et al, 2010].

These issues are an obstacle that has slowed the cloud adoption and should be addressed in order to enable the trustworthiness of cloud systems [Nguyen D. et al, 2012]. Fortunately, many of the biggest cloud providers in the market such as Microsoft, Google, and Amazon have commitments to develop the best policies and practices to secure a customer's data and privacy, also, many researchers and research centers were focusing on privacy point (see Table 1).

Moreover, the data of HIT's unlike other kind of data has strict confidentiality, privacy and security concerns [Soma A., 2011]. The Health Insurance Portability & Accountability Act (HIPAA) compliance is the most fundamental requirement when moving medical records to the cloud as a solution for this challenge. The aims of this system are:

- Reducing costs and enhances the overall efficiency;
- Effectiveness of health care delivery and insurance industry;
- Enhancing the ability of various entities in the healthcare industry to exchange information via standardization;
- Ensure the confidentiality and security of personal health information;
- Ensure portability and continuity of health insurance coverage.

Table 1. Privacy research areas in HCT

Author (s)	Privacy Research Areas	Articles Info
Samuel, O.W. et al., 2013	HIT challenges for secure delivery	**Enhanced Cloud based Model for Healthcare Delivery Organizations in Developing Countries**
Soman A. K., 2011	HIPAA implementation on cloud	**Cloud-based Solutions for Healthcare IT**
Ruoyu Wu et al.,2012	Regulatory level privacy protection	**Towards HIPAA-compliant Healthcare Systems in Cloud Computing**
Jason King et al., 2012	Organizational/System level privacy protection	**Audit Mechanisms in Electronic Health Record Systems: Protected Health Information May Remain Vulnerable to Undetected Misuse**
Thomas Trojer et al, 2012	Personal level privacy protection	**Managing Privacy and Effectiveness of Patient-administered Authorization Policies**
Matt Matlock et al., 2013	Data level privacy protection	**Systematic Redaction for Neuroimaging Data**

Interoperability

Interoperability (Data probability and integrity in some references) is one of the biggest challenges when moving healthcare systems to the cloud, the concern regarding the ability to transition to another cloud vendor or back to the healthcare organization without disrupting operations or introducing conflicting claims to the data [Li M., 2010].

With traditional IT, the healthcare organization has physical control of systems, services and data. The concern is that if a provider were to suspend its services or refuse access to data, a healthcare organization may suddenly be unable to service its patients or customers [H. Liohr, A.et al, 2009; Ortho predict web site, 2014].

The lack of probability and integrity across cloud systems could make it very challenging to migrate to a new cloud service provider. This risk highlights the need for provider agreements that address termination rights, rights to access and retrieve data at any time, termination assistance in moving to another provider to allow a breach of contract to be remedied before the provider terminates or suspends services [Yu W., 2011; Teng C. et al, 2010]. Therefore, a new approach to developing health care systems should be taken in order to design more interoperable systems. This change will result in numerous and substantial benefits to the health community. Table 2 summarizes the studies about data probability constrains and solutions in HCT.

Table 2. Interoperability research areas in HCT

Author (s)	Interoperability Research Areas	Articles Info
Alex Mu. et al., 2011	**Data Migration probability**	Opportunities and Challenges of Cloud Computing to Improve Health Care Services
Sanjay P. Ahuja et al., 2012	**Present a complete survey on moving challenges**	A Survey of the State of Cloud Computing in Healthcare
Ruoyu Wu, et al , 2012	Error rate of data transmission	Secure Sharing of Electronic Health Records in Clouds
Eman AbuKhousa et al., 2012	Interoperability and security challenges in Healthy cloud computing	e-Health Cloud: Opportunities and Challenges

As a solution for this challenge, utilize the concept of Service-Oriented Architecture for implementing the HCT. SOA aims to make services available and easily accessible through standardized models and protocols without having to worry about the underlying infrastructures, development models or implementation details. This helps achieve interoperability and loose coupling among HCT components and also among HCT users [AbuKhousa E. et al, 2012; healthcare site, 2014].

Results and Discussions

Despite of the same challenges have contributed to the slow adoption of the cloud, there are equally as many benefits for providers to embrace this new technology across the enterprise. Based on this study a lot of benefits encompass both business and clinical areas as the following:

1. **Cost**: The cloud allows for health IT managers to avoid the costs of extra on site storage and network infrastructure. In addition, it also allows for greater financial flexibility in health IT because the cloud model is based on a scalable, on demand system.

2. **Reliability:** The central storage of data allows for increased IT responsiveness and efficiency. Disaster recovery is noted as being one of the key benefits of storing information in the cloud.
3. **Portability:** The centralized platform in the cloud allows for health care providers to access vital patient data regardless of the original geographic location that their records were generated remotely from a cloud provider.

Moreover, according to this study, many challenges associated with HCT such as privacy, data probability, integration, and data migration. Figure 4 illustrates these challenges and we noticed through four years from 2010-2013, the privacy and security challenges are taking the bulk from the authors, these solicitudes due to many reasons:

- The open environment for cloud computing technology;
- Lack of security restrictions at the SaaS layer;
- Unsecure communication between cloud providers and clients;
- The weakness associated with the hypervisor layer in most cloud providers acts the cavity for the attacker.

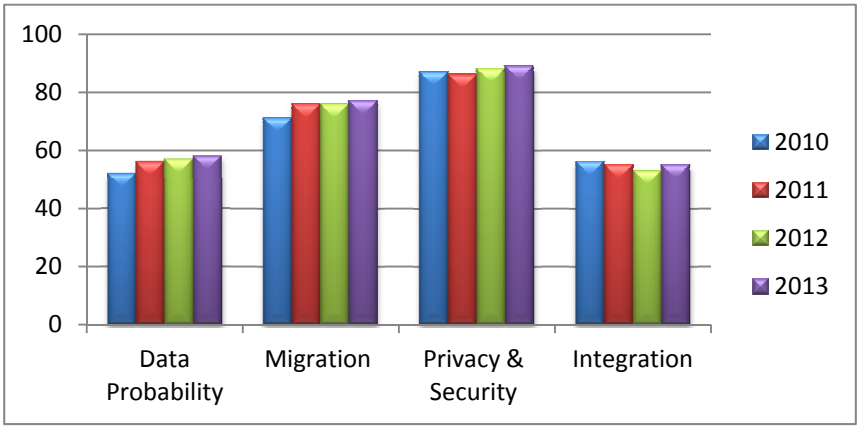


Figure 4. Statistical analysis for HCT challenges

So, a collection of precautions must be followed to overcome the privacy challenge such providing a private IP network isolation for each client, supplying a modern encryption algorithms to guarantee secure communication , and periodically, execute the server load balancing features and automated backups.

Conclusion and Future Work

The current tendency of adopting cloud computing in the medical field can improve and solve several collaborative information issues in healthcare organizations as well as cost optimizations. Standardized cloud-based applications will bring obvious advantages to patients, doctors, insurance companies, pharmacies, imaging centers, etc. when sharing data across medical organizations yielding better outcomes. Challenges such as privacy concerns and Interoperability will rise due to the cloud-computing model and deployment. Thus, the adoption of the cloud is progressing slowly. Through this survey we conclude the HCT will hopefully engender a future development of the cloud-based systems adoption, despite all of the obstructions.

So, in the future we will recommend to design a new healthcare cloud system which able to overcome all challenges. For protection, provide a novel security scheme based on quantum encryption model, and for data, probability, utilize the waterfall model to insure the life cycle of the file migration and consolidation.

Bibliography

[AbuKhousa, E.; Najati, H.A. UAE-IHC, 2012] "Steps towards Integrated E-Health Environment in UAE". In Proceedings of the 4th e-Health and Environment Conference in the Middle East, Dubai, UAE, 30 January 2012–2 February 2012.

[Buyya, J. Broberg, and A. Goscinski, 2011] "Cloud computing Principles and Paradigms," New York, USA, Wiley Press. pp. 1–44.

[Dark Government, 2014] "NSA-Embraces Cloud computing, URL: http://www.darkgovernment.com/news/nsa-embraces-cloud-computing,5:00 pm.

[Eman AbuKhousa, Nader Mohamed and Jameela Al-Jaroodi, 2012] "e-Health Cloud: Opportunities and Challenges", Future Internet 2012, 4, 621-645; doi:10.3390/fi4030621.

[Fan, L.; Buchanan, W.; Thummler, C.; Lo, O.; Khedim, A.; Uthmani, O.; Lawson, A.; Bell, D. DACAR, 2011] "Platform for eHealth Services Cloud". In Proceedings of the 4th International Conference on Cloud Computing, Miami, FL, USA, July 2011; pp. 219–226.

[Grobauer, B.; Walloschek, T.; Stocker, E., 2011] "Understanding cloud computing vulnerabilities. IEEE secures. Privacy Mag. 2011, 9, 50–57.

[Guo, L.; Chen, F.; Chen, L.; Tang, X., 2010] "The building of cloud computing environment for e-health". In Proceedings of the International Conference on e-Health Networking, Digital Ecosystems and Tech. (EDT), Shenzhen, China, 17–18 April 2010; pp. 89–92.

[H. Liohr, A. R. Sadeghi, C. Vishik, and M. Winandy, 2009] "Trusted privacy domains challenges for trusted computing in privacy-protecting information sharing". In Information Security Practice and Experience, 5th International Conference, (ISPEC'09), volume 5451 of Lecture Notes in Computer Science, pages 396 - 400.

[He, C.; Jin, X.; Zhao, Z.; Xiang, T. A ,2010] "Cloud Computing Solution for Hospital Information System". In Proceedings of the Intelligent Computing and Intelligent Systems (ICIS), Xiamen, China, 29–31 October 2010; pp. 517–520.

[Hosseini, A.; Sommerville, I.; Sriram, I., 2012] "Research Challenges for Enterprise Cloud Computing". Available online: http://arxiv.org/abs/1001.3257 (accessed on 28 June 2012).

[King M. Janson, Ben Smith, and Laurie Williams, 2012] "Audit Mechanisms in Electronic Health Record Systems: Protected Health Information May Remain Vulnerable to Undetected Misuse", International Journal of Computational Models and Algorithms in Medicine, 3(2).

[Li, M.; Yu, S.; Ren, K.; Lou, W., 2011] "Securing Personal Health Records in Cloud Computing: Patient-centric and Fine-grained Data Access Control in Multi-owner Settings". In Proceedings of the 6th International ICST Conference on Security and Privacy in Communication Networks (SecureComm 2010), Singapore, 7–9 September 2010; pp. 89–106.

[Matt Matlock, Nakeisha Schimke, Liang Kong, Stephen Macke, and John Hale, 2013] "Systematic Redaction for Neuroimage Data", IJCA,Vol.5, No. 2.

[Nguyen, D.K.; Lelli, F.; Papazoglou, M.P.; van den Heuvel, W.-J., 2012] "Blueprinting Approach in Support of Cloud Computing. Future Internet ,PP 322–346.

[Omer K., Safia M., El-Sayed M., and Abed-Albadaa S., 2013] "Cloud Computing Cryptography: state of the art", International Journal of Computer, Information Science and Engineering Vol:7 No:8, 2013

[online available] http://www.healthcare-informatics.com/article/orthopedic-care-clinic-franchise-launched\

[Online available] http://www.fiercehealthit.com/story/cloud-computing-healthcare-question-not-ifwhen/2012-01-0.

[Pearson, S., 2009] "Taking Account of Privacy when Designing Cloud Computing Services". In Proceedings of the ICSE Workshop on Software Engineering Challenges of Cloud Computing at CLOUD'09, Washington, DC, USA, 23 May 2009; pp. 44–52.

[R. Alex Mu, Hsing Kuo, 2011] "Opportunities and Challenges of Cloud Computing to Improve Health Care Services" International Journalof information sciences, No 22, Vol 3.

[Rosado, D.G.; Gómez, R.; Mellado, D.; Fernández-Medina, E.2012] "Security analysis in the migration to cloud environments". Future Internet 2012, 4, 469–487.

[Ruoyu Wu, Gail-Joon Ahn and Hongxin Hu, 2010] "Secure Sharing of Electronic Health Records in Clouds", Proceedings of the 1st ACM International Health Informatics Symposium.

[Ruoyu Wu, Gail-Joon Ahn and Hongxin Hu, 2012] "Towards HIPAA-compliant Healthcare Systems in Cloud Computing", International Journal for Computing and Algorithms, Vol. 3, No. 20.

[Samuel, O.W., Omisore, M.O., Ojokoh, B.A., and Atajeromavwo, E.J., 2013] "Enhanced Cloud based Model for Healthcare Delivery Organizations in Developing Countries", International Journal of Computer Applications (0975 – 8887) Volume 74– No.2.

[Sanjay P. Ahuja, Sindhu Mani and Jesus Zambrano, 2012] "A Survey of the State of Cloud Computing in Healthcare", Network and Communication Technologies; Vol. 1, No. 2.

[Soma A. K., 2011] "Cloud-based Solutions for Healthcare IT", Published by Science Publishers, P.O. Box 699, Enfi eld, NH 03748, USA An imprint of Edenbridge Ltd., British Channel Islands, Copyright reserved- 2011 ISBN 978-1-57808-702-0

[Srinivasa, R., Nageswara, R., and Ekusuma, K., 2009] "Cloud computing: An overview", Journal of Theoretical and Applied Information Technology (JATIT), pp. 71-76, 2009.

[Sunyaev J., Leimeister H., and Krcmar H., 2010] "Open security issues in German healthcare telematics". In HEALTHINF 2010 - Proceedings of the 3rd International Conference on Health Informatics, pages187-194.

[Teng C.C.; Mitchell, J.; Walker, C. A, 2010] "Medical Image Archive Solution in the Cloud." In Proceedings of the 2010 IEEE International Conference on Software Engineering and Service Sciences (ICSESS), Beijing, China, 16–18 July 2010; pp. 431–434.

[Thomas Trojer, Basel Katt, Ruth Breu, Thomas Schabetsberger, and Richard Mair, 2012] "Managing Privacy and Effectiveness of Patient-admin¬istered Authorization Policies", International Journal of Computational Models and Algorithms in Medicine, 3(2), 43-62.

[Yu, W.D. A, 2011] "Service Modeling Approach to Service Requirements in SOA and Cloud Computing—Using a U-Healthcare System Case". In Proceedings of the IEEE 13th International Conference on e-Health Networking, Applications and Services, Columbia, MO, USA, 13–15 June 2011; pp. 233–236.

Authors' Information

Omer K. Jasim *- Ph.D. student at faculty of computer and information sciences, Ain Shams University, ZIP (+2), Cell phone: 01154567015; e-mail: omarkj@auc-edu.org*

Iraq nationality

Safia Abbas *- she is lecturer at faculty of computer and information sciences, Ain Shams University, Cairo-Egypt; e-mail: safia_abbas@cis.asu.edu.eg*

El-Sayed M. El-Horbaty *- he is professor in faculty of computer and Information sciences, Head of Computer Science department, Ain Shams University, Abbasia, Cairo-Egypt; e-mail: Shorbaty@cis.asu.edu.org*

Abdel-Badeeh M. Salem *- he is professor in faculty of computer and information sciences, Head of BioMedical Informatics and Knowledge Engineering Research Lab, Ain Shams University, Abbasia, Cairo-Egypt; e-mail: absalem@cis.asu.edu.eg*

MATRIX "FEATURE VECTORS" IN GROUPING INFORMATION PROBLEM: LINEAR DISCRIMINATION

Volodymyr Donchenko, Fedir Skotarenko

***Abstract**: The problem of classification, clusterization or patterns recognition is one the manifestation of grouping information problem (GIP) in applied researching. It involves, beside mentioned above, the problem of recovering function, represented by empirical data (observations). Solutions of GIP largely depend of on the choice of "math representatives" of the objects under investigation. It`s usual to use a collection of real valued characteristics – "feature vector", in classification form of the GIP. Feature vector is in the essence a vector from Euclidean space R^n. This choice is due to the highly advanced ties - and correspond techniques- in mathematical structure of such type. This technique includes, particularly, spectrum of linear operator (SVD), Moore-Penrose inversion, orthogonal projectors operators for fundamental subspaces of the linear operator, Grouping operators and so on. Euclidean spaces $R^{m\times n}$ of all matrixes of fixed dimension are natural spaces of "representatives" for a great many important applied fields of investigations: speech recognition, image processing and so on. In the paper SVD and Moore – Penrose technique for $R^{m\times n}$, proposed and developed in the earlier paper of the authors published in 2012 is used for formulating and solution of linear discrimination of two classes, represented by matrix learning samples.*

***Keywords:** Feature vectors, information aggregating, matrix corteges, matrix corteges operators, Single Valued Decomposition for cortege linear operators, linear discrimination.*

***ACM Classification Keywords**: G.2.m. Discrete mathematics: miscellaneous, G.1.6. Numerical analysis, I.5.1. Pattern Recognition, H.1.m. Models and Principles: miscellaneous*

Introduction

Grouping information problem (GIP) is fundamental problem in applied investigations. There are two main form of it, namely: the problem of recovering the function, represented by their observations, and the problem of clustering, classification and pattern recognition. Examples of approaches in the field are represented perfectly in [Kohonen, 2001], [Vapnik, 1998], [Haykin, 2001], [Friedman, Kandel, 2000], [Berry, 2004]. It is opportune to notice, that math modeling is the representation of an object structure by the means of mathematical structuring. A math structure after Georg Cantoris is a set plus "ties" between its elements. Only four fundamental types of "ties" (with its combination as fifth one) exist: relations, operations, functions and collections of subsets. Thus, the mathematical description of the object (mathematical modeling) can not be anything other than representing the object structure by the means of mathematical structuring. It refers fully to so call "complex system". A "complex system" should be understanding and, correspondingly, determined, as an objects with complex structure (complex "ties"). Namely, when reading attentively manuals by the theme (see, for example, [Yeates, Wakefield, 2004], [Forster, Hölzl, 2004]) one could find correspondent allusions. "Structure" understanding is reasonable determining of a "complex systems" instead of defining them as the "objects, consisting of numerous parts, functioning as an organic whole".

In the essence, math modeling is representing by math "parts plus ties" of the object in applied field.

It is usual in GIP to represent object under consideration by the ordered collection of characteristics: quantitative (numerical) or qualitative (non numerical). Such ordered collection with real numbered characteristics is called feature vector and thus can be considered, naturally, as element of R^n. Sometimes such collection is not collection of numbers and called cortege this case. In clustering and classification problem the collection may be both qualitative and quantitative. Feature vector case is more attractive since it allows using structural diversity of Euclidean space R^n, namely: linear operations (addition and scalar multiplying), scalar product and orthogonality, norm and distance.

Euclidean space R^n is not unique, which naturally appears in applications: the space $R^{m\times n}$ of all matrixes of a fixed $m \times n$ dimension is another example. Using off R^n in applied researches is determined largely by sophisticated techniques developed for R^n - vectors handling. Namely, these are: matrix algebra, spectrum technique (Single Valued Decomposition – SVD), Pseudo Inverse by Moore – Penrose (PIM-P) [Nashed, 1978] (see, also, [Albert, 1972], [Ben-Israel, Greville, 2002]. One cannot mention in this context the outstanding contribution of N.F. Kirichenko in development of PIM-P – technique for R^n (especially, [Кириченко, 1997] [Kirichenko, 1997], see also [Кириченко, Лепеха, 2002]). Greville's formulas: forward and inverse -for PIM-P matrixes, formulas of analytical representation for disturbances of PIM-P, - are among them. Additional results in the theme as to further development of the technique and correspondent applications one can find in [Кириченко, Лепеха, 2001], [Donchenko, Kirichenko, Serbaev, 2004], [Кириченко, Крак, Полищук, 2004], [Kirichenko, Donchenko, Serbaev, 2005], [Кириченко, Донченко, 2005], [Donchenko, Kirichenko, Krivonos, 2007], [Кириченко, Донченко, 2007], [Кириченко, Кривонос, Лепеха, 2007], [Кириченко, Донченко, Кривонос, Крак, Куляс, 2009].

As to technique designing for the Euclidean space $R^{m\times n}$ as "environmental" math structure first steps have been made for example, by [Донченко, 2011], [Donchenko, Zinko, Skotarenko, 2012]. Speech recognition with the spectrograms as the representative and the images in the problem of image processing and recognition are the natural application areas for the correspond technique.

As to the choice of the collection (design of cortege or vector) it is necessary to note, that good "feature" selection (components for feature vector or cortege or an arguments for correspond functions) determines largely the efficiency of the problem solution. This phase in solving the grouping information problem is the special step of the investigation. Experience indicates that this step should be arranged in the form of recurrent selection procedures: pre-selection and subsequent improvement of the feature characteristics. Vivid examples of such approach are the next publications on [Ivachnenko, 1995] (also [Ivachnenko, 1969] with Ivachnenko's GMDH (Group Method Data Handling) and [Vapnik, 1998] with Vapnik's Support Vector Machine. Further development of the recurrent approach in feature selection through the development and systematical application of advanced PIMP technique with criteria for estimation of feature informative significance one can find in [Donchenko, Kirichenko, Serbaev, 2004], [Кириченко, Крак, Полищук, 2004], [Kirichenko, Donchenko, Serbaev, 2005], [Кириченко, Донченко, 2005], [Donchenko, Kirichenko, Krivonos, 2007], [Кириченко, Донченко, 2007], [Кириченко, Кривонос, Лепеха, 2007], [Donchenko, Krak, Krivonos, 2012]. The idea of nonlinear recursive regressive transformations (generalized neuron nets or neurofunctional transformations) due to Professor N.F Kirichenko is represented in the works referred earlier.

Correspondent technique has been designed in this works separately for each of two its basic form f the grouping information problem. The united form of the grouping problem solution is represented here in further consideration. The fundamental basis of the recursive neurofunctional technique includes the development of pseudo inverse theory in the publications mentioned earlier first of all due to Professor N.F. Kirichenko and his disciples.

The essence of the idea mentioned above is thorough choice of the primary collection and changing it if necessary by standard recursive procedure. Each step of the procedure include detecting of insignificant components, excluding or purposeful its changing, control of efficiency of changes has been made. Correspondingly, the means for implementing the correspondent operations of the step must be designed. Methods of neurofunctional transformation (NfT) (generalized neural nets, nonlinear recursive regressive transformation: [Donchenko, Kirichenko, Serbaev, 2004], [Кириченко, Крак, Полищук, 2004], [Кириченко, Донченко, Сербаєв, 2005]).

There are two basic approaches in solving to solving classification- clusterization form of GIP when Euclidean space is "environmental" space: using of recurrent procedure of k-means type and discrimination with linear discrimination as a base. First approach needs and use so called distance of conformity with classes or clusters. Variants of such distances based on advanced PIMP-technique one can find, for example, in [Кириченко, Донченко, 2007], [Донченко, 2011]. In one can find the development of the "distance of conformity" approach for $R^{m\times n}$, based on developing of PIMP-technique for Euclidean spaces of $R^{m\times n}$ - type.

The linear discrimination (LD) form for classification - clusterization variant of GIP for $R^{m\times n}$ is formulated below in the proposed paper and solved fully on the base of PIMP-technique developed the [Donchenko, Zinko, Skotarenko, 2012], has been cited earlier.

Linear discrimination as a form classification- clusterization in GIP – problem - formalization

Linear discrimination as a form of clusterization and classification of GIP-problem (Cl-Cl GIP) for Euclidean spaces of R^n - type has been discussed and solved fully on the base of PIMP – technique in [Кириченко, Кривонос, Лепеха, 2007], [Donchenko, Krak, Krivonos, 2012] including designing of recurrent selection procedure as well as criteria of informative significance components of feature vector.

In this paper we apply the ideas of papers just have been cited for formulating and solving fully linear discrimination problem for Euclidean spaces of $R^{m\times n}$ - type for two classes, represented by learning samples.

We will reference these classes by Cl_1, Cl_2 with united learning sample $X(j)\in R^{m\times n}, j=\overline{1,N}$ and with J_1, J_2 - partition of index set $\{1,...,N\}$ which corresponds leaning samples for each of the classes:

$$J_1, J_2 \subseteq \{1,...,N\}:\ J_1\cap J_2=\varnothing,\ J_1\cup J_2=\{1,...,N\},$$
$$X(j)\in Cl_k \Leftrightarrow j\in J_k,\ k=1,2,$$

$j=\overline{1,N}$

We mean by LD - problem in $R^{m\times n}$ (linear discrimination problem) of two classes Cl_1, Cl_2 represented by the parts $X(j), j\in J_1,\ X(j), j\in J_2$ of a united learning sample $X(j), j=\overline{1,N}$, the problem of designing linear functional $\varphi: R^{m\times n}\to R^1$ (discrimination function) $\Delta>0$, which would "$\Delta-$ differentiate" classes for some $\Delta>0$, in the sense, that:

$$\begin{aligned} y_j &= \varphi(X(j))\geq \Delta, j\in J_1, \\ y_j &= \varphi(X(j))\leq -\Delta, j\in J_2 \end{aligned} \qquad (1)$$

Linearity for functional φ means, that it can be uniquely represented through the inner product i.e. that $m\times n$ - matrix A exists such, that

$$\varphi(X) = (A,X)_{tr} \tag{2}$$

Dot product $(A,B)_{tr}, A = (a_{ij}), A = (a_{ij}) \in R^{m\times n}$ - is a trace inner product, determined in the standard way by the equation

$$(A,B)_{tr} = \sum_{i=\overline{1,m}, j=\overline{1,n}} a_{ij} b_{ij} ,$$

or, equivalently, by the sum of diagonal elements (trace) of matrix product $A^T B$:

$$(A,B)_{tr} = A^T B .$$

We will denote by $\Omega(\Delta)$ for any $\Delta > 0$ the subset of all vectors $y \in R^N : y^T = (y_1,\ldots,y^N)$ such, that its components $y(j), j = \overline{1,N}$ satisfy inequalities from (1):

$$\Omega(\Delta) = \{y \in R^N : y^T = (y_1,...,y_N), y_j > \Delta, j \in J_1, y_j < -\Delta, j \in J_2\} .$$

We will use also denotation $\wp_\alpha^*$ with matrix cortege

$$\alpha = (A_1,\ldots,A_N), A_j \in R^{m\times n}, j = \overline{1,N} \tag{3}$$

for linear operator from $R^{m\times n}$ to R^N , defined by the equation

$$\wp_\alpha^* Y = \begin{pmatrix} (A_1,Y)_{tr} \\ \cdots \\ (A_N,Y)_{tr} \end{pmatrix} = \begin{pmatrix} trA_1^T \\ \cdots \\ trA_N^T Y \end{pmatrix}.$$

It has been proven in [Donchenko, Zinko, Skotarenko, 2011] that $\wp_\alpha^*$ is conjugate to a so called cortege operator $\wp_\alpha : R^N \to R^{m\times n}$, defined for a matrix cortege from (3 by the equation

$$\wp_\alpha x = \sum_{j=1}^{N} x_j A_j, x \in R^N, x^T = (x_1,...,x_N) \tag{4}$$

Thus, in the notations, have been introduced earlier, the text theorem is true.

Theorem 1. In the notation introduced previously LD-problem in $R^{m\times n}$ is equivalent to the solving of conditional system of linear equations

$$\wp_{\alpha_L}^* A = y, y \in \Omega(\Delta) \tag{5}$$

with cortege α_L , designed from the matrixes of united Learning sample:

$\alpha_L = (A(1),\ldots,A(N))$.

Prove. Indeed, system of the inequalities (1) is equivalent, that real-valued vector *y* with the components from (1) and functional φ from (2) is valid next statement

$$y = \begin{pmatrix} y_1 \\ \cdots \\ y_N \end{pmatrix} = \begin{pmatrix} (A,A(1))_{tr} \\ \cdots \\ (A,A(N))_{tr} \end{pmatrix} \in \Omega(\Delta)$$

.

Then, by mentioned above theorem from [Donchenko, Zinko, Skotarenko, 2011] formula

$$\begin{pmatrix} (A,A(1))_{tr} \\ \cdots \\ (A,A(N))_{tr} \end{pmatrix}$$

define a linear operator from $R^{m\times n}$ to R^N, which are conjugate to cortege operator $\wp_{\alpha_L}$ from $R^{m\times n}$ to R^N with cortege α_L, defined by united learning sample:

$$\alpha_L = (A(1),\ldots,A(N))$$

$$\wp^*_{\alpha_L} A = \begin{pmatrix} (A,A(1))_{tr} \\ \cdots \\ (A,A(N))_{tr} \end{pmatrix}.$$

Thus

$\wp^*_{\alpha_L} A \in \Omega(\Delta)$

We denote by X Gramian matrix for the united collection of Learning sample matrixes:

$X = \left((A(i),A(j)_{tr})\right)_{i,j=\overline{1.N}}$.

The next theorem is valid in the notations have been introduced.

Theorem 2. LD-problem is equivalent to solvability of quadratic optimization problem for $y^T Z(X)y$ in domain $\Omega(\Delta)$ i.e. it is necessary and sufficient for existence of LD-problem solution that minimum $y_* \in R^N$ of quadratic form $y^T Z(X)y$ belongs to $\Omega(\Delta)$:

$$y_* = \arg\min_{y\in R^N} y^T Z(X)y, y_* \in \Omega(\Delta) \tag{3}$$

where

$$Z(X) = E_N - X^+ X,$$

and X^+ - Moore-Penrose pseudo inverse for matrix X as linear operator from R^N in R^N (see, for example, [Albert, 1972]).

Prove. Indeed, condition $\wp^*_{\alpha_L} A = y \in \Omega(\Delta)$ indicate, that for some $y \in \Omega(\Delta)$ linear equation $\wp^*_{\alpha_L} A = y$ is solvable for some $y \in \Omega(\Delta)$. This means, that y belongs to range of $\wp^*_{\alpha_L}$: $y \in \Re\left(\wp^*_{\alpha_L}\right)$.It is obvious, that

$\Re\left(\wp^*_{\alpha_L}\right) = \Re\left(\wp_{\alpha_L}\wp^*_{\alpha_L}\right) = \Re(X)$.

Belonging to linear subspace or range means that it is a fixed point of the correspond orthogonal projector. As the $\Re\left(\wp^*_{\alpha_L}\right) = \Re(X)$ correspond orthogonal projectors coincides $P_{\Re\left(\wp^*_{\alpha_L}\right)} = P_{\Re(X)}$, so

$$P_{\Re(X)} y = y \tag{6}$$

Consequently

$y = P_{\Re(X)} y$,

or

$\left(E_N - P_{\Re(X)}\right)y = 0$.

Orthogonal projector $P_{\Re(X)}$ uniquely determined by pseudo inverse for X according to the next equality

$P_{\Re(X)} = X^{+}X$.

Thus (6) one can rewrite

$\left(E_N - X^{+}X\right)y = 0$,

or, equivalently, in notation of, for example [Kirichenko,1997]

$$Z(X)y = 0 \quad . \tag{7}$$

In its turn, last equality is equivalent to $y^T Z(X)y = 0$.

Last equality means that absolute minimum of nonnegative quadratic form $y^T Z(X)y, y \in R^N$ is achieved in domain $\Omega(\Delta)$. It is equivalent, that there exists a $y_* \in \Omega(\Delta)$ which is minimum of $y^T Z(X)y, y \in \Omega(\Delta)$ and the minimum value is zero:

$y_*^T Z(X)y_* = 0, y_* \in \Omega(\Delta)$.

This is the finish if the prove.

Insolvability of the optimization problem with constraints from *Theorem 2* means insolvability LD-problem with the feature matrixes of the model. So the features need purposeful change, for the matrixes feature (matrixes "feature vector") now. So, criteria for the choice of correspondent components and means for correspondent changes must be available, just as that was exposed in [Кириченко, Кривонос, Лепеха, 2007], [Donchenko, Krak, Krivonos, 2012] for feature vector from R^n.

Conclusion

Conception of enriching the standard considering the "representatives" in Applied Math to be the feature vectors: elements from Euclidean space R^n - has been further developed in the paper (see, also, [Donchenko, Zinko, Skotarenko, 2012]). Using matrixes as the "representatives" of the real objects is main idea of the conception. This mean, that matrix instead vector represents all principal features of the objects in applied fields. Support of this concept requires the development of technologies handling with matrixes similar techniques operating with vectors from Euclidean spaces R^n. SVD-technique as well as PIMP - technique are the priority among them. The results of such type are represented in the paper. These results demanded a generalization of matrix algebra and transforming it in algebra of matrix and vector cortege as well as definition and using the linear cortege operator. Correspond results are represented in the paper of the authors [Donchenko, Zinko, Skotarenko, 2012]. Using that handling technique for matrix features ("matrix feature vectors") make it possible to put and fully solute the Linear Discrimination problem for two collection of matrixes. Correspond solution uses standard SVD and PIMP for Gramian matrix of united collections and solution of quadratic optimization in a domain of appropriate R^n. Thus, the development of matrix technique manages to reduce to existing technique for real valued vectors. Solution of Linear Discrimination Problem for matrixes is similar to correspond result for real-valued vectors in [Кириченко, Кривонос, Лепеха, 2007] or [Donchenko, Krak, Krivonos, 2012]. The two obvious application areas are worth mentioning within the context of the application of these results. These are: speech recognition and image

processing. Matrixes naturally represent the objects under consideration, namely, spectrograms and digital images.

Bibliography

[Albert, 1972] Albert A.E. Regression and the Moore-Penrose pseudoinverse.-Academic Press., 1972, 180 p.

[Ben-Israel, Greville, 2002] Adi Ben-Israel, Thomas N.E. Greville Generalized Inverses. Theory and Applications.- Second Edition.- Springer-Verlag New York, Inc. -2003.-420 p

[Berry, 2004] Michael W. Berry Editor Survey of Text Mining: Clustering, Classification, and Retrieval Springer - Verlag New York, Inc. -2004.-244 p.

[Donchenko, Kirichenko, Krivonos, 2007] Donchenko V. Kirichenko M., Krivonos Yu. Generalizing of neural nets: Functional nets of special type// International Journal Information Models and Analysis. – N 3. –V.14. – 2007. – P. 259-266.

[Donchenko, Kirichenko, Serbaev, 2004] Donchenko V.S. Kirichenko M.F., Serbaev D.P. Recursive regression transformation and dynamical systems// Proceedings: of the Seventh International Conference "Computer Data analysis and Modeling: robustness and computer intensive methods". – V.1. – September 6-10, 2004. – Minsk. – P. 147-151

[Donchenko, Krak, Krivonos, 2007] Donchenko V. Krak Ju.., Krivonos Yu. Recurrent procedure in solving the Grouping Information Problem in Applied Mathematics//ITHEA: International Journal Information Theories and Applications. – V.1.– N 1. – 20012. – P.62 -77.

[Donchenko, Zinko, Skotarenko, 2012] Donchenko V, Zinko T, Skotarenko F. "Feature Vectors" in Grouping Information Problem in Applied Mathematics: Vectors and Matrixes. - Vitalii Velichko, Alexey Voloshin, Krassimir Markov(ed.)- Problems of Computer Intellectualization.- Institute of Cybernetics NASU, ITHEA.-Kyiv, Ukraine -Sofia, Bulgaria.- 2012.- P.111-124.

[Forster, Hölzl, 2004] Forster J., Hölzl V. Applied Evolutionary economics and complex systems.-2004-- 293 p.

[Friedman, Kandel, 2000] Menahem Friedman Abraham Kandel Introduction to pattern recognition statistical, structural, neural and Fuzzy logic approaches.- World scientific publishing co. Pte. Ltd. - reprinted 2000.-329 p.

[Haykin, 2001] Haykin S. Neural Networks. A Comprehensive Foundation.-Prentice Hall, Upper Saddle River, New Jersey 07458. – 1999.– 842 p

[Ivachnenko, 1969] Івахненко А.Г. Системи розпізнавання і автоматичного управління, що самоорганізуються. – К.: Техніка.-1969.–395 с.

[Ivachnenko, 1995] Ivachnenko A.G and Muller J.-A. Present state and new problems of further GMDH development.-SAMS, 1995.- pp.3-16.

[Kirichenko, 1997] Kirichenko N.F. Analytical Representation of Perturbation of Pseudo inverse Matrices.// Cybernetics and Systems Analysis. -. March-April 1997. - Vol.33. - Number 2.- P.230-239.

[Kirichenko, Donchenko, Serbaev, 2005] Kirichenko N.F., Donchenko V.S. Serbaev D.P. Nonlinear recursive nonlinear Transformations: Dynamic Systems and Optimizations.//Cybernetics and System Analysis. - May 2005. – V.41. - № 3. – P. 364-373.

[Kohonen, 2001] Kohonen T. Self-Organizing Maps -3-d ed. - Tokyo: Springer, 2001.- 501 p

[Nashed, 1976] Nashed M. Zuhair, Votruba G.F., editors. A Unified Operator Theory of Generalized Inverse.//Proceedings of an Advanced Seminar Sponsored by the Mathematical Research Center, The University of Wisconsin, Madison, October 8-10, 1973. – New York, Academic Press, 1976.–1200 p.

[Vapnik, 1998] Vapnik, V.N. Statistical Learning Theory. New York: Wiley, 1998

[Yeates, Wakefield, 2004] Yeates D., Wakefield T. Systems Analysis and Design. - Pearson Education. – 2004. - 499 p.

[Донченко, 2011] Владимир Донченко Евклидовы пространства числовых векторов и матриц: конструктивные методы описания базовых структур и их использование.// International Journal "Information technologies & Knowledge".- 2011.- Vol. 5.- Number 3.-P.203-216.

[Кириченко, Донченко, 2005] Кириченко М.Ф., Донченко В.С. Задача термінального спостереження динамічної системи: множинність розв'язків та оптимізація//Журнал обчислювальної та прикладної математики. – 2005. –№5– С.63-78.

[Кириченко, Донченко, 2007] Кириченко Н.Ф., Донченко В.С. Псевдообращение в задачах кластеризации.//- Кибернетика и системный анализ.- 2007.- № 4.- с. 73 –92

[Кириченко, Донченко, Кривонос, Крак, Куляс, 2009] Кириченко М.Ф., Донченко В.С Кривонос Ю.Г., , Крак Ю.В.,., Куляс А.І. Аналіз та синтез ситуацій в системах прийняття рішень. - Київ: Наукова думка.-2009. 336с.

[Кириченко, Донченко, Сербаев, 2005] Кириченко Н.Ф., Донченко В.С., Сербаев Д.П. Нелинейные рекурсивные регрессионные преобразователи: динамические системы и оптимизация.// Кибернетика и системный анализ.– №3, 2005.– С. 58-68.

[Кириченко, Крак, Полищук, 2004] Кириченко Н. Ф., Крак Ю. В., Полищук А.А. Псевдо обратные и проекционные матрицы в задачах синтеза функциональных преобразователей.// Кибернетика и системный анализ –2004.–№3.

[Кириченко, Кривонос, Лепеха, 2007] Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П. Синтез систем нейрофункциональных преобразователей в решении задач классификации. // Кибернетика и системный анализ. – 2007. – №3. С. 47-57

[Кириченко, Лепеха, 2002] Кириченко Н.Ф., Лепеха Н.П. Применение псевдо обратных и проекционных матриц к исследованию задач управления, наблюдения и идентификации.// Кибернетика и системный анализ. – 2002. – № 4. – С. 107-124.

[Кириченко, 1997] Н.Ф. Кириченко Н.Ф. Аналитическое представление возмущений псевдо обратных матриц.// Кибернетика и системный анализ. – 1997. – № 2. – С. 98-107.

Authors' Information

Volodymyr Donchenko *– Professor, National Taras Shevchenko University of Kyiv. 60, Volodymyrs'ka street, Kyiv, 03680, Ukraine; e-mail: voldon@bigmir.net.*

Fedir Skotarenko – *post-graduate student, National Taras Shevchenko University of Kyiv. 60, Volodymyrs'ka street, Kyiv, 03680, Ukraine.*

STANDARDIZATION OF GEOMETRICAL CHARACTERISTICS IN GESTURE RECOGNITION

Andrew Golik

***Abstract**: This paper covers approach to obtaining geometrical characteristics, in particular defects, for gesture shown in front of web camera. Variant of standardization of defect`s structure is suggested. The paper provides detailed algorithm that allows obtaining standardized feature vector for defect of any size and structure. It is suggested to use obtained feature vectors for recognition of tactile sign language.*

***Keywords**: gesture recognition, defects, standardization, normalization, tactile sign language.*

***ACM Classification Keywords**: I.2 Artificial Intelligence, I.4 Image Processing and Computer Vision, I.5 Pattern Recognition, G.1.3 Numerical Linear Algebra.*

Introduction

Recognition of tactile sign language is an important applied task. There exist people who need such recognition systems not for entertainment, but for everyday life. They are deaf and deaf-and-dumb people. In case when person lost hearing at mature age usually it is possible to continue full-fledged life, but children, who lost hearing in the early childhood, are in much more difficult situation. It is an important social task to help such children. Of cause, availability of money for hearing aids and other equipment partially solves the problem, but children in many orphanages remain without appropriate support. So developing of recognition system, which does not require expensive devices like sensors and can be used for low cost, is an actual task. [1]

Principle of allocation and highlighting of a contour of hand on captured by web camera image is given. Geometrical characteristics, which can be effectively used for gesture recognition, namely defects, are covered. Algorithm of obtaining standardized feature vectors for defects is suggested.

Finding and highlighting of a contour of hand

The majority of approaches to finding of a contour of hand on an image are connected with pixel analysis. Usually two color models, RGB and HSV, are used. RGB is an additive color model in which red, green, and blue light are added together in various ways to reproduce a broad array of colors. HSV is one of most common cylindrical-coordinate representations of points in an RGB color model. HSV stands for hue, saturation, and value. Both models have advantages and shortcomings. However, HSV model is preferable in gesture recognition. This model allows considering hue as separate component. It is important for situations when finding a contour of hand is connected with finding areas that correspond to color of skin. In this case, smoothing and filtration, which helps to get rid of noises, are applied too. This task is not trivial because of environment conditions (lighting, background and so on) which have direct influence on results of recognition. Existing systems require detailed configuration before someone can use them in certain environment. This problem can be partially solved with usage of a red mitten.

The most useful for implementation of mentioned above functionality is OpenCV library. It is written on C++, but there are many wrappers, which allow using it in different programming environments, for example Java (JavaCV) or C# (EmguCV). The library offers great capabilities for image analysis, processing and smoothing, finding of

contours and so on. While developing of recognition system for tactile sing language, stage of finding of a contour of hand was implemented. Illustration of corresponding results is shown in Figure 1. [1]

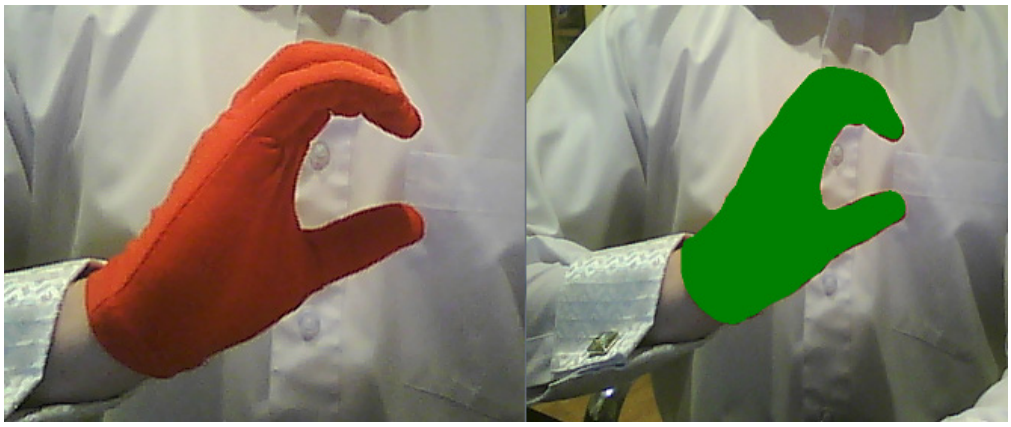


Figure 1. A contour of hand is found and highlighted.

Convex hull and defects

Information about highlighted part of an image is presented as a closed contour that is a formalized description of a captured object. The most useful information is stored in "defects". Defects are geometrical characteristics of gesture, which allow identifying of tactile sign. They are characterized by relative positioning of a contour of hand and convex hull built for this contour. In addition, defects can be considered as "curvilinear triangles" with (blue) parts of convex hull as basis (Figure 2).

Highlighted contour of hand and convex full for this contour is shown in Figure 2. The figure represents four defects, start, end and depth points of each of them. However, there are only "significant" defects on the figure and there exist many small defects most of which even cannot be seen at the first sight.

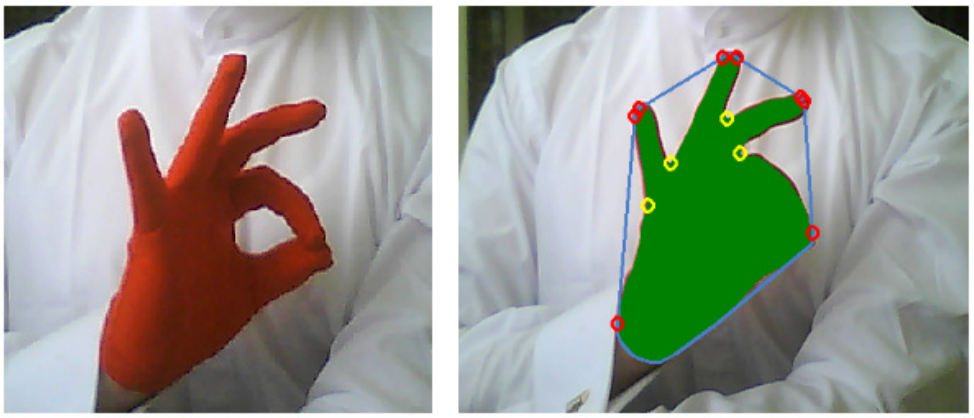


Figure 2. Finding of a contour of hand, convex hull and defects

Defects are sorted in order to find "significant" defects, which can be effectively used for recognition. This stage is rather complex because criteria of importance of defect are not trivial. Following parameters can be used for sorting: length and depth of defect, relation of length to depth and so on. The most efficient approach is to use

rules like "if length more than A, depth more than B and relation of length to depth is more than C then defect is significant". [1, 2]

Formation of standardized feature vectors for structure of defect

As mentioned in previous section there are numerical characteristics which can be obtained for defects, namely length, height, area, perimeter and so on. In addition, usage of relative values shows their efficiency on practice. However, structure of defect deserves special attention. The paper provides an approach to formation of feature vectors that represent structure of defect in the most adequate way.

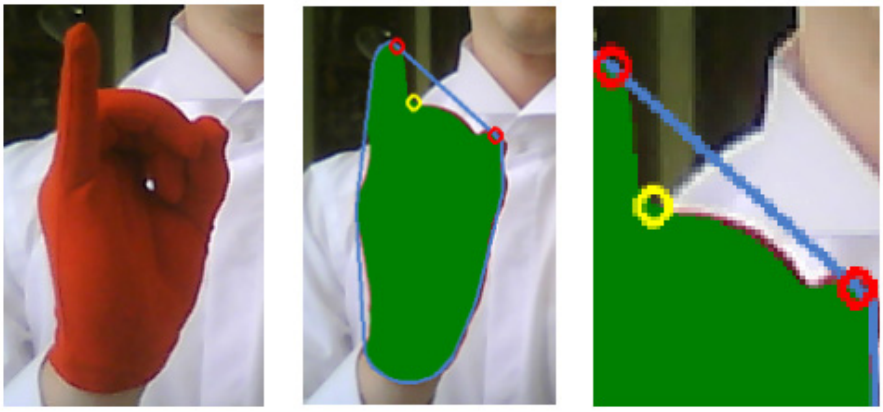


Figure 3. Tactile sign "I" with highlighted defect.

Gesture with one significant defect is shown in Figure 3. Structure of defect is described by positions of points of a contour within it. It should be represented in such way that would allow not only displaying positions of mentioned points, but also comparing structures of different defects and finding similarity degree between them. Algorithm of formation of feature vectors, which meet all the requirements, is given below:

1. Obtain equation of line that passes through start $c_{start}(x_1, y_1)$ and end $c_{end}(x_2, y_2)$ points of defect:

$$(y_2 - y_1)x + (x_1 - x_2)y + (x_2 y_1 - x_1 y_2) = 0;$$

$$\mathrm{A} = y_2 - y_1; \mathrm{B} = x_1 - x_2; \mathrm{C} = x_2 y_1 - x_1 y_2;$$

$$\mathrm{A}\,x + \mathrm{B}\,y + C = 0;$$

2. Having coordinates of points of a contour within defect $c_i(x_{i,} y_i), i = \overline{start, end}$ for each point obtain distance between this point and line from step 1 (Figure 4).

$$d_i = \frac{|\,\mathrm{A}\,x_i + \mathrm{B}y_i + \mathrm{C}\,|}{\sqrt{(\mathrm{A}^2 + \mathrm{B}^2)}}, i = \overline{start, end};$$

Numerical vector is obtained $d = (d_{start}, ..., d_{end})$. Size of this vector d_{size} equals to amount of points of a contour within defect.

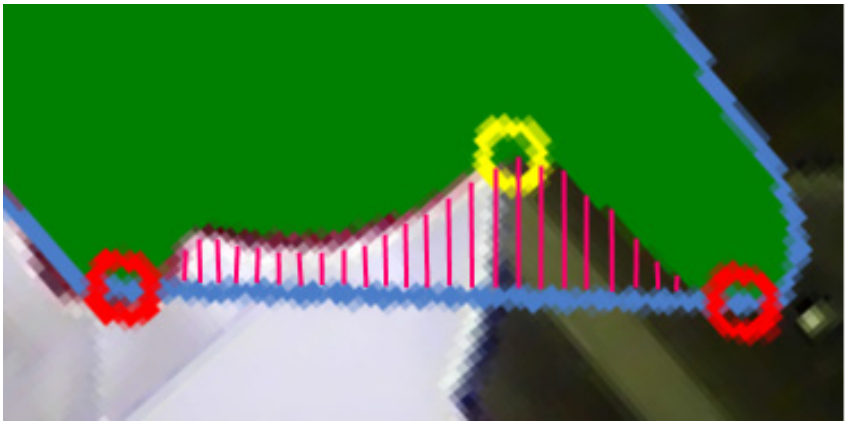


Figure 4. Generalized illustration of second step of the algorithm

3. Standardize vector d .

Third step of the algorithm should be covered more detailed. First, necessity in standardization appears because of different dimensions d_{size} of feature vectors. It is naturally that amount of points of a contour within defect differs for each demonstration of gesture. In order to solve this problem following algorithm of standardization of feature vector is suggested:

1. Set etalon dimension $d_{etalonSize}$ for feature vectors.
2. Compare dimension of current feature vector d_{size} with $d_{etalonSize}$. If $d_{size} \geq d_{etalonSize}$, move to step 3 a), else move to 3 b).
3. a) Divide dimension of current feature vector d_{size} by $d_{etalonSize}$. Round result to bigger number $amount = ceil(d_{size} / d_{etalonSize})$. Starting from the first element of vector d group elements in groups of size $amount$ (or less in case when there are not enough elements).

$$\begin{aligned}
&\text{set}_1 = \{d_{start}, \ldots, d_{start+amount-1}\},\\
&\text{set}_2 = \{d_{start+amount}, \ldots, d_{start+j*amount-1}\},\\
&\ldots\\
&\text{set}_j = \{d_{start+(j-1)*amount}, \ldots, d_{start+j*amount-1}\},\\
&\ldots\\
&\text{set}_{etalonSize} = \{d_{start+(d_{etalonSize}-1)*amount}, \ldots, d_{start+d_{etalonSize}*amount-1}\};
\end{aligned}$$

Find average value for elements in each group:

$$\begin{aligned}
&v_j = \frac{(d_{start+(j-1)*amount} + \ldots + d_{start+j*amount-1})}{amount},\\
&j = \overline{1, amount};
\end{aligned}$$

b) Increase quantity of feature vector elements on $d_{size} - 1$ by inserting additional elements between existing ones. Values for created elements are obtained by calculating average for two adjacent elements.

$$v_1 = d_{1,}$$
$$v_2 = \frac{(d_1 + d_2)}{2}, v_3 = d_2,$$
$$v_4 = \frac{(d_2 + d_3)}{2}, v_5 = d_3,$$
...
$$v_k = \frac{(d_{k/2} + d_{k/2+1})}{2}, v_{k+1} = d_{k/2+1},$$
...
$$v_{2d_{size}-2} = \frac{(d_{d_{size}-1} + d_{d_{size}})}{2}, v_{2d_{size}-1} = d_{d_{size}};$$
$$(\mathrm{k}/2+1) \le d_{size};$$

If $2d_{size} - 2 \ge d_{etalonSize}$ move to step 3 a), else repeat step 3 b).

4. Normalize feature vector by dividing all its elements by maximal element of the vector:

$$v_{\max} = \max_{i=1,d_{etalonSize}}(v_i),$$
$$v_i = \frac{v_i}{v_{\max}},$$
$$v = (v_1, v_2, ..., v_i, ..., v_{d_{etalonSize}}), v_i \in [0,1];$$

Normalization can solve problems that are connected with different distances between hand and web camera so this step is necessary.

Conclusion

Overall, standardized feature vector has etalon dimension $amount$ and its elements belong to interval from 0 to 1 (Figure 5). The feature vector can be used for finding similarity degree between defects. It is suggested to use Euclidian, ellipsoidal or orthogonal compliance distances with obtained feature vector [3].

Figure 5. Illustration of standardized feature vector

Bibliography

1. Donchenko V.S., Golik A.O. Usage of matrix feature vectors in recognition of tactile sing language. // SAIT 2013 System Analysys and Information Technologies, May 27-31, 2013, Kyiv, Ukraine.

2. Golik A.O. Recognition of tactile sign language: designing the feature vectors based on amount and structure of defects. XIX International Conference "Problems of Decision Making under Uncertainties"(PDMU -2012),– April 23-27, 2012.– Mukachevo, Ukraine, ст. 83.

3. Donchenko V.S., Golik A.O. Usage of ellipsoidal and orthogonal distances in speech and gesture recognition. // "Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics & Mathematics", №1, 2013.

4. Donchenko V.S., Golik A.O. Matrix feature vectors in speech and gesture recognition // International Journal "Information Theories and Applications", Vol. 20, Number 2, 2013, p.143-155.

5. Golik A.O. Designing the feature vectors: implementation of speech recognition algorithms using parallel computing capabilities. Visn., Ser. Fiz.-Mat. Nayky, Kyiv Univ. im. Tarasa Shevchenka. – 2012. – N 3. – P. 141-150. (in Ukrainian).

7. Donchenko V.S., Golik A.O. Using of parallel computing technology in speech recognition problems // XVII International Conference "Problems of Decision Making under Uncertainties"(PDMU -2011),– May 23-27, 2011.–Skhidnytsia, Ukraine.

8. .Donchenko V.S., Golik A.O. Implementation of algorithms of speech recognition using parallel computing capabilities // XVI International Conference "Problems of Decision Making under Uncertainties"(PDMU -2010),– October 4-8, 2010.– Yalta, Ukraine.

Authors' Information

Andrew Golik – *post-graduate student, Taras Shevchenko National University of Kyiv, Department of System Analysis and Decision Making Theory, e-mail: analytical@ukr.net*

Major Fields of Scientific Research: computer vision (especially speech and gesture recognition), artificial intelligence, pattern recognition, algorithms, web development.

FUZZY NEURAL NETWORKS FOR EVALUATING THE CREDITWORTHINESS OF THE BORROWERS

Natalia Shovgun

Abstract*: The problem of assessing the creditworthiness of the borrower is considered. The application of fuzzy neural networks for this problem solution, fuzzy neural networks TSK and Mamdani was suggested. The experimental investigations of application of these networks for our task were carried out and comparison with classical methods was performed. The modification of adaptation and learning algorithms of fuzzy neural networks was suggested.*

Keywords*: fuzzy neural networks, credit rating, fuzzy logic*

ACM Classification Keywords*: H.*4 Information systems applications; H.4.2. Types of Systems Decision Support

Introduction

The main activity of commercial bank is a credit activity. Lending provides almost half of bank profits; however, it is inextricably linked with risk. Credit risk is connected to possible misconduct by the borrower and it is one of the most significant risks of commercial banks. Consumer loans to individuals is a basic banking products. The bank assesses the creditworthiness of a potential borrower before lending. This is a method to minimize the losses of the bank. Careful selection of borrowers and effective assessment of creditworthiness is the main way of assessing and reducing credit risk. Information for decision making about lending may be inaccurate, incomplete, and information about the borrower may be such that it is difficult formalized. The analysis of existing methods of credit analysis showed the feasibility of using the methods based on fuzzy logic. These methods can work with both quantitative and qualitative characteristics and decision-making process is based on a comprehensible rules base.

For example, the technique of assessing the creditworthiness of individuals using the method of paired comparisons and fuzzy systems with Mamdani-type logical conclusion was considered in [Kuznetsov, 2007]. But the downside of fuzzy inference is that they can not learn automatically. Parameters and type of membership functions, which describe factors of creditworthiness, is given by an expert that's why it may be inadequate. Fuzzy neural networks (FNN) combine the advantages of fuzzy inference systems and neural networks - the ability to adapt and automatic learning, and the ability to interpret the process results.

To analyze the creditworthiness of borrowers applied fuzzy neural network with output Mamdani [Zaychenko, 2008] and fuzzy neural network with output Sugeno.

The work is devoted to the study of the FNN in the problem of assessing the creditworthiness of the borrower. A comparison of the results for the credit assessment by FNN TSK with classical methods such as logic model was performed and with popular in recent years Bayesian networks. Also provided are methods for setting up the rule base on which to base make a decision on lending.

Methods for assessing the creditworthiness of the borrower and their features

Based on questionnaires borrower the bank needs to decide - whether to grant credit. Each such form can be represented as a vector $\{X_1, X_2, ..., X_i, ..., X_M\}$, where X_i – some way the formalized data of borrower and the parameters of loan. This vector is the input of network. The decision on lending to the borrower is the output of network.

Most commonly used to solve this problem is using linear or logistic regression [Duffie, 2003]. Using linear regression we have function that determines the credit rating by approximating of a linear function wich argument is the vector characteristics of the borrower, i.e.:

$$p = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots a_N \cdot x_N,$$

where a_0 – the free term; a_i, $i = 1,..., N$ – weights of borrower characteristics; x_i - characteristics of the borrower.

All regression methods are sensitive to the correlation between the characteristics, so in the model should not be strongly correlated independent variables. In addition, the regression coefficients are not giving enough information about mechanism of influence characteristics of the borrower on the risk.

Bayesian networks (BN) are used in situations with some uncertainty. BN is a triple $N =< V, G, J >$, where V is a set of variables, G is a directed acyclic graph whose nodes correspond to random variables modeled process, J is a joint probability distribution of variables $V = \{X_1, X_2, ..., X_i, ..., X_N\}$. Bayesian networks, which may be presented with discrete and continuous variables is called hybrid Bayesian networks. Details about the BN can find at [Dawid, 2007].

In FNN results is obtained by using fuzzy logic, but the corresponding membership function is customized using learning algorithms FNN. Thus, the network uses a priori information for find new knowledge and it is logically transparent to the user.

We consider two different FNN. FNN with Mamdani-type fuzzy rules use next base of rules:

R_i: if x_1 is A_{i1} and x_2 is A_{i2} and … and x_n is A_{in} then y_i is C_i,

where x_i and y_i are input and output variables of the network, A_i and C_i are input and output fuzzy sets.

FNN with Sugeno-type fuzzy rules (FNN TSK) use next base of rules:

$$R_k: \text{if } x_1 \text{ is } A_1^{(k)};\ x_2 \text{ is } A_2^{(k)};\ ...;\ x_n \text{ is } A_N^{(k)}, \text{ then } y_i = p_{i0} + \sum_{j=1}^{N} p_{ij} x_j \ ,$$

where $A_i^{(k)}$ is fuzzy sets of variable x_i, i=1,2…N (data of borrower) for rule R_k with membership function

$$\mu_A^{(\kappa)}(x_i) = \frac{1}{1 + (\frac{x_i - c_i^{(k)}}{\sigma_i^{(k)}})^{2b_i^{(k)}}}.$$

For network training is used back-propagation algorithm. To find the parameters of membership function is used a gradient or genetic method. In the gradient method for configuring the parameters of membership function can be used resilient propagation method [Riedmille, 1992] to reduce the learning process. Each of the considered algorithms has its drawbacks. Thus, the gradient algorithm is highly dependent on the initial conditions, and genetic frequently converge to local optima. The author proposed to use a hybrid algorithm in which the initial

approximation (initial values of membership functions) is found by using genetic algorithm, and only then it is considered as the starting point for the gradient algorithm.

Scaling up the rules is the most efficient algorithm among the algorithms adapting FNN. The algorithm proposed in [Kruglov, 2002], based on an assessment of the accuracy of approximation. In this algorithm, we add a rule, if the existing knowledge base gives too large an error for the current point.

In [Juang, 1999] proposed an adaptation algorithm based on firing strength of rules, which is faster than algorithm based on approximation accuracy. In this algorithm, a new rule is added if the condition: $I = \arg\max_{1\le k\le r} F^k(x) \le F_{in}$, where $F^k(x(t)) = w_k$ is firing strength of rule k for input vector t, and F_{in} is a pre-specified threshold that decays during the learning process. In new rule the parameters of membership functions are set as follows: $c_j^{(k)} = x_j(t)$, $b_j^{(k)} = 1$, $\sigma_j^{(k)} = \beta \cdot \prod_{j=1}^{N} \frac{1}{1+(\frac{x_j - c_j^{(I)}}{\sigma_j^{(I)}})^{2b_j^{(I)}}}$. That is, we add a rule, if none of the existing rules do not describe well enough the current input vector. For the efficiency is proposed to make additional checks: if $f_o = \min(f_1, f_2, ..., f_m) > R$, where $f_k = \left\| \vec{c}^{(k)} - \vec{x}^i \right\|$, $k = 1, m$ is the distance between the centers of each membership function for each rules and the current point, $R = const$ then new rule is generated. Thus increasing the control of the number of rules. Accordingly the optimal network structure is building and thus the training requires less time.

Experimental results

The data sample of one of Ukrainian banks, which consists of 1,000 samples is used for credit analysis using the proposed methods. The feasibility of using FNN can be seen from Figure 1.

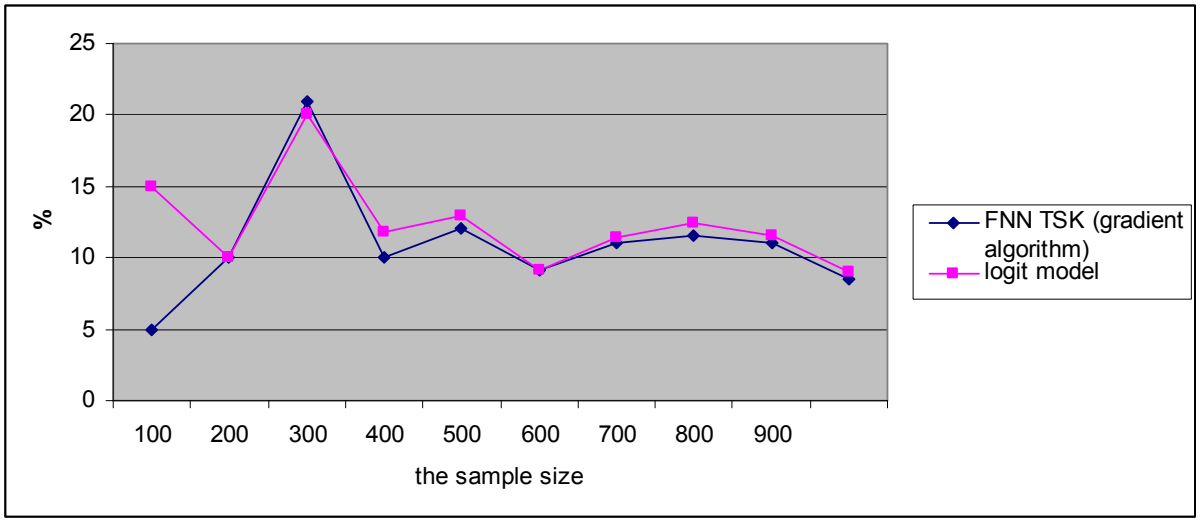


Figure 1. The errors dependence on volume of a sample

As you can see from the figure the percentage of incorrect classifications decreases with increasing training set. Errors are smaller when was using FNN TSK. To prevent re-training FNN TSK should be edit the complexity of network structure by adapting the algorithm parameters in accordance with the size of the training set. Building such an algorithm may be the subject of further research.

The data set consists of values: the borrower's age, sex, marital status, number of dependent, income, work experience, realty, monthly payment and reply. After the correlation analysis with using programs Netica (http://www.norsys.com/netica.html) was constructed the Bayesian network - Figure 1.

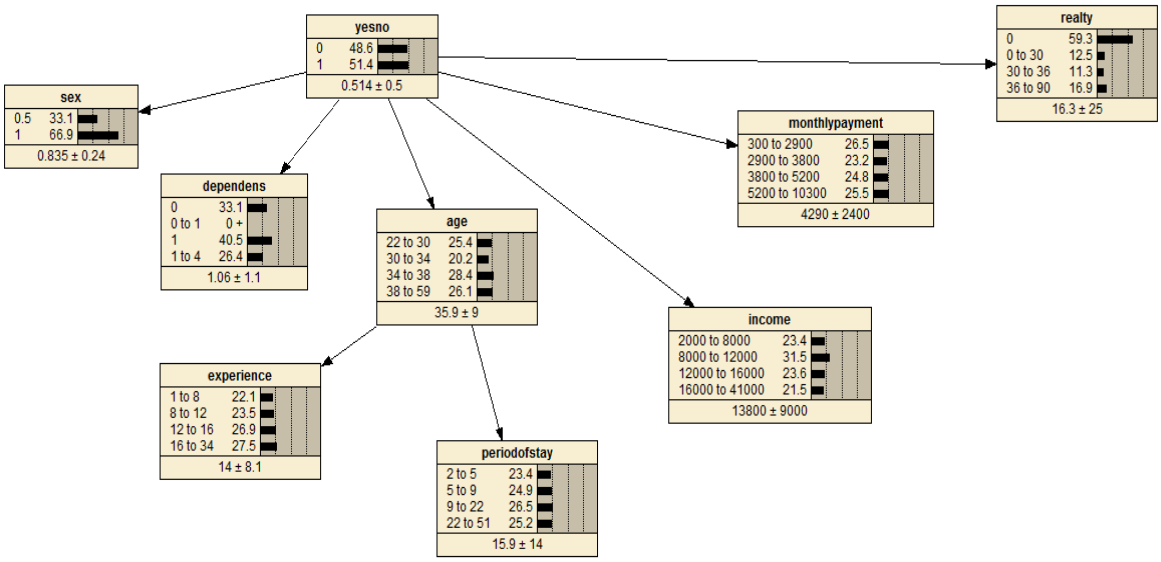


Figure 2. The structure of Bayesian network

The experemental results on figure 3.

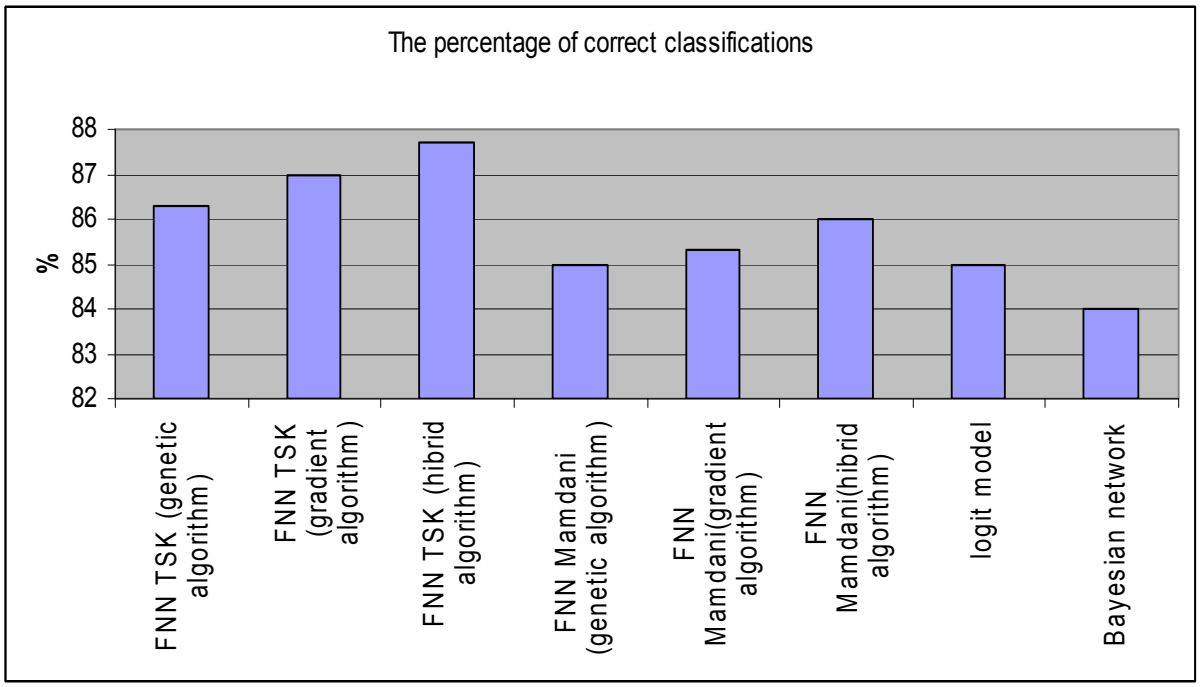


Figure 3. The results of the credit assessment of different methods

The best results has the FNN TSK (hibrid algorithm). Since the percentage of true classification isn't dependent on the type of membership function (Figure 4) we can speak of an automated construction of rules base on which a decision-making on lending is done. For example, if we trained the rule base with only two rule:

Rule 1: If the "Age" is "Low", "Sex" is "Female", "Married" is "No", "Number of dependents" is "High", "Income" is "Low", "Experience" is "High", "Residence time" is " High "," Monthly Payment " is "Low ", "Answer" is "No".

Rule 2: If the "Age" is "High", "Sex" is "Male", "Married" is "Yes", "Number of dependents" is "Low", "Income " is "High", "Experience" is "Low" , "Residence time" is "Low", "Monthly Payment" is "High", "Answe" is "Yes".

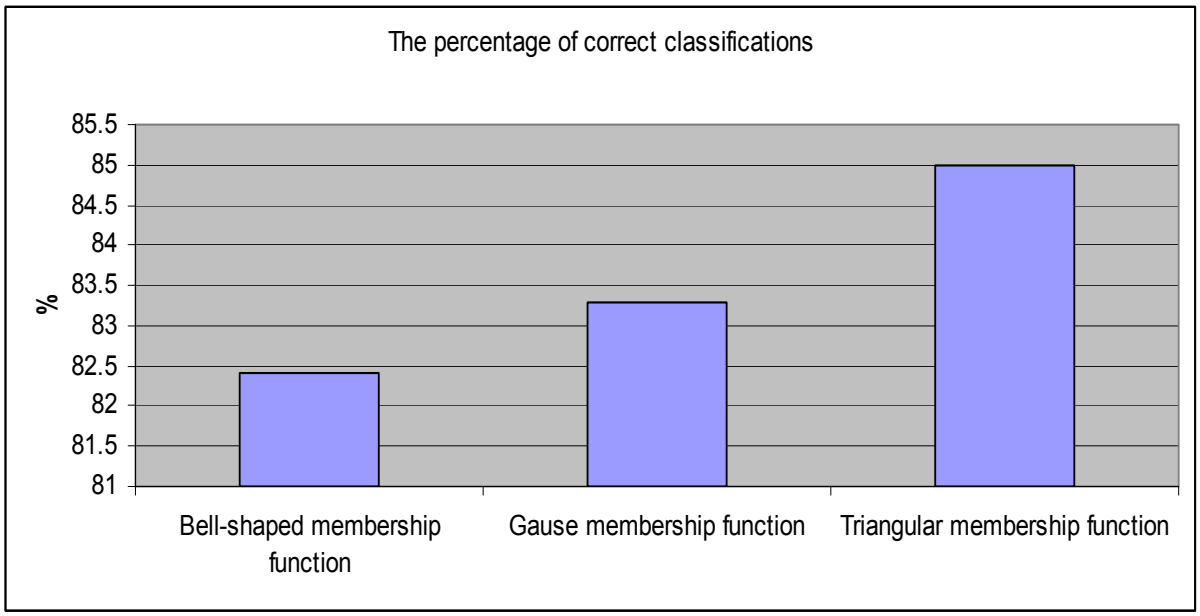


Figure 4. The results of the credit assessment for different membership function

We have a clear interpretation of the process of obtaining the decision making, credit institutions are given the opportunity to evaluate and adjust credit terms to offer the borrower an alternative parameters of lending.

Conclusion

The article considers the practical application of fuzzy neural networks to the problem of assessing the creditworthiness of the borrower. The results are compared with the classical method such as a logit model and some new as bayesian network. The best percentage of true classifications showed FNN TSK with hybrid (combination of gradient and genetic methods) learning algorithm. A new adaptation algorithm for fuzzy neural network was proposed, so we can build base rules automaticaly. As a result, there is an optimal network structure construction in accordance with the training set.

Acknowledgement

"The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua)."

Bibliography

[Dawid, 2007] Philip Dawid, Steffen L. Lauritzen, David J. Spiegelhalter Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks; Springer Science, 2007, 325 p.

[Duffie, 2003] Darrell Duffie, Kenneth J. Singleton Credit Risk: Pricing, Measurement, and Management, Princeton University Press, 2003, 396 p.

[HÁJEK,2008] Petr HÁJEK, Vladimír OLEJ MUNICIPAL CREDITWORTHINESS MODELLING BY NEURAL NETWORKS // Acta Electrotechnica et Informatica Vol. 8, No. 4, 2008, pp. 10 - 15

[Juang, 1999] CF Juang and CT Lin A recurrent self-organizing neural fuzzy inference network// IEEE transaction of neural networks, Volume 10, Number 4, pp. 828–845, July 1999

[Kruglov, 2002] Kruglov VV, Borisov VV, Artificial Neural Networks; Theory and Practice 2nd ed., Stereotype M: Hotline Telecom, 2002. 382 p.

[Kuznetsov, 2007] Kuznetsov L., Evaluation of credit history of individuals on the basis of fuzzy models // Managing large systems, 2007, № 21, pp. 84-106

[Riedmille, 1992] Riedmiller M., Brawn H. RPROP - a fast adaptive learning algorithms, Technacal Report // Karlsruhe: University Karlsruhe, 1992.

[Zaychenko Yu, 2008] Zaychenko Yu.P., Credit risk estimation on the base of fuzzy logic // System Research and Information Technologies, 2010, № 2, pp. 37–54

Author's Information

***Shovgun Natalia**, PhD student, The National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, email:* shovgun@gmail.com

Major Fields of Scientific Research: Neural network and fuzzy set

PROBLEM AND MATHEMATICAL MODELS FOR RESCUE TECHNICS ACQUISITION

Vitaliy Snytyuk, Pavlo Kucher

***Abstract**: In this paper the problem decision technology for resque technics acquisition with use multiobjective optimization, method of the variant's consecutive analysis and evolutionary modeling is considered. Models, serving information-analytical base of the integral objective forming, is suggested.*

***Keywords**: evolutionary modeling, objective function, rescue technics.*

***ACM Classification Keywords**: I.2 Artificial Intelligence, H.4 Information Systems Applications, J.6 Computer-aided Engineering*

Description of the problem-solving area

Actuality of the resque technics acquisition problem (RTAP) is defined by the increase dynamics of situations, in which necessary is its use, as well as increase of environments technogenic danger. In practice, the RTAP problem decision is taken by responsible person, coming from own experience. In consequence of this by performing the resque work often is an absence necessary toolbox in general, or impossibility of the problem performing fully.

In present time significantly extended assortment fire-prevention and resque product, taken off restrictions on import foreign technics, but exists the certain deficit of financial resource. It is impossible also not to notice of wide functionality and maximum power necessity.

Obviously that problem of the RTAP has much common aspects with the known problem of the bin packing [Lodi, 2002]. The bin packing problem is concluded in accomodation object predestined form by such way that number used container was most or volume object was least. In problem of the RTAP objective function of bin packing problem changes in restrictions on overall dimensions elements. The objective functions are functionality, power, cost, other features of RTA elements. So, priority problem is a forming integral objective function and presentations of the potential decisions of the problem. Aspects of its solving are offered below.

Problem of the resque technics acquisition

Let the sets $X=\{X_1,X_2,...,X_n\}$ presents the assortment of the resque technics. Each element from X belongs to one of the set classes $C=\{C_1,C_2,...,C_k\}$, where $k<<n$. Assume, that in complete set must enter equipment from each of $\{C_1,C_2,...,C_m\}$ classes, $m<k$, i.e. $\{X_{i_1}^1,X_{i_2}^1,...,X_{i_{j_1}}^1\}\subset C_1,...,\{X_{i_1}^m,X_{i_2}^m,...,X_{i_{j_m}}^m\}\subset C_m$. each element of X will to correspond with value set:

$$X_q \to < F_{1_q},F_{2_q},F_{3_q},a_q,b_q,c_q >, \tag{1}$$

where F_{1_q} − functionality value for q element; F_{2_q} − its power value; F_{3_q} − its cost; a_q,b_q,c_q − its overall dimensions, $q=\overline{1,n}$.

We shall do the simplifying remarks. Let all elements have a form of the right-angled parallelepiped and they must be placed in right-angled bit. Besides, in the bit must be one element from each class.

The RTAP is reduced to multiobjective optimization problem

$$F_1(x) \to \max, F_2(x) \to \max, F_3(x) \to \min, \tag{2}$$

where $x = (x^1_{i_1}, x^2_{i_2}, ..., x^m_{i_m}), x^j_{i_j} \in C_j$ by restriction

$$F_1(x^j_{i_j}) > 0, F_2(x^j_{i_j}) > 0, 0 < F_3(x^j_{i_j}) < F_3^{j_{\max}}, \tag{3}$$

$$0 < a_q(x^j_{i_j}) < \max\{a,b,c\},\ 0 < b_q(x^j_{i_j}) < \max\{a,b,c\},\ 0 < c_q(x^j_{i_j}) < \max\{a,b,c\}, \tag{4}$$

where a,b,c − bit's overall dimensions.

It is known that such problems refer to NP-hard problems. But, obviously that in problem (2)-(4) can be made suggestions, simplifying process of its solving. We consider rational to use the ideas of the multiobjective optimization problems decision [Chernoruzkiy, 2005], [Voloshin, 2006], method of the variant's consecutive analysis [Volkowitch, 1993] and evolutionary modeling [Michalewicz, 1996].

Information-analytical models of complex systems

As the basis of the efficient problem (2)-(4) solving lays such preconditions:

1. Forming a models set, which will allow realizing the objective function identification.

2. The development integral objectives function, which values reception will allow installing the preferences on the variants set.

We shall consider the problem of the forming the models set, which work out an information-analytical research basis. It is known that by the complex systems construction traditionally [Timchenko, 1991] use the models of the construction, operation and development.

In our case the construction model is such:

$$M_s < X_1, X_2, ..., X_n >, \tag{5}$$

where n − the number of RTA elements. The construction model is a basis, which is intended for forming an element s set and structures by RTA acquisition.

The operation model

$$M_f = < G_1, G_2, ..., G_n >, \tag{6}$$

where $G_i, i = \overline{1,n}$, − transformations, which is realized by i element, and $Y_i = G_i(I_i, R_i, P_i), Y_i$ − same feature, which is defined by transformation G_i and pointing to its result, I_i − a priori information about RTA types, their scale and possible consequence, R_i − material and energy facility required for operating the element X_i and receptions values Y_i, P_i − features of transformation process $< I_i, R_i > \to Y_i, i = \overline{1,n}$.

The third model − development model will present, using belonging elements to classes

$$M_d = < (X^1_{i_1}, X^1_{i_2}, ..., X^1_{i_{j_1}}), ..., (X^m_{i_1}, X^m_{i_2}, ..., X^m_{i_{j_m}}) > \tag{7}$$

where m is the number of RTA elements classes, which execute like functions. Elements from each subset can be ranked on functionality, power and cost levels. Possible also are variants of the overall dimensions order.

The offered models form the basis for receipt of objective function, which will used by decision making for choice of the RTA completing optimum variant in conditions of resources deficit.

Construction features of integral objective function

The RTA problem has a features, to which concern multiobjectivity, different dimension objective functions values, weak structuring. We shall consider aspects of the integral objective function, coming from the known methods of solving of the multiobjective optimization problems [Larichev, 2003]. Notice that objective functions (2) can be both constant and analytical dependences.

1. Main objective function method. Assume main objective function to be a cost of the RTP element. Then problem (2)-(4) is converted to such type:

$$F_3(x) \to \min,\ x = (x^1_{i_{l_1}}, x^2_{i_{l_2}}, .., x^m_{i_{l_m}}),\ x^m_{i_{l_j}} \in C_j, \tag{8}$$

$$x \in D,\ D = \{x / F_{i\min} < F_i(x),\ i = \overline{1,2}\} \tag{9}$$

and (4) is executed. In the problem (8) - (9), $F_{j\min}, i = \overline{1,2}$, – minimum possible values i^{th} odjective function. In that way we get the multiobjective optimization problem. Its solving in case of known values F_1, F_2, F_3 for all elements is reduced to searching

$$x^*_1 = \max_{x \in D'} F_3(x), \tag{10}$$

where D' is the area, in which are executed restrictions (3) and (4). If $x^*_1 \in D$, then solution is found, if no – do search

$$x^*_2 = \max_{\substack{x \in D' \\ x \neq x^*_1}} F_3(x)\ . \tag{11}$$

If $\exists x^*_i :\ x^*_i = \max_{x \in D'} F_3(x),\ x^*_i \in D$, then problem has a solution, otherwise the solution is absent.

2. Method of linear convolution. The necessary conditions to realization of the method are:

- Normalization of objective functions values;
- Determination weight coefficients of objective functions.

Then integral objective function and problem will be such:

$$F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x) - \alpha_3 F_3(x) \to \max, \tag{12}$$

where $\alpha_i > 0, i = \overline{1,3}, \sum_{i=1}^{3} \alpha_i = 1$. If the objective functions values and integral objective function value on the elements (from RTA) of control set are known, then coefficients $\alpha_1, \alpha_2, \alpha_3$ can be calculated, for instance, on least squares method. However, this is not always possible, more so that most likely in array of initial data will exist a multicollinearity factor and result will be biased. At other times necessary to use a processing techniques for expert estimations.

3. Method of ideal point. The point (x^*_1, x^*_2, x^*_3) is ideal, if $x^*_i = \max_{x \in D} F_i(x), i = \overline{1,3}$. The solution of oneobjective optimization problems - ideal point will be founded. Then the solving process is concluded in searching for of such point:

$$x^* = Arg \min_{x \in D} \Big(\sum_{i=1}^{3} (F_i(x) - x^*_i)^2 \Big)^{\frac{1}{2}}. \tag{13}$$

Objective functions values must be normalized and if objective functions have a weight coefficients then problem (13) will be such:

$$x^* = Arg\min_{x\in D}(\sum_{i=1}^{3}\alpha_i(F_i(x)-x_i^*)^2)^{\frac{1}{2}}, \quad (14)$$

where $\alpha_i > 0, i = \overline{1,3}, \sum_{i=1}^{3}\alpha_i = 1.$

Exist and other methods for decision of the multiobjective optimization problems such as choice on number of dominant objective function, method of the consequtive concessions, consequtive entering the restrictions and etc, but all of these require attraction to additional information, which can be not. Therefore for our problem solving we stopped on afore-cited three methods.

Preliminary steps for shortening variants number of solving problem

1. Removing possible variant of the problem solving, which strictly dominated at least one of other variant. We shall notice that such operation can be executed at the beginning to initially realization of searching of the problem solving, if the power of variants set is relatively small. If this is not so, that checking for dominating is realized in process of the problem solving for each element separately.

2. Necessary to realize preliminary check, does not exist such element RTA that

$$(a_q > \max\{a,b,c\}) \vee (b_q > \max\{a,b,c\}) \vee (c_q > \max\{a,b,c\}) \quad (15)$$

does not exist such RTP elements set that

$$(\sum_{q=1}^{3}a_q > \max\{a,b,c\}) \vee (\sum_{q=1}^{3}b_q > \max\{a,b,c\}) \vee (\sum_{q=1}^{3}c_q > \max\{a,b,c\}). \quad (16)$$

If elements or elements sets satisfying (15) or (16) accordingly exist that their necessary to delete a priori, or in process of the problem solving. Similarly, using scheme of the consequtive analysis variant, we delete the variants, the total functionality or power which less minimum possible, as well as that, which the cost exceeds the possible value.

Main directions of problem solving

Since is necessary to find the function optimum, given tabular, under specified restrictions, and about characteristic which nothing not known, then we introduce rational using evolutionary modeling. The choice of the evolutionary modeling method is a researcher prerogative.

Assume that we use the genetic algorithm [Holland, 1994]. It is known that its realization accompanies two problems: forming of objective function and presentation of the potential solutions as binary chromosomes. In our problem objective function is already received. For forming chromosomes-solutions we shall offer such approach. Since solution is a set with m elements, then length of the chromosome will be m. Each its position corresponds to one RTP element. All elements of the chromosome belong to one class.

Each element has 3 fragments. The first fragment corresponds to functionality value, the second – to a power, but the third − to a cost. Thereby, chromosome-solution will have $3m$ fragments. On initial stage all features element values were normalized, their values are found in $[0,1]$. Further all known procedures of the genetic algorithm are used. We shall neither notice that got solution can not correspond to nor one potential variant. Then necessary to find nearest to it on criterion of the minimum middle square distances. Genetic algorithm application is preferably, when known a particular objective functions values. For solving of the problem also rational is an using evolutionary strategy [Rechenberg, 1994].

Conclusion

The considered problem of resque technics acquisition is a complex multiobjective problem. Its complexity depends on quality RTA elements and carriers, to which they will, are installed. The new samples of technics, their evolution point to optimality of RTA problem solving. Technology, which is offered in this paper, is based on element of three components: multiobjective optimization, consequtive variant analysis and evolutionary modeling and unites their advantage in itself.

Bibliography

[Chernoruzkiy, 2005] И.Г. Черноруцкий. Методы принятия решений. − СПб.: BHV, 2005. − 416 с.

[Holland, 1994] J.H. Holland. Adaptation in natural and artificial systems. An introductory analysis with application to biology, control and artificial intelligence. – London: Bradford book edition, 1994. – 211 p.

[Larichev, 2003] О.И. Ларичев. Теория и методы принятия решений. − М.: Логос, 2003.

[Lodi, 2002] A. Lodi, S. Martello, D.Vigo. Recent advances on two-dimensional bin packing problems. Discrete Appl. Math., 2002. − Vol. 123. − P. 379-396.

[Michalewicz, 1996] Z. Michalewitcz. Genetic Algorithms+Data Structures=Evolution Programs. − Springer Verlag, Berlin, Heidelberg, New-York, 1996. − 387 p.

[Rechenberg, 1994] I. Rechenberg. Evolutionsstrategie "94". – Stuttgart-Bad GannStatt: Frommann Halzboog. – 1994. − 434 p.

[Timchenko, 1991] А.А. Тимченко, А.А. Родионов. Информатика системного проектирования объектов новой техники. − К.:, Наукова думка, 1991. − 231 с.

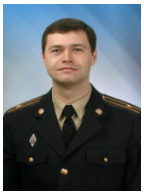
[Volkowitch, 1993] Модели и методы оптимизации надежности сложных систем // В.Л. Волкович, А.Ф. Волошин,…− К.:, Наукова думка, 1993. − 312 с.

[Voloshin, 2006] А.Ф. Волошин, С.О. Мащенко. Теория принятия решений. − К.: Киевский университет, 2006. − 304 с.

Authors' Information

Vitaliy Snytyuk *– Professor, Head of Department, Cherkassy State Technological University, Shevchenko Ave. 460, Cherkassy, Ukraine, 18006; e-mail: snytyuk@gmail.com*

Major Fields of Scientific Research: Decision making processes under uncertainty, Evolutionary modeling



Pavlo Kucher*– Senior lecturer, Academy of Fire Safety named after Chernobyl Heroes, Onoprienko Str., 8, Cherkassy, Ukraine, 18034; e-mail: kucherpavel@bigmir.net*

Major Fields of Scientific Research: Decision making processes under uncertainty, Evolutionary modeling

MODEL FOR ASTRONOMICAL DATING OF THE *CHRONICLE* OF HYDATIUS: RESULTS FOR THE INTERVAL (600-1000)

Jordan Tabov

Abstract: *This article presents the details and the results of the application of a 'soft' model for astronomical dating of the seven eclipses mentioned in the Chronicle of Hydatius, for the time interval (600, 1000), i.e. since 600 AD till 1000 AD.*

The analysis of the date shows, that in the interval (600, 1000) the following two septets (i.e. groups of seven) of eclipses are the best candidates for dating of the eclipses mentioned in the Chronicle of Hydatius:

I. Septet: 923-Nov-11, 939-Jul-19, 968-Dec-22, 972-Sep-25, 978-Jun-08, 983-Mar-01, 985-Jul-20.

II. Septet: 923-Nov-11, 939-Jul-19, 968-Dec-22, 991-Sep-26, 978-Jun-08, 983-Mar-01, 985-Jul-20.

Keywords*: Chronicle of Hydatius, astronomical dating, eclipses, soft model, fuzzy information.*

ACM Classification Keywords: *I.5 PATTERN RECOGNITION; I.5.1 Models*

1. Introduction

For the eclipses in the past we have two main sources of data: historical chronicles and astronomical tables of the dates of the past eclipses. The dating of a certain eclipse, mentioned in the historical chronicles, is in fact identifying it with a certain eclipse from the astronomical tables, which should be such that the parameters of the eclipse (place of the observation, day, month, hour, phase) mentioned in the chronicle coincide with the parameters of the eclipse from the tables. However often the dating of historical eclipses is problematic.

One of the most famous chronicles, containing information about historical eclipses (and sometimes data about their basic parameters), is that of Hydatius ([Idatii, 1619], [Idatii, 1634], [Idatii, 1861], [Hydatii, 1894]), in which seven eclipses are mentioned: five solar and two lunar.

A soft model for astronomical dating of these seven eclipses is suggested in the paper [Tabov, 2013]; details and results of its application for the time interval (300, 600) are presented in the paper [Tabov & Umlenski, to appear]. Here we give the similar details and results for the time interval (600, 1000).

In the framework of this model the information about the eclipses is systematized in two main parts of the model:

- "Template" ("image" of the initial described by the author "septet" of eclipses), including: 1) date (day and month) and day of the week – for every one of the eclipses – and 2) intervals (in years) between the eclipses, according to the text of the Chronicle, and
- "Distance".

What does mean in this case the term "soft model"?

It is natural to expect, that some of the data in the *Chronicle* may be incorrect. Therefore we consider the Template ***not as an exact image*** of the initial septet of eclipses described by the author, but as a fuzzy image of those seven real eclipses described by Hydatius. We assume that the inaccuracy of the data is small, i.e. that the parameters (dates, days of the week, etc.) in the Template do not differ much from the corresponding parameters of the initial (real) septet of eclipses, but are in some sense "close" to them.

For a more precise definition of this "closeness" in the paper [Tabov, 2013] is proposed a formula for the "distance" from the Template to any septet of real (happened in the past) eclipses. It allows us to calculate and compare the distances from the Template to all septets of real eclipses and to choose several "closest" to the Template (at the shortest distance to it) septets of real eclipses.

These septets are the target of the procedure of dating in the proposed model; they should be subject to further analysis and individual comparisons to determine which one of them is the most likely prototype of the Template, and thus its most probable dating.

2. Template of the eclipses in the Chronicle of Hydatius

Template

(This Template is built up from the parameters of the seven eclipses **H1, H2, H3, H4, H5, H6** and **H**7, described by the text in the chronicle of Hydatius, with notation and representation according to [Tabov, 2013]).

H1 Solar eclipse on November 11, Monday.

H2 Solar eclipse on July 19, Thursday.

H3 Solar eclipse on December 24, Tuesday.

H4 Lunar eclipse on September 27. It was seen in the East (Eastern parts of the Empire) and was not seen in the West.

H5 Solar eclipse on June 9, Wednesday. Time – from the 4th hour till the 6th hour. Phase – about 0.4 – 0.5 (like 5- or 6- day moon).

H6 Lunar eclipse on March 2, Friday.

H7 Solar eclipse on July 20, Monday. Time – from the 3^{d} hour till the 6th hour. Phase – about 0.4 (like 5-day moon).

Here the eclipses **H1, H2, H3, H4, H5, H6** and **H**7 are visible in the region of Mediterranean (Jerusalem, Constantinople and Caves).

The Template includes also the interval between the eclipses. Let t_i be the length of the interval (in years) between H_i and H_{i+1}. It is clear from the text of the *Chronicle*, that $t_1 = 16$, $t_3 = 5$ and $t_6 = 1$, and that the approximate values of t_2, t_4 and t_5 are **29, 7** and 5 years, respectively.

In the Chronicle there are some additional calendar- and astronomical data, which also should be included in the Template: 1) According to the *Chronicle*, the day of Easter in the year of $\mathbf{E_5}$ was on March 28; 2) The eclipse $\mathbf{E_4}$ was seen only in the Eastern parts of the Roman Empire and was not seen in the Western parts.

3. The problem for dating the eclipses in the Chronicle

Our target is: to determine several most appropriate septets of eclipses from the astronomical tables of the eclipses from AD 600 till AD 1000, from which after additional analyses should be selected the best candidate for identification with the real eclipses mentioned in the *Chronicle*. These septets should be at a least possible "distance" from the Template; the "distances" from the Template can be calculated by means of the following rules, suggested in [Tabov, 2013] (they will be used further).

Let $\mathbf{G_E = \{E_1, E_2, \ldots, E_7\}}$ be a set of seven (or a septet of) eclipses, which occurred in the past.

How much, or at what extent this septet differs from the set of eclipses $\mathbf{G_H = \{H1, H2, \ldots, H7\}}$ – differs in the astronomical parameters, described above in the Template?

For searching the answer of this question we use the "metric", suggested in [Tabov, 2013], which "models" the "closeness" of particular eclipses or of a group of eclipses respectively to **H1, H2, ..., H7** и $\mathbf{G_H}$. This metric is important for the application of the Template described above for dating ancient eclipses and especially for the "measuring" the "closeness" of $\mathbf{G_E}$ to $\mathbf{G_H}$.

Let **E** be a certain eclipse from the List of the eclipses in the past.

Recall the rules for "scores" for the "closeness" of **E** respectively to each one of the eclipses **H1 - H7,** as they given in [Tabov, 2013] with the minor changes suggested in [Tabov & Umlenski, to appear].

Let ***m*** be a fixed positive number.

Scores for evaluation of the closeness of E to the eclipse H1:

The total score $\mathbf{e_1}$ for the closeness of an eclipse **E** to the eclipse **H1** is the sum of the scores for **1-1** and **1-2**:

1-1 The date of the eclipse H1 is November 11th. If the date of E is:

- November 11 => score for **1-1:** ***m*** points;
- November 10 or 12 => score for **1-1:** 0.9***m*** points;
- November 9 or 13=> score for **1-1:** 0.5***m*** points;
- Another day => score for **1-1:** 0 points.

1-2 If the score for **1-1** is 0 points, the score for **1-2** is also 0 points; if the score for **1-1** is different from 0, the score for **1-2** is determined by the following rules. Taking into account, that the day of the week on which the eclipse **H1** occurred is Monday, if the day of the week on which occurred **E** is:

- Monday => score for **1-2:** 0.5***m*** points;
- Tuesday or Sunday => score for **1-2:** 0.4***m*** points;
- Wednesday or Saturday => score for **1-2:** 0.3***m*** points;
- Another day => score for **1-2:** 0 points.

The rules for the determination of the scores for the closeness to the other four solar eclipses - **H2, H3, H5** and **H7** – are omitted, because they are completely analogous to that for the case of **H1**; different are only the dates and the corresponding days of the week.

The rules for the determination of the scores for the closeness of a lunar eclipse **E** to the lunar eclipses **H4** and **H6** are different:

Scores for evaluation of the closeness of E to the eclipse H4:[1]

The total score $\mathbf{e_4}$ for the closeness of a lunar eclipse **E** to the lunar eclipse **H4** is the sum of the scores for **4-1, 4-2** and **4-3**:

4-1 The date of the eclipse **H4** is September 27. If the date of **E** is:

- September 27 => score for **4-1:** ***m*** points;
- September 26 or 28 => score for **4-1:** 0.9***m*** points;
- September 25 or 29 => score for **4-1**: 0.5***m*** points;
- Another day => score for **4-1**: 0 points.

[1] These rules differ insignificantly from the version in [Tabov, 2013], but are identical with the version in [Tabov & Umlenski, to appear].

4-2 If ***E*** was seen in the Eastern part of the Empire (Jerusalem),

the score for **4-2** is 0.5***m*** points, otherwise 0 points.

If the score for **4-1** is 0 points, the scores for **4-3** and **4-2** are also 0 points.

4-3 If ***E*** was not seen in the Western part of the Empire,

the score for **4-3** is 1.5***m*** points, otherwise 0 points.

Scores for evaluation of the closeness of E to the eclipse H6:

The total score $\mathbf{e_6}$ for the closeness of a lunar eclipse **E** to the lunar eclipse **H6** is the sum of the scores for **6-1** and **6-2**:

6-1 The date of the eclipse **H6** is March 2. If the date of E is:

- March 2 => score for **6-1:** ***m*** points;
- March 1 or 3 => score for **4-1:** 0.9***m*** points;
- February 28/29 or March 4 => score for **6-1:** 0.5***m*** points;
- Another day => score for **6-1:** 0 points.

If the score for **6-1** is 0 points, the score for **6-2** is also 0 points; if the score for **6-1** is different from 0, the score for **6-2** is determined by the following rules. Taking into account, that the day of the week on which the eclipse **H6** occurred is Friday, if the day of the week on which occurred **E** is:

- Friday => score for **6-2:** 0.5***m*** points;
- Wednesday or Friday => score for **6-2:** 0.4***m*** points;
- Tuesday or Saturday => score for **6-2:** 0.3***m*** points;
- Another day => score for **6-2:** 0 points.

Scores for evaluation of the lengths of time intervals between the eclipses $\mathbf{E_1, E_2, \ldots, E_7}$

Let ***n*** be a fixed positive number.

The intervals are in years and are equal to the differences between the years (in the Julian calendar) in which the respective eclipses occurred.

Denote by $\boldsymbol{f_i}$ the score for the closeness of the interval between $\boldsymbol{E_i}$ and $\boldsymbol{E_{i+1}}$ to the interval between $\boldsymbol{H_i}$ and $\boldsymbol{H_{i+1}}$.

If the interval between $\mathbf{E_1}$ and $\mathbf{E_2}$ is:

16 years => $\boldsymbol{f_1} = \boldsymbol{n}$ points, 15 or 17 years => $\boldsymbol{f_1} = 0{,}9\boldsymbol{n}$ points, 14 or 18 years => $\boldsymbol{f_1} = 0{,}4\boldsymbol{n}$ points, in other cases $\boldsymbol{f_1} = 0$ points.

The rules for calculation of the scores $\boldsymbol{f_3}$ and $\boldsymbol{f_6}$ are similar.

If the interval between $\mathbf{E_2}$ and $\mathbf{E_3}$ is:

29 years => $\boldsymbol{f_1} = 0.2\boldsymbol{n}$ points, 28 or 30 years => $\boldsymbol{f_1} = 0{,}1\boldsymbol{n}$ points, in the other cases $\boldsymbol{f_1} = 0$ points.

The rules for calculation of the scores $\boldsymbol{f_4}$ and $\boldsymbol{f_5}$ are similar.

The uncertain length of the intervals between the successive eclipses $\mathbf{H_2}$ and $\mathbf{H_3}$, $\mathbf{H_4}$ and $\mathbf{H_5}$, and $\mathbf{H_5}$ and $\mathbf{H_6}$ create additional difficulties for adequate evaluation of the "closeness" of $\mathbf{G_E}$ to $\mathbf{G_H}$. More significant deviations of the interval between $\mathbf{E_i}$ and $\mathbf{E_{i+1}}$ from the interval between $\mathbf{H_i}$ and $\mathbf{H_{i+1}}$ in more than two cases should be subject of a special attention.

Scores for Easter in the year of $\mathbf{E_5}$

Let **p** be a fixed positive number; by g denote the score for Easter in the year of $\mathbf{E_5}$.

If in the year of $\mathbf{E_5}$ Easter was on

- March 28 => $\boldsymbol{g = p}$;
- March 27 or 29 => $\boldsymbol{g} = 0{,}9\boldsymbol{p}$;
- March 26 or 30 => $\boldsymbol{g} = 0{,}5\boldsymbol{p}$;
- Another day => $\boldsymbol{g} = 0$.

4. Closeness of a set GE of 7 eclipses to the septet GH

Let $\mathbf{G_E = \{E_1, E_2, \ldots, E_7\}}$ be a set of seven eclipses. We define a "distance" of $\mathbf{G_E}$ to the septet $\mathbf{G_H = \{H1, H2, \ldots, H7\}}$ ("of Hydatius") by the astronomical parameters described above for the group $\mathbf{G_H}$ and according to the rules for giving scores given above.

We define the "distance" $\boldsymbol{d}$ from $\boldsymbol{G_E}$ to $\boldsymbol{G_H}$ in the following way:

$$\boldsymbol{d} = 12\,\boldsymbol{m} + 3{,}6\,\boldsymbol{n} + \boldsymbol{p} - (\boldsymbol{e_1} + \boldsymbol{e_2} + \ldots + \boldsymbol{e_7} + \boldsymbol{f_1} + \boldsymbol{f_2} + \ldots + \boldsymbol{f_6} + \boldsymbol{g}).$$

It is easy to check that in case of coincidence of the respective parameters for the eclipses of $\mathbf{G_E}$ and $\mathbf{G_H}$ the distance $\boldsymbol{d}$ is equal to 0. The less is $\boldsymbol{d}$, the "closer" is the septet $\mathbf{G_E}$ to $\mathbf{G_H}$.

5. Searching for the closest (to GH) septet of eclipses GE

The proposed formula for calculation of the distance from $\mathbf{G_E}$ to $\mathbf{G_H}$ is an essential part of our model for astronomical dating. It is natural to combine it with different methods for determining the closest to $\mathbf{G_H}$ "septets" of eclipses in a given historical period – for example, in the time interval from AD 600 to AD 1000.

A brief description of a possible approach how to "search" for suitable "septets" at shortest distance from $\mathbf{G_H}$ and its application are given below.

The first step is the reduction of the list of all eclipses of the period (assuming that this is the interval from AD 600 to AD 1000) to its part **L,** containing only the eclipses visible from the Mediterranean region (Jerusalem, Constantinople, Caves).

From this reduced list L we select seven sets of eclipses $\mathbf{G^1}$, $\mathbf{G^2}$, …, $\mathbf{G^7}$*:* the set $\mathbf{G^1}$ contains only those eclipses of **L**, whose date is "around the date of **H1**", i.e. about November 11, and more precisely, in the framework of the proposed Template, on the days from 9 to 13 November inclusive. Similarly, we select the other sets $\mathbf{G^2}$, $\mathbf{G^3}$, …, $\mathbf{G^7}$. The result is represented in **Table 1**.

Set	O Description of the eclipses in the set
Set $\mathbf{G^1}$ (of **H1**)	**H1** Solar eclipse on November 11, Monday. The set $\mathbf{G^1}$ contains the solar eclipses from **L** which occurred on November 09, 10, 11, 12 and 13. For these eclipses we also say that they are of type **H1.**
Set $\mathbf{G^2}$ (of **H2**)	**H2** Solar eclipse on July 19, Thursday. The set $\mathbf{G^2}$ contains the solar eclipses from **L** which occurred on July 17, 18, 19, 20 and 21. For these eclipses we also say that they are of type **H2.**
Set $\mathbf{G^3}$ (of **H3**)	**H3** Solar eclipse on December 24, Tuesday. The set $\mathbf{G^3}$ contains the solar eclipses from **L** which occurred on December 22, 23, 24, 25 and 26. For these eclipses we also say that they are of type **H3.**
Set $\mathbf{G^4}$ (of **H4**)	**H4** Lunar eclipse on September 27. It was seen in the East (Eastern parts of the Empire). The set $\mathbf{G^4}$ contains the lunar eclipses from **L** which occurred on September 25, 26, 27, 28 and 29, seen in the East (in Jerusalem). For these eclipses we also say that they are of type **H4.**

Set $\mathbf{G^5}$ (of **H5**)	**H5** Solar eclipse on June 9, Wednesday. The set $\mathbf{G^5}$ contains the solar eclipses from **L** which occurred on December 07, 08, 09, 10 and 11. For these eclipses we also say that they are of type **H5.**
Set $\mathbf{G^6}$ (of **H6**)	**H6** Lunar eclipse on March 2, Friday. The set $\mathbf{G^6}$ contains the lunar eclipses from **L** which occurred on February 28 and 29 and on March 01, 02, 03, and 04. For these eclipses we also say that they are of type **H6.**
Set $\mathbf{G^7}$ (of **H7**)	**H7** Solar eclipse on July 20, Monday. The set $\mathbf{G^7}$ contains the solar eclipses from **L** which occurred on July 18, 19, 20, 21 and 22. For these eclipses we also say that they are of type **H7.**

Table 1. The sets $\mathbf{G^1}$, $\mathbf{G^2}$, …, $\mathbf{G^7}$

In order to analyse the number and the distribution of the solar eclipses of the types **H1**, **H2**, **H3**, **H5** and **H7** in the interval AD 600-1000, we arrange them in chronological order in **Table 2**:

Type H1	Type H2	Type H3	Type H5	Type H7
		604-Dec-26		
			606-Jun-11	
			625-Jun-10	
	836-Jul-17			
			894-Jun-07	
			913-Jun-07	
	920-Jul-18			**920-Jul-18**
923-Nov-11				
	939-Jul-19			**939-Jul-19**
942-Nov-11				
		949-Dec-22		
	966-Jul-20			**966-Jul-20**
		968-Dec-22		
			978-Jun-08	
	985-Jul-20			**985-Jul-20**

Table 2. The eclipses of the types **H1** – **H3, H5** and **H7** in the interval AD 600 – 1000

From the septets of eclipses we are interested in (7 eclipses, one of each type **H1** – **H7)** we should choose several with highest scores according to the scheme of the model. Every such septet should be in a certain interval of length 100 years. Hence, in order to have in a certain interval of length 100 a septet with a high score, in this interval should present eclipses of at least five of the types **H1** – **H7.** Then among them there should be at least 3 solar ones.

From **Table 2** it is clear that:

In the interval 600-800 there are solar eclipses of two types: **H3** and **H5,** and there is no solar eclipse of the types **H1**, **H2**, and **H7;**

In the interval 800-800 there is solar eclipses of two types: **H2** and **H5,** and there is no solar eclipse of the types **H1**, **H2**, and **H7.**

Consequently in order to select septets with high scores, it is sufficient to consider only the eclipses from the interval AD 890-1000.

From the list **L** we find successively ***in the interval AD 890-1000***:

Eclipses from the set $\mathbf{G^1}$ (of type **H1**): 923-Nov-11, 942-Nov-11.

Eclipses from the set $\mathbf{G^2}$ (of type **H2**): 939-Jul-19, 966-Jul-20, 985-Jul-20.

Eclipses from the set $\mathbf{G^3}$ (of type **H3**): 949-Dec-22, 968-Dec-22.

Eclipses from the set $\mathbf{G^4}$ (of type **H4**): 953-Sep-25, 972-Sep-25, 991-Sep-26 (visible in Jerusalem, but not visible in Constantinople and Caves).

Eclipses from the set $\mathbf{G^5}$ (of type **H5**): 894-Jun-07, 913-Jun-07, 978-Jun-08.

Eclipses from the set $\mathbf{G^6}$ (of type **H6**): 918-Feb-28, 937-Feb-28, 956-Feb-28, 964-Mar-01, 983-Mar-01.

Eclipses from the set $\mathbf{G^7}$ (of type **H7**): 920-Jul-18, 939-Jul-19, 966-Jul-20, 985-Jul-20.

Now for every one of the listed above eclipses from the sets $\mathbf{G^1}$ – $\mathbf{G^7}$ in the interval AD 890-1000 we find the score for its "closeness" to the respective element of the Template: for the eclipses from $\mathbf{G^1}$ – to **H1**, from $\mathbf{G^2}$ – to **H2,** and so on.

According to the proposed model, if **E** is an arbitrary solar eclipse, the score for its closeness to **H1** should be calculated in the following way:

Scores for evaluation of the closeness of E to the eclipse H1:

The total score $\mathbf{e_1}$ for the closeness of an eclipse **E** to the eclipse **H1** is the sum of the scores for **1-1** and **1-2**:

1-1 The date of the eclipse H1 is November 11th. If the date of E is

- November 11 => score for **1-1:** ***m*** points;
- November 10 or 12 => score for **1-1:** 0.9***m*** points;
- November 9 or 13=> score for **1-1:** 0.5***m*** points;
- Another day => score for **1-1:** 0 points.

1-2 If the score for **1-1** is 0 points, the score for **1-2** is also 0 points; if the score for **1-1** is different from 0, the score for **1-2** is determined by the following rules. Taking into account, that the day of the week on which the eclipse **H1** occurred is Monday, if the day of the week on which occurred **E** is

- Monday => score for **1-2:** 0.5***m*** points;
- Tuesday or Sunday => score for **1-2:** 0.4***m*** points;
- Wednesday or Saturday => score for **1-2:** 0.3***m*** points;
- Another day => score for **1-2:** 0 points.

Applying these rules for **E = 923-Nov-11** (November 11, 923 was in Tuesday) we find: Score **1-1**: **m**; for **1-2**: 0.4**m; Total**: 1.4 **m.**

Similarly:

E = 942-Nov-11 (Friday). Score for **1-1**: **m**; for **1-2**: 0; **Total**: **m.**

Closeness to H2, H3 and so on:

According to the proposed model, if **E** is an arbitrary solar eclipse, the rules for calculating the score for its closeness to **H2** are similar to that for the closeness to **H1**. Applying them to the eclipses of the set $\mathbf{G^2}$ in the interval AD 890-1000, we find the scores for their closeness to the respective element of the Template – **H2**:

E = 920-Jul-18 (Tuesday). Score for **2-1**: 0.9**m**; for **2-2**: 0.3**m**; **Total**: 1.2**m.**

E = 939-Jul-19 (Friday). Score for **2-1**: **m**; for **2-2**: 0.4**m**; **Total**: 1.4**m.**

E = 966-Jul-20 (Friday). Score for **2-1**: 0.9**m**; for **2-2**: 0.4**m**; **Total**: 1.3**m.**

E = 985-Jul-20 (Monday). Score for **2-1**: 0.9**m**; for **2-2**: 0**m**; **Total**: 0.9**m.**

For the eclipses of the set $\mathbf{G^3}$ we obtain:

E = 949-Dec-22 (Saturday). Score for **3-1**: 0.5**m**; for **3-2**: 0; **Total**: 0.5**m.**

E = 968-Dec-22 (Tuesday) Score for **3-1**: 0.5**m**; for **3-2**: 0.5**m**; **Total**: **m.**

For the eclipses of the set $\mathbf{G^5}$ we obtain:

913-Jun-07 (Monday, Easter on March 28). Score for **5-1**: 0.5***m*;** for **5-2**: 0.3***m;***

Score for Easter in the year of E_5**: p; Total**: 0.8**m + p.**

978-Jun-08 (Saturday, Easter on April 11). Score for **5-1**: 0.9***m*;** for **5-2**: 0;

Score for Easter in the year of $\mathbf{E_5}$**: 0; Total**: 0.9**m.**

For the eclipses of the set $\mathbf{G^7}$ we obtain:

920-Jul-18 (Tuesday). Score for **7-1**: 0.5***m*;** for **7-2**: 0.4**; Total**: 0.9**m.**

939-Jul-19 (Friday). Score for **7-1**: 0.9***m*;** for **7-2**: 0; **Total**: 0.9**m.**

966-Jul-20 (Friday). Score for **7-1**: ***m*;** for **7-2**: 0; **Total**: **m.**

985-Jul-20 (Monday). Score for **7-1**: ***m*;** for **7-2**: 0.5***m***; Total**: 1.5**m.**

Now for the lunar eclipses H4 and H6

According to the proposed model, if **E** is an arbitrary lunar eclipse, the score for its closeness to **H4** should be calculated in the following way:

Scores for evaluation of the closeness of E to the eclipse H4:[1]

The total score $\mathbf{e_4}$ for the closeness of a lunar eclipse **E** to the lunar eclipse **H4** is the sum of the scores for **4-1, 4-2** and **4-3**:

4-1 The date of the eclipse **H4** is September 27. If the date of **E** is

- September 27 => score for **4-1:** ***m*** points;
- September 26 or 28 => score for **4-1:** 0.9***m*** points;
- September 25 or 29 => score for **4-1**: 0.5***m*** points;
- Another day => score for **4-1**: 0 points.

4-2 If ***E*** was seen in the Eastern part of the Empire (Jerusalem),

the score for **4-2** is 0.5***m*** points, otherwise 0 points.

4-3 If ***E*** was not seen in the Western part of the Empire,

the score for **4-3** is 1.5***m*** points, otherwise 0 points.[2]

If the score for **4-1** is 0 points, the scores for **4-3** and **4-2** are also 0 points.

Applying them to the eclipses of the set $\mathbf{G^4}$ in the interval AD 890-1000, we find the scores for their closeness to the respective element of the Template – **H4**:

---

[1] These rules are slightly different from the respective version in the paper [Tabov, 2013], but are identical with the version in the paper [Tabov & Umlenski, to appear].

[2] The above mentioned difference is in this rule,

953-Sep-25 Score for **4-1**: 0.5**m; for 4-2**: 0.5**m;** for **4-3**: 0; **Total**: **m.**

972-Sep-25 Score for **4-1**: 0.5**m;** for **4-2**: 0.5**m;** for **4-3**: 0; **Total**: **m.**

991-Sep-26 Score for **4-1**: 0.9**m;** for **4-2**: 0.5**m;** for **4-3**: 1.5**m; Total**: 2.9**m.**

Similarly for the eclipses of the set **G^6 in the interval AD** 890-1000 we find the scores for their closeness to the respective element of the Template – **H6:**

918-Feb-28 Score for **6-1**: 0.5***m*;** for **6-2**: 0.4***m;*** **Total**: 0.9**m.**

937-Feb-28 Score for **6-1**: 0.5***m*;** for **6-2**: 0; **Total**: 0.5**m.**

956-Feb-28 Score for **6-1**: 0.5***m*;** for **6-2**: 0.3***m;*** **Total**: 0.8**m.**

964-Mar-01 Score for **6-1**: 0.9***m*;** for **6-2**: 0; **Total**: 0.9**m.**

983-Mar-01 Score for **6-1**: 0.9***m*;** for **6-2**: 0.4***m;*** **Total**: 1.3**m.**

According to our analysis of the data in the *Chronicle*, three of the intervals between the seven eclipses are most probably exact, or almost exact – between **H1** and **H2**, between **H3** and **H4** and between **H6** and **H7**. In accordance with this vision the model suggests higher scores for closeness of the respective intervals in a septet and in the Template. Therefore we will form preliminary suitable pairs of eclipses among the chosen above, so that the interval between the eclipses in every such pair is close \in length) to the respective interval in the Template; this procedure will make easier the further choice of septets of eclipses with highest scores.

We start with pairs of eclipses from **G^1** – **G^2** (here the points for the interval are determined by the following rule: If the interval between **E_1** and **E_2** is: 16 years => ***f_1*** = ***n*** points, 15 or 17 years => ***f_1*** = 0,9***n*** points, 14 or 18 years => ***f_1*** = 0,4***n*** points, in other cases ***f_1*** = 0 points.):

E_1 = **923-Nov-11 (Score** 1.4 **m) & E_2 = 939-Jul-19 (Score** 1.4 **m)**

The interval between **923-Nov-11 & 939-Jul-19** equals **16 years =>** The score for this interval equals **n.**

Total score for the pair **923-Nov-11 & 939-Jul-19 =>** 2.8**m + n.**

Similarly:

E_1 = 942-Nov-11 (Score m) & E_2 = 966-Jul-20 (Score 1.3 **m)**

The interval between **942-Nov-11 & 966-Jul-20** equals **24 years =>** the score for this interval equals **0.**

Total score for the pair **942-Nov-11 & 966-Jul-2 =>** 2.3**m.**

Similarly for the pairs of eclipses from **G^3** – **G^4** (here the points for the interval are determined by the following rule: If the interval between **E_3** and **E_4** is: 5 years => ***f_3*** = ***n*** points, 4 or 6 years => ***f_3*** = 0,9***n*** points, 3 or 7 years => ***f_3*** = 0,4***n*** points, in the other cases ***f_1*** = 0 points.):

E_3 = 428-Dec-22 (Score 1.3 **m) & E_4 = 432-Sep-25 (Score m)**

The interval between **428-Dec-22 & 432-Sep-25** equals **4 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **428-Dec-22 & 432-Sep-25 =>** 2.3**m +** 0.9**n.**

E_3 = 447-Dec-23 (Score 1.4 **m) & E_4 = 451-Sep-26 (Score** 1.4 **m)**

The interval between **447-Dec-23 & 451-Sep-26** equals **4 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **447-Dec-23 & 451-Sep-26 =>** 2.8**m +** 0.9**n.**

E_3 = 447-Dec-23 (Score 1.4 **m) & E_4 = 489-Sep-25 (Score m)**

The interval between **447-Dec-23 & 489-Sep-25** equals **34 r. =>** The score for this interval equals **0.**

Total score for the pair **447-Dec-23 & 451-Sep-26 =>** 2.4**m.**

E_3 = 949-Dec-22 (Score 0.5 **m) & E_4 = 953-Sep-25 (Score m)**

The interval between **949-Dec-22 & 953-Sep-25** equals **4 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **949-Dec-22 & 953-Sep-25 =>** 1.5**m +** 0.9**n.**

E_3 = 968-Dec-22 (Score m) & E_4 = 972-Sep-25 (Score m)

The interval between **968-Dec-22 & 972-Sep-25** equals **4 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **968-Dec-22 & 972-Sep-25 =>** 2**m +** 0.9**n.**

E_3 = 968-Dec-22 (Score m) & E_4 = 991-Sep-26 (Score 2.4 **m)**

The interval between **968-Dec-22 & 991-Sep-26** equals **23 r. =>** The score for this interval equals **0.**

Total score for the pair **968-Dec-22 & 991-Sep-26 =>** 3.4**m.**

Pairs of eclipses from **G^6 – G^7** (here the points for the interval are determined by the following rule: If the interval between **E_6** and **E_7** is: 1 year => f_3 = ***n*** points, 0 or 2 years => f_3 = 0,9***n*** points, 3 years => f_3 = 0,4***n*** points, in other cases f_1 = 0 points):

E_6 = 918-Feb-28 (Score 0.9 **m)** & **E_7 = 920-Jul-18 (Score** 0.9 **m)**

The interval between **918-Feb-28 & 920-Jul-18** equals **4 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **918-Feb-28 & 920-Jul-18 =>** 1.8**m +** 0.9**n.**

E_6 = 937-Feb-28 (Score 0.5 **m) & E_7 = 939-Jul-19 (Score** 0.9 **m)**

The interval between **937-Feb-28 & 939-Jul-19** equals **2 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **937-Feb-28 & 939-Jul-19 =>** 1.4**m +** 0.9**n.**

E_6 = 964-Mar-01 (Score 0.9 **m) & E_7 = 966-Jul-20 (Score m)**

The interval between **964-Mar-01 & 966-Jul-20** equals **2 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **964-Mar-01 & 966-Jul-20 =>** 1.9**m +** 0.9**n.**

E_6 = 983-Mar-01 (Score 1.3 **m) & E_7 = 985-Jul-20 (Score** 1.5**m)**

The interval between **983-Mar-01 & 985-Jul-20** equals **2 r. =>** The score for this interval equals **0.9n.**

Total score for the pair **983-Mar-01 & 985-Jul-20 =>** 2.8**m +** 0.9**n.**

For calculating the scores of the septets we have to take into account the following rules for giving scores for the intervals:

If the interval between **E_2** and **E_3** is: 29 years => f_2 = 0.2***n*** points, 28 or 30 years => f_2 = 0,1***n*** points, in the other cases f_2 = 0 points.

If the interval between **E_4** and **E_5** is: 7 years => f_4 = 0.2***n*** points, 6 or 8 years => f_4 = 0.1***n*** points, in the other cases f_4 = 0 points.

If the interval between **E_5** and **E_6** is: 5 years => f_5 = 0.2***n*** points, 4 or 6 years => f_5 = 0.1***n*** points, in the other cases f_5 = 0 points.

6. Conclusion

The analysis of the above results shows that highest scores have (and consequently are most perspective) the following two septets:

I. Septet 923 & 939 (Score 2.8**m + n), 968 & 972** (Score 2**m +** 0.9**n), 978 (**Score 0.9**m), 983 & 985** (Score 2.8**m +** 0.9**n).**

If the interval between **E_2** and **E_3** is 29 years = > f_2 = 0.2***n.***

If the interval between **E_4** and **E_5** is 6 years = > f_4 = 0.1***n.***

If the interval between $\mathbf{E_5}$ and $\mathbf{E_6}$ is 5 years = > $\boldsymbol{f_5}$ = 0.2***n.***

Total score: 2.8**m + n +** 2**m +** 0.9**n +** 0.9**m +** 2.8**m +** 0.9**n +** 0.2 ***n*** + 0.1 ***n*** **+** 0.2 ***n*** **= 8.5 m + 3.3 n.**

II. Septet. 923 & 939 (Score 2.8**m + n), 968 & 991** (Score 3.4**m), 978 (**Score 0.9**m) 983 & 985** (Score 2.8**m +** 0.9**n).**

If the interval between $\mathbf{E_2}$ and $\mathbf{E_3}$ is 29 years = > $\boldsymbol{f_2}$ = 0.2***n.***

If the interval between $\mathbf{E_4}$ and $\mathbf{E_5}$ is -14 years = > $\boldsymbol{f_4}$ = 0.

If the interval between $\mathbf{E_5}$ and $\mathbf{E_6}$ is 6 years = > $\boldsymbol{f_5}$ = 0.1**n.**

Total score: 2.8 **m + n +** 3.4 **m +** 0.9 **m +** 2.8**m +** 0.9**n + 0.2 n + 0.1n =** 9.9 **m +** 2.2 **n.**

For ***m = n = p = 10*** we have:

$$12\,m + 3{,}6\,n + p = 166;$$

the distance from **I. Septet** to G_H equals

$$d = 12\,m + 3{,}6\,n + p - (e_1 + e_2 + \ldots + e_7 + f_1 + f_2 + \ldots + f_6 + g)$$

$$= 166 - (8.5\,\mathbf{m} + 3.3\,\mathbf{n}) = 166 - 118 = 48,$$

and the distance from **II. Septet** to G_H equals

$$d = 12\,m + 3{,}6\,n + p - (e_1 + e_2 + \ldots + e_7 + f_1 + f_2 + \ldots + f_6 + g)$$

$$= 166 - (9.9\,\mathbf{m} + 2.2\,\mathbf{n}) = 166 - 121 = 45.$$

Hence, according to our model, **II. Septet** has advantage before the **I. Septet** and the other septets; for more precise results and conclusions further investigations are necessary; they should involve in particular analysis of the time and the phases of the eclipses.

Bibliography

[Tabov, 2013] J. Tabov. Model for Astronomical Dating of the Chronicle of Hydatius. International Journal INFORMATION MODELS & ANALYSES, Vol. 2/ 2013, No. 1, 23-36. ISSN 1314-6416 http://www.foibg.com/ijima/vol02/ijima02-01-p03.pdf

[Hydatii, 1894] Hydatii Lemici. Continuatio chronicorum Hieronymianorum. Monumenta Germaniae Historiae: Auct. Antiq., XI (ed. Mommsen). Apud Weidmannos, Berolini, MDCCCXCIV.

[Idatii, 1619] Idatii Episcopi Chronicon, et Fasti Consvlares. Lutetiae Parisiorum. Ex Officina Nivelliana. Apud Sebastianvm Cramoisy, via Iacobaea, sub Ciconiis, M.DC.XIX.

[Idatii, 1634] Idatii Episcopi Chronica. In: Prudencio de Sandoval. Historias de Idacio Obispo que escrivio poco antes que España se perdiese: De Isidoro Obispo de Badahoz, De Sebastiano Obispo de Salamanca, De Sampiro Obispo de Astorga, De Pelagro, Obispo de Quiedo. Impresso en Pamplona, por Nicolas de Assiayn Impressor del Reyno de Navarra, M.DC.XXXIV.

[Idatii, 1861] Idatii Episcopi Chronicon. Patrologia Latina v. 51, 1861.

[Tabov & Umlenski, to appear] "Soft" model for dating of the eclipses in the Chronicle of Hydatius: results in the interval AD 300-600. To appear.

Authors' Information

Jordan Tabov *– Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, Acad. G.Bonchev Str. Block 8, 1113 Sofia, Bulgaria; e-mail: tabov@math.bas.bg*

Major Fields of Scientific Research: Applications of mathematics and informatics in the humanities, Didactics of mathematics and informatics

ESSAY ON ORDER

Karl Javorszky

***Abstract**: We offer a definition for that elusive term, "order". Order so much directs our perception, cognition and intellect that we should take a look at the infrastructure of how the human mind builds up its view of the world. Order is shown to be a combined statement about properties of things and how they are placed. The interdependence between "where" and "what" has been split for processing by the brain due to evolutionary pressure. Similarity is neurologically preferred above the properties of dissimilarities and is used as an intercultural tool of communication. We present a model that combines place and property attributes and integrates the dissimilarity properties of the present state of the world to the alternatives. We re-introduce the differences that we were instructed to ignore while we have learnt basic dexterity with numbers. We influence, and encourage dissolving, a deep-seated posthypnotic suggestion of culture: "The foreground, the similarities are important."*

***Keywords**: artificial intelligence, theoretical physics, indexed, multi-valued logical statements; stepwise additivity of {.t.|.f.}; accounting concepts in arithmetic procedures; concurrent usage of logical operators; consolidation among contradicting sentences, Minkowski space model.*

Introduction

An essay is „a short literary composition dealing with a subject analytically or speculatively; an attempt or endeavor, effort; a test or trial" [1]. The essay presented here attempts to introduce to the International Conference on Natural Information Technologies a method of thinking that is rooted in the natural processing of information, as it happens in actual life.

After some 35 years of professional work in clinical psychology, one may feel emboldened to say that one has gained some insights into how information processing takes place in actual life. The human brain is the information processing mechanism we use to evaluate – among other things – the contributions to this Conference on the subject of Natural Information Technologies.

Looking attentively to the way the human brain functions educates one on the subject of information processing. The perceptional mechanisms that are the infrastructure of thinking are a product of the development of the human race. Nature, by the evolutionary pressure of selection, has made sure that only such brains will be inside of cranium of participants of this Conference, which obey its rules. One may presuppose that the participants of this Conference are of sound mind and are capable of reasoning, communicating and understanding.

The present attempt of offering an idea has good chances of finding resonance. The reason for optimism is that the participants of this Conference are, with regard to neurological capacities, quite able to understand what is communicated in the present essay. The task is to show to the willing, prepared and interested participants the idea in such a fashion that they can catch and absorb it. One learns in psychology the technique of "reframing" a system of thoughts, by giving a different background to them. We attempt the introduction of dynamism to the rational way of seeing the world. Re-learning fundamentals does not happen overnight. We encounter what is termed in the trade "resistance", which means a wish for consistency in one's identity, being cautious of perceiving something in a different light, an unwillingness to discard the well-trained methods of evaluating

something, the reluctance to leave the well-trodden paths of one's thinking, the comfort one derives of being at home in one's way of doing things.

Knowing that the participants of this Conference are blessed each with a brain that is capable of learning something new, and assuming that the participants participate because they are willing, prepared and interested in the subject of natural information technologies, the task of the essayist is to present the subject in a fashion that invites an interested participant to make the effort of thinking it thru. The contribution of the reader of this essay would be, then, to check his [2] conservatisms with regard to what constitutes rational thinking, as opposed to natural thinking. The title of the Conference: "Natural Information Technologies" does hint that there exists an opposition between it and information technologies that are not really natural. This contrast is the subject of the present essay.

Artifacts

An essay being a short work, only a few, necessarily selective and cursory, sentences introduce the term "artifact" as understood in psychology and neuroscience [3]. An artifact is an unavoidable bias that is present during an experiment. An example would be our inability to determine the weight of a living organism with the precision usual in Physics, because the living organism continuously interacts with its surroundings, e.g. by breathing, sweating, eating and discharging, etc.; the artifact of living imposes a bias on the exactitude of measurements. Similarly, the artifact of social interactions makes it impossible to determine in a conclusive fashion, once and for all, the opinion of the general public with respect to the Punic Wars: as long as there were, are and will be historians, there will be differing opinions on the role of Hannibal. Artifacts are a fact of life and when planning experiments in social psychology, one will always try to discount them.

Among the artifacts of perception, one will have to accept the influences of optical illusions, because we cannot escape the fact that our visual apparatus does process sensory input the way it does. We can deal with neurological artifacts: we have learnt to discount the Doppler Effect; we also know that it is not the Sun rising and setting, although the fact of our stationary place on a rotating body generates this sensory impression. Once we have realized that our brain generates neurological, psychological or mental artifacts, we are ready to discount their effects and feel enlightened by not falling prey to the bias our neurology imposes on our perception and naïve thinking.

With regard to rational thinking, there is still an extended catalogue of perceptional and cognitive artifacts that need some reevaluation. The present essay will enumerate the most obvious among the cultural conventions that we have learnt to accept as axiomatic. There is no danger of insanity in contemplating the ways the human brain builds up its view of the world. We in our culture of the 21st century smile about the resistance scientists and the general public of bygone ages have offered against ideas that went against the socio-cultural axioms of the respective ages. We can glide over the arguments against the idea that the Earth is a big globe: no one today offers the argument that the "antipodes" would fall off the Earth, if it were actually round, and so on. The ideas that appeared hair-raising of the day have made it eventually into both common-sense and scientific thinking. A fine example for accepting the formerly 'unthinkable' as a self-evident axiom is the discovery that children as young as newborns have a neurology and try to maximize sensual pleasures; this thought, perceived as scandalous, was encountering strong resistance in the cultural environment of the last years of the Monarchy.

What counts as a self-evident, rational truth is deeply dependent on the cultural environment of the day/decade/generation. The fact that a Conference is called that dedicates itself to Natural Information Technologies gives rise to the hope that our present day/decade/generation is ready to question what is the present cultural agreement on Information Technologies and in which ways would be a modification necessary and acceptable in order to make the presently orthodox view of Information Technologies to be changed and

become a system of thoughts and insights that merits the name Natural Information Technologies. Working on the artifacts of perception and cognition in the domain of rational thinking could help in this endeavor.

Factors Improving Survival

Among the many faculties that determine, which individual is among the fittest that survive, the ability to recognize similarities is of high prominence. The ability to think in analogies is to recognize that a present impression is similar to an impression that the organism has encountered previously, and this faculty is the basis both for the memory and the ability to learn. No wonder then that our culture lays a great emphasis on similarities. In fact, our abstract, rational thinking is based on a picture of the universal unit, which is devoid of any properties and similar to replicates of its own shape. We base our rational thinking on the idea of the unit, denoted as "1", and we live in the idealized world, where every sum is made up of 1s. We decree that it is rational to believe that the basic building block of our concepts is one of unit properties. This idea may be pleasing to the neurology, and makes thinking a simple matter of dealing with uniform units, yet we may be seduced by insight to improve on it.

Our ability to focus on a specific thing of interest is good for survival. Pointing out one specific instance of something is good if we want to hunt it down and eat or impregnate it, but as always, this advantage has also its drawbacks. By single-mindedly neglecting the surroundings of the target object we gain survival advantages, so we promote this technique as a superb tool of thinking rationally. We save the trouble of having to decide, which of the aspects of the mental picture the important one is, but we incur the costs of not training decisions, and generally domesticate our thinking into believing that a well-founded logical picture of the world is free of conflicting, even contradictory results of evaluations.

We can very well distinguish between objective and subjective. One is outside and factual; the other is personal, intimate and not so easy to communicate. So we use the mental techniques of contrasting, and differentiate between what is the foreground and what is the background. No communication can be understood unless it relates to its specific background; yet in a rational discourse it is a cultural taboo to switch between that what is clear, circumscribed, defined and that what is the subjective background, the insinuation and the debatable. Only the poets have the liberty to use the connotations freely, by others we usually see in the infusion of background information a sign of a troubled mind. We are trained – and used - to believe that a mind works correctly, if it restricts its public communications to elements of the foreground [4]. Yet, poets are a part of Nature.

Small children, even animals, are able to distinguish between bigger and smaller, more or less, quiet or loud, bright or dim. This faculty of discrimination can be observed by far earlier than the ability to make additions. A child can distinguish between extents or amounts that we describe by 2, 3, 5 (size dolls, heaps of chocolate, etc.) far earlier that it can figure out that 2+3=5. The ability to use the logical relations {<|=|>} is at the disposal of a child of some 24-36 months, but we believe formal logic begins with its education in the marvels of the logical relations within the set of logical sentences that are based on the operator {+} within the domain of {=}. Somehow, we have come to look down on the simple ability to discern on size – or extent, or intensity, etc. – as a low-level, proletarian thing that every moron can do and which is not thinking at all.

These are but a few of the neurological artifacts of our brain that influence our thinking. We shall now look into ways to counterbalance the illusions that they create in our view of the world.

Counteracting the Effects of Evolution

The biases in our perception of the world show that we as genus 'homo sapiens' are part of the animal kingdom and are subject to evolutionary pressures. Our efforts to recognize and counterbalance the artifacts of our

neurology show that we insightful, intelligent animals. The task is now to identify areas of certainties in our picture of the world which are nothing more than artifacts of our nervous system.

Prominent among the areas needing a re-evaluation is our concept of what is a unit. The newborn encounters the world at first by means of its skin. Tactile sensations have an impact on our convictions that cannot be overestimated. The skeptic's "I'll believe it when I see it" is trumped by "Pinch me, I can't believe my eyes". We ascribe reality to an object which is – at least conceptually – accessible to be hold in one's hands. We are used to separate the logical categories of "logical object" and "logical relation". If we discuss – in a traditional manner – matters of combinatorics, we understand well the question, how many distinct logical relations can be maximally generated using n objects with some symbols on them. Reversing the direction of concluding is unusual: 'Having x logical relations, at least how many fractions of objects are necessarily present?' borders on poetry. The proposal to use the number of logical relations as unit and deduct the number of actual objects as a result of combinatorics, or, more practically, to use a unit that is a half-way freak between a purely logical entity – like a possibility – and a logical entity, like the unit in terms of objects, that could – if it was real – be touched [5] is a step towards overcoming a dichotomy that is present in our culture, because we know the difference between reality and imagination.

Similarly, we have learnt to distinguish well between the place of an object and the object itself. It is self-evident that a thinking person can distinguish between the direction the Sun is shining from and the light it emits; the time of the tide and the water. Only imbeciles would confound where something is to be found and what it is. The rational view of the world is extremely snobby against members of society who are unable or unwilling to respect the properties of places, e.g. eating or rioting in a church, talking about secrets in an inappropriate place. The properties of the place and the properties of the objects are connected by culture: there are types and categories of things that do and do not match with types and categories of places. This distinction is not purely man-made: we observe animals to discharge waste according to topological criteria. The concept of territoriality itself is however not constant over all kinds of animals: there are nomadic versions that survive well.

Let us come forward with an educational tool which can be used to de-condition our conditioned reflexes and culturally motivated prejudices about what is rational and acceptable in a picture of the world. Before presenting the tool, we first have to introduce the collection of logical objects we shall use and the main method of dealing with the collection of logical objects.

The Thing-As-Such and Its Place

If a solution to a problem is working well, the animal will repeat it. Humans find it useful to unify and to abstract. Discarding the particularities of an object allows building a more general concept of the kind of object this particular one is a representative of. Kant has repeated this step-by-step process to its ultimate end, where he arrived at the idea of the thing-as-such ('t-a-s'). A purely logical concept, the t-a-s is the idea of a thing without any of the properties of an actual thing. There might be, however, some advantages in doing one step less in the process of abstraction and leave a bunch of things as such with some basic properties of type, like the modality [6]. Overdoing the abstraction defeats its purpose of having a concept that is useful and helpful in achieving some goals [7]. Here, we shall use the t-a-s in its appearance as the mathematical unit of "1" but allow for multiples of the unit to be of a distinct modality [8]. Genetics teaches us that Nature does not go to the Shannon extreme of abstracting into "this" and "not this", but keeps at least 4 markers in use and distinguishes between left, middle and right among places. One of the differences between Information Technologies and Natural Information Technologies may be that the latter does not continue schematizing and abstracting to the very end, but allows for a small number of types and kinds of t-a-s to remain distinct. This concept is used in psychology under the

name of "archetypes", of which there are several, the last step of abstraction into an idea of the archetype-as-such not being done.

The management of the place-related properties of an object appears to be done by different regions of the brain than utility-related properties of the thing. We experience the well-established distinction between topical and sensual properties of the objects that make up our world by pairing the attributes of things in a mental experiment. That the place is clean does not mean that things from there can be eaten; edible things loose on the other hand their property of nourishment if they are or were located on a place which by its properties negates the edibility of things. The strict distinction between where something is to be found and what kind of a thing it is becomes immediately evident in the resistance against the idea that it is useful to *sort* things according to their properties.

Sorting is a central concept of the model to be presented. This essay attempts to drive the reader's attention towards sorting and resorting data sets. Normally, people treat the place of an individual thing as less relevant to the thing as its other properties. There is a reason for this condescending attitude towards the collection of comparisons {<|=|>} that assign a place to each element of the data set. After all, it is only sorting and ordering, and if the object were in a different multitude, its place would be a different one [9]. We overcome this critique by assuring that the individual elements are and remain in one and the same collection of individual elements; therefore the place an element will have depends only on its properties and the concept of order we impose on the whole of the collection.

We demonstrate the idea of order on a collection of things-as-such. It is obvious that a spider maintains a clear idea of what is order, and we cannot discuss the fact that spiders are a part of Nature. The task is then to figure out, what logical rules a spider obeys while establishing what to do so that its web is well-ordered. In order to be able to discuss, what is well-ordered we first have to discuss what order is as such. To do so, we introduce a collection of things-as-such and order and reorder them under several of their aspects.

Aspects of Order

There is no overall social convention about what is order; much less about what would be an ideal order. In order to have an intellectual concept, which is neutral and accessible to all, we propose to use the results of a sort on a data set to declare that the data set is in order after the sort has been conducted [10]. We point to the data set in its state as a sorted one and say "the data set is presently in order <…>", where we insert between the symbols < and > the sorting criteria which were used to bring the data set into this specific order.

To keep the discussion simple and general at the same time, we use as elements of the data set things-as-such which have retained some of their qualities that is, are not completely indistinguishable from each other. To be precise, we use numbers in the range 1 to 16. Learning from Nature that there is a dichotomy of two versions of the same idea [11], we build pairs of the things-as-such and call these *a* and *b.* The tool we demonstrate concepts of order on is then a collection of numbers in the range 1 to 16, always two together, where we call the left one *'a'* and the right one *'b'*. One may visualize the collection as summands in an addition of the form *'a+b=c'*. To put it simply, we generate the 136 smallest additions [12] and sort and resort them.

The starting order of the collection will be dependent on the way we have programmed the loops which generate *a* and *b*. If the collection has been generated so, that the data set is ordered as (1,1), (1,2), (1,3), (1,4) and so on, we speak of an order SQ_{ab}, if the elements come generated in the sequence (1,1), (1,2), (2,2), (1,3) and so on, we speak of an order SQ_{ba}.

The sort has assigned to each element a place in a linear sequence 1..136. We see that element (1,1) is in both of the sorts we have encountered so far on place 1; place 2 is occupied in both orders SQ_{ab} and SQ_{ba} by element (1,2) [13].

We shall return to the conflicting assignments of places to elements (and of elements to places) by means of imposing an order that is different to the presently prevailing order, as this search for compromise between logical contradictions is the main theme of the present essay. Before doing so, we shall introduce the aspects of additions we shall be using.

One has been told at the age of 6 by Teacher, that the important thing on mastering rational thinking is that one neglects the differences among the summands and between additions if only the result of *'a+b=c'*, namely c is the same. To look into the effects of $a_1 \neq a_2$ while $a_1 + b_1 = c = a_2 + b_2$ has the emotional connotation that a) one has not understood what additions are all about, and b) one disregards Teacher's instructions. Massive resistance against the proposal to a) split hairs on differences of additions with identical results, and then b) sort on these differences, and then c) re-sort and then d) watch the patterns of items moving together, is to be expected.

We use the describing aspects of *a, b, c=a+b, u=b-a; k=b-2a, t=2b-3a, q=a-2b, w=2a-3b, s=17-(a+b|c)* to build a numeric table of alternatives among order concepts.

We have now introduced the data set we are using to demonstrate concepts of Natural Information Technologies. Reader is invited to generate his version of the data set [14], as looking up the numbers simplifies understanding the following discussion on order and disorder.

How Come This Was Not Discussed Yet

Information Technologies has had it easy and comfortable; Natural Information Technologies is a complicated and demanding business. How nice would it be, if all sentences were always true; everything could be seen as more of one unitary unit; only the foreground, not the contrasting alternatives, was important; one could discard differences without any consequences; there were an infinite number of equally valid logical statements: This is the mental world society has built up, intercultural, to be able to deal with problems Nature has confronted us with. The simplifications are partly for neurological reasons: the overriding importance of the similarity property among mental concepts; the determined focus on the contents of the foreground. Acculturation taught us to actively disregard the background, the ungrammatical something enveloping the realm of the rational. The endorphin boost caused by having understood something is visible on young children who enjoy learning. To rely on a codified catechism of what is reasonable and what is to be understood adds a social component to the self-concept: one belongs to the educated if one has understood something that is universally defined as important to understand. Our education is a great tool for endorphin production, as it teaches us to think in a fashion that is the right fashion. Many factors contribute to the stability of a world view; one encounters the continuity/stability in the form of resistance if one suggests a change to the habitual ways of seeing the world.

The history of rational thinking begins with correctly deciding, which of two alternative stimuli of appetence is of more utility: discerning, recognizing differences is a process that happens in the moment, the intelligence needs no memory to evaluate the sensory input [15]. Memory comes in, when after recognizing that b is bigger than a, the animal remembers whether a movement to or from b used to be more useful. Abstracting from most of the properties of a and b and describing them by means of multiples of a standard unit is a product of intelligence which beasts of prey that hunt in groups probably master. The visualizing ability of the brain encounters its natural limits usually near a dozen objects. Only thanks to the technology of the last decades have we the chance to introduce something basically new to concepts of logic. It is impossible for humans to investigate the patterns observable during reordering a set from one into a different order by using his fingers and his brain; methods of paper and pencil fail. Alone, to raise the topics one needs a computer to be able to present a possible way of consolidating the differences between Information Technology and Natural Information Technology. The task is roughly comparable to discussing the individual paths of each bird or fish in a swarm that exercises its

maneuvers. In historic perspective, there was simply no way of formulating the subject of the present essay in previous generations, as they lacked the accounting power of computers.

Natural Information Technology includes that what Wittgenstein has contrasted the subject of his Treatise to, namely that what is not the case. We propose a concept of the world, in which some logical sentences are sometimes true, in dependence of the spectator's decision, which aspect of the world is relevant for him in this moment. The world which we discuss in this essay consists of only two kinds of entities and each kind can have only 16 variants. By tradition, we have become used to reading the usual order of the set under the aspect of the two summands being interchangeable ('commutative') and their sum being the important aspect. We re-introduce the differences we were instructed to ignore at our first schooling in rational thinking, at Elementary School.

Contradictions and Compromises

By using more describing aspects of the pair of logical entities that are puppets we play with, we greatly expand the complexity of the question. By bringing the collection in a sorting order, we can say that SQ_{ab} is the case. In that moment, under this understanding, we know about each element, which is its place; concurrently, we also know about each place, which element is presently in occupation of it.

Contrasting to 'SQ_{ab} is the case' would be the state of the world if 'SQ_{ba} is the case'. Then, element e_i would not be on place p_j, but rather on a different place. The general idea of a resort is that most of the elements lose their place attribute's logical property of .t. and gain the logical property of .t. attached to a different place attribute. This is usually a process in several steps. In his everyday life, the participant has experienced and understood the logical processes involved in a reorder. The area of the brain wherein are stored the experiences relating to places is not directly connected to the parts of the brain that manage quality properties of the mental contents. One has to actively encourage the idea that this essay discusses the places of logical elements. These are pictured in the abstract world as the result of a sorting procedure. The sorting connects places to quantities [16].

Natural Information Technology makes use of the numeric constants that are the result of sorting the collection of additions. Nature appears to agree to the rules classical logic, Information Technology has elaborated. The transition is very smooth; because the mechanism is self-explanatory once one has set into motion the accounting machine. Once one has cleared the logical resistance of accepting *two* equally valid versions of what is the case, the technical solution of the consolidation comes quite naturally, as the numbers are very helpful. Any two of sorting orders can be resorted into each other and the extent of truthfulness of the alternative results of '$SQ_{\alpha\beta}$ is the case vs. $SQ_{\gamma\delta}$ is the case' can at any moment be numerically evaluated and registered.

The rational world, as we are used to understand it, is stable, even rectangular. The space concept this essay proposes is generated by change resp. movement and transformation. The basic idea of a unit in Natural Information Technology is the standard extent of disorder during a resort from a given order into a different order. The fundamental vision transmitted in this essay is a continued logical debate, about which of the views of a and b are more relevant, in this moment, given this history. The movement of the elements merges place- and quantity-related attributes together, creating practicable space concepts. Which of the possible orders is in existence, and to what degree, can at all stages be numerically tracked. What we direct the attention of the reader to is the group of elements that are in interdependence – in a common category - with each other [17]. During a reorder, elements usually are included in a sub-collection that moves together. These sub-collections can be seen as noumena in a logical connection, goods in transit, filaments or strings connecting places in linear, planar and spatial structures with amounts of *a* and *b* and of their aspects.

The numbers are forthcoming enough to supply a standard kind of chain, which is of length 3 and accommodates 4 distinct logical states in each of its readings. The accounting mechanism translates up to 72 differing readings of the order history of the set into 20 or 21 actually distinguishable logical prescriptions.

Let this essay conclude with an optimistic note. The concept is rooted in psychology: that the logical relations {<|=|>} and {=,+} are both accessible to the brain, and using them concurrently is only a matter of training; its content is communicable, as it deals with elements that have a well-known life in abstract thinking: the new trick is to watch the patterns that appear while reshuffling from a specific order into a different one; the demand for a numeric model to simulate Nature's ideas about order is urgent, as applied biology is advancing forward on empirical knowledge. May this essay support Information Technology into becoming a Natural Information Technology.

Conclusion

There is a natural order governing the functioning of the human brain. Computers, as potent reincarnations of paper and pencil, allow us to keep track of what is order, at least in the arithmetic based on natural numbers. Combining place and quality attributes can be achieved by distinguishing elements among each other by comparing them on the basis of {<|=|>}. Sorting is ordering. As there are several sorting alternatives at our disposal, their results will in many cases be contradictory.

We go into details within a central concept of logic, namely that a logical system has to be free of contradictions, and be, in effect, a huge system of tautologies. We allow contradictions to try to exist, without ultimate success.

We consolidate the logical contradictions by assuming that there is a continual search for compromise among contradictions. This position allows us to see Nature as having a dynamic will of its own, that follows its own rules. The order concepts as inferred from natural numbers show contradictions with regard to $a+b=c$ if concurrently subjected to the rules of {<|=|>}. The interaction of the two logical principles allows the contradictions to be consolidated. It is possible to resort a data set into a different sorting order.

The details of the resorting show that there is a third describing attribute to noumena: after the amount (name, extent) and the place (in a given order) there appears the category of *logical connection* as an attribute of the 136 individual elements we model Nature with. The logical connection exists in that group of elements that move together during a resort from a given order to a different order. These strings connect elements in a remarkable fashion.

The world view offered by this extension to arithmetic allows building new concepts in many areas of science. The Minkowski model is well supported by the numbers. One uses a moderately extensive data set of steps each element makes while being resorted from every possible order into every other possible order. The system gets complicated once one enters the combinatorics of not all alternatives being possible at all times.

The rules of selecting the next among the possible states of the system are connected to the ties in the comparisons. There can history enter as a logically circumscribed fact. The fact that elements that would be normally within a tie in a random sequence are ordered is equivalent to stating that previously such-and-.such order has been the case. This method specifically distinguishes one-of-four within a basic – spatial – statement of one-in-three. The basic translation mechanism of theoretical genetics appears to have been found.

Remarks

[1] Mobile-dictionary.reverso.net.

[2] Or her; „he" always means „s/he".

[3] See e.g.: en.m.wikipedia.org/wiki/Cognitive neuroscience of visual object recognition.

[4] Wittgenstein has declined to discuss the background of that, what can be expressed rationally.

[5] Or heard, smelled, tasted, etc.

[6] E.g. sound as such, smell as such, touch as such, movement as such, etc.

[7] See also: Musil, R.: Der Mann ohne Eigenschaften (The Man without Qualities).

[8] That is, we assume 3 to be intrinsically distinct to 1+1+1 or 1+2.

[9] Among the blind, the one-eyed ...

[10] This would be a deictic definition of a concept. We use the deictic method in this essay.

[11] E.g.: female and male versions of Homo sapiens.

[12] From 1+1 to 16+16, where a ≤ b.

[13] Reader is invited to determine the place of (15,16) and (16,16) in both SQ_{ba} and SQ_{ba} on his own.

[14] One may also download the data set from the site www.tautomat.com.

[15] This means that the logical operation based on {<|=|>} is more archaic than {=,±}.

[16] At school, pupils' line up sorted once on name, a different time on height. Their unique SSN is a quantitative attribute, as it is made up of multitudes of the unit '1'.

[17] The reader is invited to draw the succession diagram that shows by which way (1,3) gets from place 3 to place 4 in the resort ab→ba.

[18] Thanks to Gordana Dodig-Crnkovic for pointing out the term „noumenon" for 't-a-s'

Bibliography

The core statement of this essay is a proposal to include into arithmetic the use of numeric constants generated by sorting the natural numbers 1..16 according to some criteria of a and b. There is no literature on the most probable state the system is in during resorts.

As a speculative work of literature, an essay does not need bibliographical references. As a work of logic, Wittgenstein has set a precedent in the foreword to his Tractatus.

Authors' Information

Karl Javorszky *– affiliated to FIS since 1995. Works as a clinical psychologist in Vienna since 1979. Address: Landhausgasse 4/23, A-1010 Wien, Austria; e-mail: karl.javorszky@gmail.com*

Major Fields of Scientific Research: General theory of order and disorder; interaction between sequences and mixtures; a rational model of theoretical genetics; granularity algebra [see: Zaragoza Lectures, on invitation by P.C.Marijuán, in 1994], possible uses of disjunction operator M [a low-level arithmetic tool like division, yielding categories of disjunction].

More material relating to the model presented in this essay can be found in the e-lectures given at the FIS chat room in WS 2012/13, titled "Learn to Count in Twelve Easy Steps".

ESTIMATION OF PEAK SUSTAINABLE POWER CONSUMPTION FOR SEQUENTIAL CMOS CIRCUITS

Liudmila Cheremisinova, Arkadij Zakrevskij

***Abstract**: The reliability and the cost of electronic circuits are closely connected to the maximum power dissipated by them. Tools for evaluating the worst case power consumption of sequential circuits is becoming a primal concern for designers of low-power circuits. In the paper the task of estimation of peak sustainable power for CMOS synchronous sequential circuit is considered when its automaton description in the form of Finite State Machine (FSM) is available. The method is based on finding out the simple directed cycles of FSM state transition graph closely related with test sequences for simulating the sequential circuit for sustainable power estimation.*

***Keywords**: low-power design, power consumption, CMOS circuits, peak power estimation.*

***ACM Classification Keywords**: B.6.1 Logic design: Design Style – Sequential circuits; B.7.3 Integrated Circuits: Reliability and Testing – Test Generation*

Introduction

In the VLSI (Very Large Scale Integration) chip design performance, area and cost were historically the major considerations. But in the last years power consumption has become the major issue in electronic research, it is being given increased weight age in comparison to area and speed because of two main reasons:

- Increasing use of portable and battery operated electronic devices which have limited battery life;
- Continuous increase in chip density resulting in VLSI circuits that contain up to hundreds of millions of transistors;
- Topicality of high performance computing resulting in VLSI circuits that have clock frequencies in the GigaHertz range.

Excessive power dissipation (the unit of power used throughout the paper is energy per clock cycle) in integrative circuits causes their overheating degrading the performance and reducing chip life. To prevent circuits from these consequences discouraging their usage, the chips need costly packaging and cooling arrangements. The Semiconductor Industry Association technology roadmap [SIA, 2014] has identified low power design techniques as a critical technological need in semiconductor industry today.

The development of methods and software tools that can help designers to optimize digital circuits for power consumption has received increasing attention. Accurate and efficient power estimation during design phase is required. The appropriate tools must have efficient means to estimate the power consumed by a circuit on different design phases. At present an increasing attention is focused not only on transistor-level design but on higher levels of abstraction because early power estimation is important in VLSI circuits, because it has a significant impact on the reliability of the circuits under design. And in the process of optimizing circuits for low power a designer is interested in knowing the effects of specific design techniques on the power consumption of the projected circuit. With the relevant information about power characteristics designer can redesign or correct a circuit in early design stages if it is found to consume more power than expected.

The estimation of power in digital circuits has become a significant problem, especially for present day semiconductor technologies. Currently, the simplest and most direct power estimation can be done by circuit simulation when the monitoring of the power supply current is done. So, power consumption values are determined which depend on the given vector set. There are circuit-level power estimators available as commercial tools. For example, the most known SPICE [Nagel, 1973]. But the simulation results are highly related to the input patterns given to the circuit [Kang, 1986]. Simulation methods suffer from two major drawbacks. First, they are very time consuming, especially for large circuits (because to produce a meaningful power estimate the required number of simulated vectors is usually high). Second, it needs to know the set of input patterns when the power for a designed circuit embedded in a large system is to be calculated. Thus, the calculated power may be erroneous because some of input patterns used for estimation may never occur during normal (or verified critical mode) operation.

Using simulators, power is measured for a specific set of input vectors (often chosen randomly), and can be referred to average power consumption. Many investigations were focusing on the average power estimation [Arasu, 2013; Chou, 1996; Ghosh, 1992; Najm, 1994; Wang1, 1996]. The proposed methods are not only simulation-based but probabilistic methods are very popular too. However, the average dissipation is not the only reason of circuit failure. Another critical factor that affects the chip reliability is the value of maximum (or peak) power dissipation that can cause excessive heat generation resulting in permanent damage or temporary circuit failure. Unlike average power estimations in which signal switching probabilities are sufficient to compute the average power [Chou, 1996; Ghosh, 1992; Najm, 1994], maximum power is associated with a specific starting circuit state and a specific input pattern sequence that produce such a power. Although the problem of estimation of maximum power in VLSI circuits is essential for determining the appropriate packaging and cooling techniques, optimizing the power and ground routing networks, there are a limited number of papers devoted to the problem of maximum power estimation of combinational and sequential circuits (most of them are cited in the papers [Kumthekar, 1998; Wu, 2006]).

Static CMOS logic style is used now for the vast majority of logic gates in digital integrated circuits because they have good technological parameters and good power dissipation characteristics. Many ASIC methodologies allow only complementary CMOS circuits custom designs use static CMOS for 95% of the logic [Zimmermann, 1997]. The reliability and the cost of electronic circuits are closely connected to the maximum power dissipated by them. Power and switching activity estimation for sequential circuits is significantly more complex task than that for combinational circuits because power value depends not only on input patterns but on the state the circuit is in. Tools for evaluating the worst case power consumption of sequential circuits are becoming a primal concern for designers of low-power circuits.

In the paper the task of estimation of peak sustainable power for CMOS synchronous sequential circuit when its automaton description in the form of Finite State Machine is available. The proposed method is based on finding out the simple directed cycles of FSM state transition graph (STG) closely related with input patterns for simulating the sequential circuit for sustainable power estimation.

Peak power estimations for sequential logic circuits

The total power dissipated in a CMOS logic gate consists of two basic components: static and dynamic power. In a typical CMOS circuit, most of the power dissipated is dynamic power while static power makes up a small part of the total power dissipated [Balasubramanian, 2007]. The dynamic power component normally dominates in CMOS system-on-chip and accounts for roughly 75% of the total power consumption [Benini, 2002]. The dynamic power dissipation is defined as the power spent in charging or discharging of the nodal capacitances during a high to low or low to high transition at the output node. The dynamic power dissipated is directly proportional to

the circuit switching activity, which is evaluated by the number of gate outputs that toggle (change state) in the circuit. Therefore, the total switching activity is the parameter that needs to be maximized for maximum power dissipation or the peak switching frequency is used to represent the peak power consumption in the circuits.

In the first papers devoted to circuit peak power estimation, the problem was treated as estimating the maximum power that the circuit may consume within any clock cycle. The problem in the case of combinational circuit is equivalent to looking for the maximum-power-consuming vector pair among all possible input vector pairs. For sequential circuits, on the other hand, the activity depends on the initial memory state as well as the primary input vectors, so, the state ought to be added to the pair of input vectors when looking for the maximum power estimation for sequential circuit. Further different design requirements of present VLSI chips make things more complicated. Now three types of peak power are used in the context of sequential circuits [Hsiao, 2000]:

1) Peak single-cycle power;
2) Peak *l*-cycle power;
3) Peak sustainable power.

Their time durations are one clock cycle, *l* consecutive clock cycles and an infinite number of cycles, respectively.

Peak single-cycle power consumption corresponds to the highest switching activity generated in the circuit under the test during one clock cycle. Accurate estimation of maximum power consumption for a combinational circuit involves finding a pair of input vectors which, when applied successively, maximize the number of toggles, among all possible input vector pairs. The estimate of peak single-cycle power dissipation can be used as a lower-bound for worst-case power dissipation in the circuit because found pair of input vectors can be applied one after another repeatedly causing the estimated power as an average for an indefinite time.

For a combinational circuit with n primary inputs, there are $(2^n)2 = 4^n$ possible two input vector sequences to be considered. For a sequential circuit with m memory elements, this number of sequences increases up to $m\,4^n$. The power is controlled by initial memory state vector $\boldsymbol{s}_1$ and input vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The state $\boldsymbol{s}_1$ and input vector $\boldsymbol{x}_1$ initialize all gate outputs and determine the next state $\boldsymbol{s}_2$. Then, the vector $\boldsymbol{x}_2$ and the state $\boldsymbol{s}_2$ switch some of the gate outputs, which accounts for the power dissipation. So in that case a three-tuple ($\boldsymbol{s}_1$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$) must be found that maximizes the instantaneous power consumption.

Peak *l*-cycle switching activity is a measure of the peak average power dissipation over a contiguous sequence of *l* vectors. The *l*-cycle power is related with the sequence ($\boldsymbol{s}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,…, $\boldsymbol{x}_l$) of the length *l*+1. When *l* is equal to 3, the peak *l*-cycle power is the same as the peak single-cycle power dissipation, and with *l* increasing the average power is expected to decrease.

Peak sustainable power is a measure of the peak average power that can be sustained indefinitely over many clock cycles [Hsiao, 1997] it is called as maximum average power too. The peak *l*-cycle power serves as an upper bound to peak sustainable power. And both estimates have sense only for sequential circuits.

The peak average power can be defined as follows. The average power dissipation E_i is maximum if 1) there exist such unrestrictedly long sequence of clock cycles which is characterized by the average power E_i (average power dissipated per a clock cycle); 2) there does not exist the available unrestrictedly long sequence with average power greater than E_i.

Many efforts have been done to attack the problem of peak power estimation. The proposed methods are based on transformation to a weighted max-satisfiability problem on a set of multi-output Boolean functions [Devadas, 1992]; calculation of maximum average length cycles of a weighted directed graph [Manne, 1995]; propagation of signal uncertainty waveforms throughout the circuit [Kriplani, 1992]; automatic test pattern generation techniques [Wang1, 1996]; Monte Carlo based statistical techniques for maximum current estimation [Wang2, 1996]; genetic search algorithms [Hsiao, 1997]; the asymptotic theory of Extreme Order Statistics [Wu, 2006]; ant colony

optimization techniques [Liu, 2009] and others approaches. The majority of the obtained results are devoted to the problem of peak single-cycle power estimation some of them are applicable only for combinational circuits. The difference between combinational and sequential circuits is the memory elements issue. The states of sequential circuits cannot be assigned to arbitrary values but only to reachable ones. If the initial state is initialized to any arbitrary value during the peak power estimation, then the power value will be overestimated since unreachable states are not allowed.

The majority of peak power estimation approaches are based in any event on simulation. The simulation-based power calculation procedure is comprised of three phases: generation (may be randomly) of a sequence of input patterns to be tested (it should have statistically significant size to make conclusions); simulation of the tested circuit on the sequence of input patterns estimating power dissipation on each clock cycle and then calculation of the average value of power dissipation. For sequential circuits the initial sequence of input patterns should start from some reachable state (it may be reset state). The difficulties of usage of simulation-based method for peak power estimation are: 1) the need to generate such a sequence of input patterns that ensures energy critical mode of circuit operation (otherwise we do not get estimate of peak power dissipation); 2) the simulation process is very time consuming because of the great number of simulated vectors for large circuits to produce a meaningful power estimate; 3) the necessity to initialize the tested sequential circuit, to start simulation from a reachable state; 4) baffling complexity of the task because a sequential circuit can be considered as a series of combinational circuits with different initial reachable states.

Problem statement

High level synthesis produces a combined description of the data-path and control logic. The well-known representation of control logic is FSM state transition graph (STG). At the level of logic design a gate-level netlist is generated from a FSM, so a circuit structure is reflected by an appropriate FSM structure.

In this paper the focus is on the upper bound to peak sustainable power for CMOS synchronous sequential circuit. We made the assumption that the circuit automaton description in the form of FSM is available. We seek for test sequences of input vectors that are the candidates to be tested for peak sustainable power dissipation in sequential circuit. The test sequences are derived from augmented STG of the given FSM. The switching frequency is used to evaluate the peak power consumption in the circuit in the process of constructing test sequences that would cause this peak power value. We compute peak sustainable power by finding the average switching frequency for a cycle sequence of FSM transitions that can be repeated infinitely for a long time.

The proposed approach constructs test sequences which are only suspicious (maybe highly) to consuming the peak power, that is because: 1) we are not interested in detail of the target circuit structure and use only its global structure – STG of FSM; 2) the correlation between switching frequency used for test sequence estimate and the actual peak power is indefinite; 3) the process technology is not taken into account. Once the test sequences have been determined circuit-level simulation should be performed to accurately determine the associated values of average power dissipation and to choose the most of them corresponding to peak sustainable power dissipation.

Let T_i denote test sequence represented by k-cycle sequence ($\boldsymbol{s}^i$, $\boldsymbol{x}_1^i$, $\boldsymbol{x}_2^i$,…, $\boldsymbol{x}_k^i$,), where $\boldsymbol{s}^i$ is a FSM internal state represented by a Boolean vector of states of memory elements, $\boldsymbol{x}_j^i$ is a Boolean vector of input variables representing a FSM input state at the j-th clock cycle The values of $\boldsymbol{s}^i$ and $\boldsymbol{x}_1^i$ initialize the circuit at the first clock cycle, before the process of estimating the series of switching's in the circuit.

The paper proposes a methodology for finding out a cyclic sequence that 1) is allowable (although it may have low-probability); 2) may be repeated many times; 3) is the most power-consuming deriving the maximal average (sustainable) power consumption.

Let a k-cycle test sequence T= ($\boldsymbol{s}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,…, $\boldsymbol{x}_k$) be termed cyclic if its simulating generates the sequence ($\boldsymbol{s}$, $\boldsymbol{x}_1$, $\boldsymbol{s}_1$, $\boldsymbol{x}_2$, $\boldsymbol{s}_2$, …, $\boldsymbol{s}_{k-1}$, $\boldsymbol{x}_k$, $\boldsymbol{s}_k$) such that $\boldsymbol{s} = \boldsymbol{s}_k$. It is accepted that a single cycle power estimation metrics (evaluates the power dissipated during any one cycle) is equal to the sum of signal switching's of all circuit nodes.

The method of search for peak power test sequences

Let FSM (A, B, S, Ψ, Φ) be given, where A is the set of input symbols, or input alphabet; B is the set of output symbols, or output alphabet; S is the set of states, or internal alphabet; $\Psi : A \times S \to S$ is the transition function mapping a state and the input symbol to the next state; $\Phi : A \times S \to B$ ($\Phi : S \to B$ for an automaton of Moore type) is the output function mapping a state and the input symbol to the output symbol.

The more obvious representation of an automaton is the state transition graph G that is a directed graph whose vertices correspond to the automaton states, and edges – to the transitions between the states. Any edge of the graph is marked with input symbols, which cause the corresponding transition, and with output symbols going with this transition (in the case of Mealy's automaton). In the case of Moore's automaton, the output symbols mark the vertices corresponding to the states where the automaton is producing those symbols.

We suppose that all the automaton symbos are encoded by Boolean variables. At that, the state symbols $a \in A$, $b \in B$ and $s \in S$ are replaced by the vectors of Boolean variables:

$$a \to \boldsymbol{x} = (x_1, x_2, \dots, x_n);$$

$$b \to \boldsymbol{y} = (y_1, y_2, \dots, y_m);$$

$$s \to \boldsymbol{z} = (z_1, z_2, \dots, z_k).$$

The functions Ψ and Φ are transformed into the vector functions $\boldsymbol{\psi}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{z}^+$ and $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{y}$, and those into the system of $m + k$ Boolean functions.

Let the vertices $v_i \in V$ and edges $t_i \in T$ of the STG G = (V, T) are marked with codes $\boldsymbol{z}$ (and with codes $\boldsymbol{y}$ in the case of Moore type automaton), and the edges $t_i \in T$ of the STG are attached with codes $\boldsymbol{x}$ and $\boldsymbol{y}$ of input and output symbols. Then let a weight to each edge in the STG be attached which indicates the power dissipation estimate and is equal to the number of input, output and internal variables (from the vectors $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$) changing their values as result of automaton transition between two adjacent states. In other words, an edge weight equals to the total toggle count (variables bit changes) per the appropriate state transition calculated as the sum of Hamming distances between codes of input, output and internal states touched with the transition.

For example, let the following sequence of transitions take place:

$$\Psi(a_i, s_c) = s_d, \Phi(a_i, s_c) = b_p, \Psi(a_j, s_d) = s_e, \Phi(a_j, s_d) = b_q,$$

and $s_c \to \boldsymbol{z}_c = (000)$, $s_d \to \boldsymbol{z}_d = (010)$, $s_e \to \boldsymbol{z}_e = (101)$, $a_i \to \boldsymbol{x}_i = (00)$, $a_j \to \boldsymbol{x}_j = (01)$, $b_p \to \boldsymbol{y}_p = (11)$, $b_q \to \boldsymbol{y}_q = (00)$. Then the weight of the edge (s_d, s_e) between the states s_d and s_e equals 6:

$$\mathrm{d}(\boldsymbol{z}_d, \boldsymbol{z}_e) + \mathrm{d}(\boldsymbol{a}_i, \boldsymbol{a}_j) + \mathrm{d}(\boldsymbol{b}_p, \boldsymbol{b}_q) = \mathrm{d}(010, 101) + \mathrm{d}(00, 01) + \mathrm{d}(11, 00) = 6.$$

When looking for cyclic test sequences with maximal average peak power we may content ourselves only with simple directed cycles, i.e. a closed directed walk with no repetitions of vertices or edges allowed. This assumption is reasonable because 1) any directed cycle other than simple one can be constructed as a superposition of some simple directed cycles, and 2) among simple cycles constituting no simple one, there

exists always one better than the composite cycle; 3) we are interested of the most power-consuming directed cycles.

We search for simple directed cycles $C_i = (v_0^i, t_1^i, v_1^i, t_2^i, v^i_2, \ldots, t_k^i, v_0^i)$ of the state transition graph $G = (V, T)$, where $v_j^i \in V$ and $t_j^i \in T$ are vertices and edges of the graph. For each obtained simple directed cycle, weights are calculated for all its edges and then average value of the weight per an edge $t_j^i \in C_i$ which will be the estimate of power dissipation of sequential subcircuit which implements the automaton cyclic sequence corresponding to the simple directed cycle C_i. The estimate carries the comparative character in the sense that it can be used only for comparing different automaton cyclic sequences with each other to compute the maximum average cycles. The approach must construct test sequences which are only highly suspicious to consuming the peak power. These generated test sequences should be applied into a commercial power calculation tool to estimate real power dissipation.

Test sequence $(\boldsymbol{z}^i, \boldsymbol{x}_1^i, \boldsymbol{x}_2^i, \ldots, \boldsymbol{x}_k^i)$ corresponds to a found i-th simple directed cycle $(v_0^i, t_1^i, v_1^i, t_2^i, v^i_2, \ldots, t_k^i, v_0^i)$, where $\boldsymbol{z}^i$ is the starting (and ending) sequential circuit memory state, $\boldsymbol{x}_j^i$ are input patterns causing the transitions from the state $\boldsymbol{z}_{j-1}^i$ to $\boldsymbol{z}_j^i$ (that mark the vertices v_{j-1}^i and v_j^i of the state transition graph $G = (V, T)$) so, for each two consecutive vertices of the cycle, there exists an edge directed from the earlier vertex to the later one.

An example of the search for peak power test sequences

Let consider an example of FSM of Moore type with six states, the corresponding state transition graph $G = (V, T)$ is shown in Figure 1. The vertices of the directed graph G correspond to the automaton states, and edges – to the transitions between the states. A vertex $v_i \in V$ is marked with the code $\boldsymbol{z}_i$ of automaton state s_i and the code $\boldsymbol{y}_i$ (in accordance with the output function $\Phi : S \to B$) of the output symbol b_i in the form $\boldsymbol{z}_i/\boldsymbol{y}_i$. An edge of the graph G corresponds to the automaton transition and is marked with the code of input symbol, which causes the transition. For instance, the vertex v_1 is marked with two vectors 000/01 where 000 is the code $\bar{z}_1\ \bar{z}_2\ \bar{z}_3$ of the automaton state s_1 and 01 is the code $\bar{y}_1\ y_2$ of the output symbol in the state s_1. Two edges from the vertex v_1 marked with 00 and 11 correspond to automaton transition to states s_5 and s_6 under input signals $\bar{x}_1\ \bar{x}_2$ and $x_1\ x_2$.

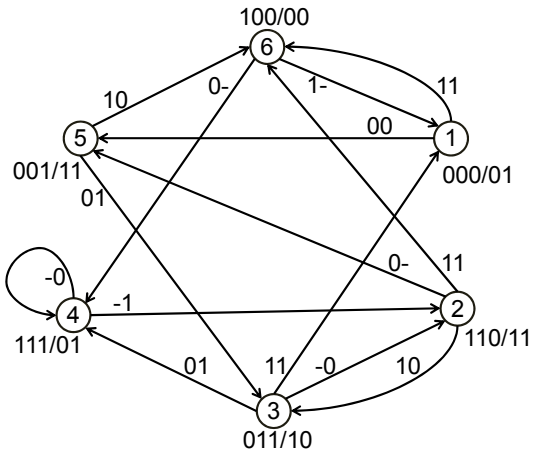


Figure 1. State transition graph of Moore's FSM

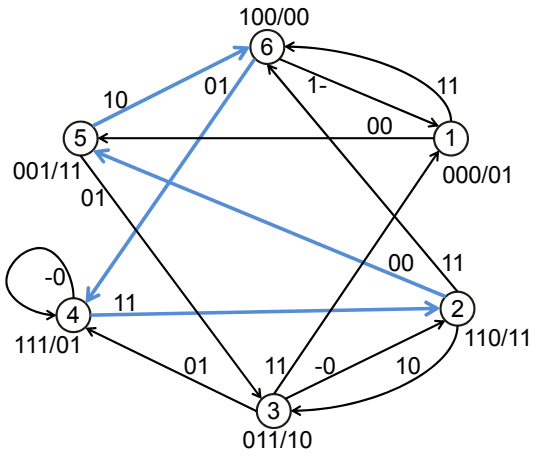


Figure 2. The best directed cycle of the state transition graph of FSM of Moore type

The graph G has 14 simple directed cycles whose vertices are enumerated in the second column of the below shown Table 1. For each cycle, the total toggle count is calculated which consists of the following three values:

1) the sum of maximum numbers of value changes of internal variables z_i; 2) the sum of maximum numbers of value changes of input variables x_i and 3) the sum of maximum numbers of value changes of output variables y_i as the result of appropriate automaton transition (third, fourth and fifth columns of Table 1).

Then the average number of toggle counts per an edge for all cycles is computed (the sixth column of Table 1) which serves as the power estimate of a cycle and allows to range the cycles according to their potentials to be a candidate to be the test for worst-case power dissipation. In considered case we can choose, for instance, the following three test sequences corresponding to the best directed simple cycles 8, 6, 7:

1) $(z_1\, z_2\, z_3;\ x_1\, x_2;\ \ \bar{x}_1\, x_2;\ x_1\ \bar{x}_2;\ \ \bar{x}_1\, x_2)$;
2) $(z_1\, z_2\, z_3;\ x_1\, x_2;\ x_1\ \bar{x}_2;\ \ \bar{x}_1\, x_2;\ x_1\, x_2;\ \ \bar{x}_1\ \bar{x}_2)$;
3) $(z_1\, z_2\, z_3;\ x_1\, x_2;\ x_1\ \bar{x}_2;\ \ \bar{x}_1\, x_2;\ \ \bar{x}_1\ \bar{x}_2;\ x_1\ \bar{x}_2;\ \ \bar{x}_1\, x_2)$.

The directed cycle 8 (the best candidate for estimating maximal average (sustainable) power consumption) is shown at the state transition graph (Figure 2).

Table 1. The toggle characteristics of simple directed cycles of the state transition graph

	Vertices	Total number of variable toggles			
		state	input	output	average
1	1, 6	2	2	2	3
2	1, 5, 6	4	4	4	4
3	1, 5, 3, 2, 6	6	7	6	3,8
4	1, 5, 3, 4, 2, 6	6	4	8	3
5	4, 2, 6	4	4	4	4
6	4, 2, 3, 1, 6	8	8	6	4,4
7	4, 2, 3, 1, 5, 6	10	8	8	4,3
8	4, 2, 5, 6	8	6	4	4,5
9	1, 5, 3	4	4	4	4
10	4, 2, 5, 3, 1, 6	10	8	6	4
11	2, 3	4	2	2	4
12	2, 3, 4	4	4	4	4
13	2, 5, 3	6	4	2	4
14	2, 5, 3, 4	6	4	4	3,5

Conclusion

The task of estimation peak sustainable power for sequential circuit is simplified when its initial automaton description is known. In this case it is shown how it can find out candidate sequences of input patterns that ensure energy critical mode of circuit operation. The approach allows avoiding time consuming generation of the great number of simulated vectors to produce a meaningful power estimate of worst case power consumption.

Acknowledgement

The paper is published with partial support by the project ITHEA XXI of the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua).

Bibliography

[Arasu, 2013] S.A.K. Arasu, F.E.Josy, N.Manibharathi, K.Rajasekaran. BPNN Based Power Estimation of Sequential Circuits. In: Intern. Journal of Advanced Research in Computer Science and Software Engineering, 2013, vol. 3, no. 11, pp. 256–260.

[Balasubramanian, 2007] P. Balasubramanian, C.H. Narayanan, K. Anantha. Low Power Design of Digital Combinatorial Circuits with Complementary CMOS Logic. In: Intern. Journal of Electronics, Circuits and Systems, 2007, vol. 1, no. 1, pp. 10–18.

[Benini, 2002] L. Benini, G.De. Micheli. Logic Synthesis for Low Power. In: Logic Synthesis and Verification. Ed. S. Hassoun, T. Sasao and R.K. Brayton. Boston, Dardrecht, London: Kluwer Academic Publishers, 2002, pp. 197–223.

[Chou, 1996] T. Chou, K. Roy. Accurate Power Estimation of CMOS Sequential Circuits. In: IEEE Trans. VLSI Systems, vol. 4, no. 3, 1996, pp. 369–380.

[Devadas, 1992] S. Devadas, K. Keutzer, and J. White. Estimation of Power Dissipation in CMOS Combinational Circuits Using Boolean Function Manipulation. In: IEEE. Transaction on Computer-Aided Design, March, 1992, no. 11, pp. 373–383.

[Ghosh, 1992] A. Ghosh, S. Devadas, K. Keutzer, J. White. Estimation of Average Switching Activity in Combinational and Sequential Circuits. In: 29th ACM / IEEE Design Automation Conference. Tech. Dig, June 1992, pp. 253–259.

[Hsiao, 1997] M.S. Hsiao, E.M. Rudnick, and J.H. Patel. K2: An estimator for peak sustainable power of VLSI circuits. In: Proc. Intern. Symposium on Low Power Electronics and Design, Aug. 1997, pp. 178–183.

[Hsiao, 1999] M.S. Hsiao. Peak Power Estimation Using Genetic Spot Optimization for Large VLSI Circuits. Proc. Design, Automation and Test in Europe, 1999, pp. 175–179.

[Hsiao, 2000] M. S. Hsiao, E. M. Rudnick, and J. H. Patel. Peak power estimation of vlsi circuits: New peak power measures. IEEE Trans. on VLSI Systems, 2000, no. 8(4), pp. 435–439.

[Kang, 1986] S.M. Kang. Accurate simulation of power dissipation in VLSI circuit. In: IEEE Journal of Solid-State Circuits, 1986, vol. 21, no. 5, pp. 889–891.

[Kriplani, 1992] H. Kriplani, F. Najm, and I. Hajj. Maximum Current Estimation in CMOS Circuits. In: 29th ACM/IEEE Design Automation Conference, Anaheim, CA, June 8-12, 1992, pp. 2–7.

[Kumthekar, 1998] B. Kumthekar, E. Macii, M. Poncino, F. Somenzi. Simulation-Based Re-Synthesis of Sequential Circuits for Peak Sustainable Power Reduction. In: Proc. Intern. Workshop on Logic Synthesis. Lake Tahoe, CA, 7–10 June 1998, pp. 392–397.

[Liu, 2009] Yi-Ling Liu, Chun-Yao Wang, Yung-Chih Chen and Ya-Hsin Chang. A Novel ACO-based Pattern Generation for Peak Power Estimation in VLSI Circuits. In: 10th Intern. Symposium on Quality Electronic Design (ISQED 2009). San Jose, CA, 16–18 March 2009, pp. 317–323.

[Manne, 1995] S. Manne, A. Pardo, R. Bahar, G. Hachtel, F. Somenzi, E. Macii and M. Poncino. Computing the Maximum Power Cycles of a Sequential Circuit. In: Proc. of Design Automation Conf., 1995, pp. 23–28.

[Nagel, 1973] L.W. Nagel, D.O. Pederson. Spice (simulation program with integrated circuit emphasis). In: Technical Report UCB/ERL M382. EECS Department, Berkeley: University of California, April 1973.

[Najm, 1994] F.N. Najm. A survey of Power Estimation Techniques in VLSI Circuits. In: IEEE Trans. on VLSI, 1994, no. 12, pp. 446–455.

[SIA, 2014] Semiconductor Industry Association (SIA) ITRS Roadmap, available: http://www.sia-online.org/backgrounders_itrs.cfm.

[Wang1, 1996] C. Wang, K. Roy, T. Chou. Maximum Power Estimation for Sequential Circuits using a Test Generation-based Technique. In: Proc Custom Integrated Circuits Conf., 1996, pp. 229–232.

[Wang2, 1996] C.-Y. Wang and K. Roy. Maximum Current Estimation in CMOS Circuits Using Deterministic and Statistical Techniques. In: IEEE Trans. on VLSI Systems, 1998, no. 3, pp. 134–140.

[Wu, 2006] Q. Wu, Q. Qiu, M. Pedram. Estimation of Peak Power Dissipation in VLSI Circuits Using the Limiting Distributions of Extreme Order Statistics. In: IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2006, vol. 20, no. 8, pp. 942–956.

[Zimmermann, 1997] R. Zimmermann, W. Fichtner. Low-power logic styles: CMOS versus pass-transistor logic. In: IEEE Journal of Solid-State Circuits, July 1997, vol. 32(7), pp. 1079–1090.

Authors' Information

Liudmila Cheremisinova *– The United Institute of Informatics Problems of National Academy of Sciences of Belarus, principal, Surganov str., 6, Minsk, 220012, Belarus; e-mail: cld@newman.bas-net.by*

Major Fields of Scientific Research: Discrete mathematics, Logic design automation

ПАМЯТИ ЧЛЕН-КОРРЕСПОНДЕНТА НАН БЕЛАРУСИ АРКАДИЯ ДМИТРИЕВИЧА ЗАКРЕВСКОГО

24 февраля 2014 года ушел из жизни Аркадий Дмитриевич Закревский – член-корреспондент НАН Беларуси, академик Международной академии информации, информационных процессов и технологий, доктор технических наук, профессор, главный научный сотрудник Объединенного института проблем информатики НАН Беларуси.

Аркадий Дмитриевич Закревский – талантливейший ученый, выдающийся белорусский и советский кибернетик – специалист в области прикладной дискретной математики, информатики и логического проектирования. Он стоял у истоков рождения кибернетики, является основателем одной из самых известных школ логического проектирования в Советском Союзе и в мире. Его уход из жизни – невосполнимая потеря для белорусской науки.

А.Д. Закревский родился 22 мая 1928 г. в Ленинграде в семье служащих. Получив в Красноярском ремесленном училище связи профессию радиооператора, он, начиная с 1943 г., восемь лет работал радистом в изыскательских отрядах "Желдорпроекта" в районах Заполярья, Сибири, Монголии (Таймыр, Игарка, Норильск, р. Турухан, а затем Забайкалье, Чита, Монголия). В 1949 г. в г. Улан-Баторе (Монголия) сдает экстерном экзамены за курс средней школы и, получив аттестат зрелости, в 1951 г. поступает на физический факультет Томского госуниверситета. Через два года Аркадий Закревский переходит на только что организованный радиофизический факультет и с отличием оканчивает его по специальности

"физик-радиоэлектроник". В 1956 г. он поступает в аспирантуру к профессору В.Д. Кессениху, ведет преподавательскую работу в университете, подбирает группу студентов-энтузиастов и вместе с ними начинает заниматься новой (не только в стенах университета, но и в СССР) наукой – кибернетикой.

Узнав, что в г. Пензе начинается серийный выпуск ЭВМ Урал-1, А.Д. Закревский в 1957 г. добился выделения ее Томскому госуниверситету, а затем, критически оценив это последнее достижение вычислительной техники СССР, предлагает проект оригинальной вычислительной машины с той же производительностью, что и Урал-1 (который требовал для своего размещения зал в 100 кв.м), но всего на 18 триггерах. По этому проекту потом обучалось не одно поколение студентов-кибернетиков и не только в Томском госуниверситете, но и в Севастопольском приборостроительном институте. Затем разрабатывает логическое расширение ЭВМ Урал-1 (L-машину), на несколько порядков повышающее ее производительность при решении логических задач. Изготовление макета L-машины послужило темой дипломных работ студентов-кибернетиков первого выпуска в Томском госуниверситете.

С 1956 г. Аркадий Закревский проводил исследования в области технических приложений логики в Томском госуниверситете и Сибирском физико-техническом институте, руководил установкой первой в Сибири ЭВМ Урал-1 (1958 г.), заложил основы компьютерной дискретной математики, разработал технологию экспериментального статистического исследования алгоритмов на ЭВМ.

В 1960 г. Аркадий Закревский защищает диссертацию на соискание ученой степени кандидата физико-математических наук на тему "Матричный метод синтеза релейных схем". О высоком уровне диссертации красноречиво свидетельствует тот факт, что ее первый оппонент д.т.н., проф. А.Б. Сапожников, зачитав свой отзыв, предложил присвоить соискателю степень не кандидата, а сразу – доктора наук.

Оценив на основе собственного опыта трудоемкость программирования задач логического синтеза в машинных кодах, А. Закревский приходит к выводу о необходимости создания специализированного языка программирования, и в 1962 г. он разрабатывает Логический язык для представления алгоритмов синтеза релейных устройств – ЛЯПАС – первый отечественный язык программирования, ориентированный на решение логических задач и реализованный на ЭВМ.

Итогом плодотворной научной деятельности А. Закревского явилась его первая монография "Алгоритмический язык ЛЯПАС и автоматизация синтеза дискретных автоматов" (1966 г.), по которой он в 1967 г. защищает в Институте автоматики и телемеханики АН СССР (г. Москва) диссертацию на соискание ученой степени доктора технических наук. Для языка ЛЯПАС были разработаны системы программирования для разных типов ЭВМ, которые нашли широкое применение во многих организациях СССР, а также за рубежом: в Польше, Германии (ГДР и ФРГ), Чехословакии, Югославии, США. Интерес к языку ЛЯПАС за рубежом был вызван появлением перевода на английский язык сборника научных статей с описанием языка ЛЯПАС и алгоритмов синтеза дискретных автоматов, представленных на этом языке, – «LYaPAS, A Programming Language for Logic and Coding Algorithms» (Ed. by M. Gavrilov and A. Zakrevskij), ACM Monograph Series, New York, London, 1969.

В 1959–1971 гг. А.Д. Закревский – ассистент, старший научный сотрудник, заведующий лабораторией счетно-решающих устройств Сибирского физико-технического института, а затем профессор, заведующий кафедрой математической логики и программирования, которая была им организованна на радиофизическом факультете Томского госуниверситета. В 1971 г. он публикует широко известную в СССР фундаментальную монографию «Алгоритмы синтеза дискретных автоматов». Под его руководством создается первая в СССР система автоматического синтеза дискретных автоматов, принятая комиссией по кибернетике при Президиуме АН СССР и эксплуатировавшаяся в 1970–1980-е гг. на многих предприятиях министерств электронной промышленности и радиоэлектроники. Так было положено начало широкому фронту исследований в области компьютерной дискретной математики и

логического проектирования дискретных устройств и систем. Эти исследования органически влились в проблематику известной в то время школы М.А. Гаврилова по теории релейных схем и конечных автоматов, первое заседание которой состоялось в марте 1964 в Томске, а второе (в Комарово, вблизи Ленинграда) было целиком посвящено ЛЯПАСу.

В 1971 г. Аркадий Дмитриевич с группой сотрудников переезжает в Минск и организует в Институте технической кибернетики АН БССР лабораторию системного программирования и логического синтеза, впоследствии переименованную в лабораторию логического проектирования. В 1972 г. он избирается членом-корреспондентом АН БССР. Более двадцати лет А.Д. Закревский являлся заведующим организованной им лаборатории, передав затем управление лабораторией своему ученику, оставаясь главным научным сотрудником и научным руководителем множества научных и научно-технических тем и проектов.

Характерной чертой научного творчества А.Д. Закревского является сочетание широты охвата рассматриваемых проблем (вплоть до поиска аналогий в соседних областях) со строгостью и глубиной исследования (с предложениями практически эффективных методов их решения, доведенных до алгоритмической, а зачастую и программной реализации). Отличительной особенностью научной школы Аркадия Закревского стало развитие логико-комбинаторного подхода, основанного на формулировке задач проектирования в виде оптимизационных логико-комбинаторных задач на функциональных и структурных моделях объектов проектирования. Такой подход позволил развить теоретические основы кибернетики в самых разных областях. Трудно перечислить все, что было сделано Аркадием Дмитриевичем, поэтому ограничимся лишь кратким перечнем основных направлений кибернетики, где он оставил наиболее глубокий след:

1. *Логическая теория дискретных устройств* (применение помехоустойчивого кода Хемминга к синтезу надежных логических схем; визуально-матричный метод минимизации булевых функций; аппарат матричных логических уравнений для решения задач анализа, синтеза и диагностики неисправностей программируемых логических матриц; эффективные методы декомпозиции булевых функций и дискретных автоматов, кодирования внутренних состояний синхронного и асинхронного автоматов). А.Д. Закревским был предложен матричный аппарат для логического анализа, синтеза и диагностики дискретных устройств на базе программируемых логических матриц, результаты этих исследований представлены в его монографии «Логический синтез каскадных схем» (М.: Физматлит, 1980).
2. *Автоматизация программирования логических задач* (язык и системы программирования ЛЯПАС, эффективные для решения логико-комбинаторных задач и нашедшие широкое применение во многих организациях страны и за рубежом: в Польше, ГДР, Чехословакии, Югославии, ФРГ, США).
3. *Автоматизация логического проектирования* (эффективные методы минимизации и реализации полностью и слабо определенных булевых функций многих переменных, минимизации систем булевых функций в классе полиномов Жегалкина и Рида – Маллера, методы реализации систем многозначных частичных функций поляризованными полиномами Рида – Маллера, диагностирования константных неисправностей в EXOR-схемах, а также ряд систем автоматизированного логического проектирования, внедренных в производство (ЦКБ «Алмаз» (Москва), НИИЭВМ, НПО «Интеграл» и др.).
4. *Логические основы интеллектуальных систем* (методы решения больших систем логических уравнений, линейных и нелинейных; методы нахождения кратчайших решений неопределенных и несовместимых систем линейных логических уравнений; обобщение методов теории булевых функций на конечные предикаты; методы индуктивного и дедуктивного вывода в приложении к

распознаванию образов в пространствах бинарных и конечнозначных признаков, выявления импликативных закономерностей в этих пространствах; экспертные системы логического распознавания образов).

5. *Автоматизация проектирования систем логического управления* (язык описания параллельных алгоритмов логического управления ПРАЛУ; методы верификации, моделирования и реализации параллельных алгоритмов; понятия параллельного и секвенциального автоматов и основанные на них методы синтеза устройств логического управления; методы композиции и декомпозиции параллельных автоматов и описаний на языке ПРАЛУ).
6. *Комбинаторные задачи дискретной математики* (техника вычислений в булевом пространстве; комбинаторный базис логического проектирования – комплекс эффективных методов и программ решения комбинаторных задач над логическими матрицами и графами, имеющих множество полезных практических интерпретаций; техника вычислений в пространстве конечных предикатов, методы их минимизации и декомпозиции; высокоэффективные методы решения систем логических уравнений).

Результаты теоретических исследований А.Д. Закревского легли в основу ряда программных комплексов автоматизации решения логико-комбинаторных задач: логического проектирования дискретных управляющих устройств в базисе СБИС, программной и аппаратной реализации параллельных алгоритмов управления и экспертной системы логического распознавания.

Аркадий Дмитриевич, стоявший у истоков рождения кибернетики в Советском Союзе, свою научную деятельность успешно сочетал с педагогической. Он читал курсы лекций собственной разработки по математической логике, дискретной математике, теории вероятностей, теории автоматов, теории графов, теории булевых функций, программированию, методам логического проектирования и др. в ТГУ (Томск), БГУ и БГУИР (Минск). Созданная им научная школа логического проектирования первоначально зародилась в Томске, а затем стала действовать в Минске, Севастополе и Кишиневе. Его работы были известны не только в странах бывшего СССР, но и за рубежом, он руководил и входил в программные и организационные комитеты многих международных конференций и симпозиумов.

Талантливый организатор и руководитель, А.Д. Закревский подготовил 34 кандидата и восемь докторов технических и физико-математических наук. Он был неизменным руководителем постоянно действующего научного семинара ОИПИ НАН Беларуси по логическому проектированию и воспитывал своих учеников личным примером и тщательным рецензированием всех статей, ими написанных.

Результаты научных исследований А.Д. Закревского представлены более чем в 540 научных публикациях, в том числе в 26 монографиях (7 из них – в дальнем зарубежье), получивших широкую известность в стране и за рубежом. Он являлся одним из авторов книги "Наука Беларуси в XX столетии (2001 г.)". Под его научной редакцией было издано более 50 монографий и научных журналов, ежегодных сборников трудов по логическому проектированию и программированию (1975 – 2001 гг.), материалов научно-технических конференций (1975 – 1995 гг). Около 30 статей и заметок было издано о его жизни и научной деятельности.

Научную работу Аркадий Дмитриевич успешно совмещал с научно-организационной: являлся членом советов по защите диссертаций в ОИПИ НАН Беларуси, БГУ, Института электроники Латвийской АН (1980–1990); членом Ученого и Научного советов ОИПИ НАН Беларуси; членом редколлегий журналов "International Journal on Information Theories and Applications" (София, Болгария), "Управляющие системы и машины" (Киев, Украина), "Информатика" (ОИПИ НАН Беларуси, Минск), сборника "Проблемы защиты информации" (БГУ, Минск). А.Д. Закревский являлся участником важнейших научно-организационных мероприятий и событий в Институте и НАН Беларуси, связанных с развитием, направлениями и

повышением результативности научных исследований; председателем подкомиссии по автоматизации логического проектирования при совете по кибернетике Президиума АН СССР (1970– 1990 гг.). За научно-организационную деятельность А.Д. Закревский награжден медалями, Почетными грамотами Верховного Совета БССР (1978, 1988 гг.).

Аркадий Дмитриевич запомнится многим не только как выдающийся ученый, но и как опытный организатор туристских мероприятий (во время проведения школ-семинаров), пеших и лыжных однодневных походов в окрестностях г. Томска и Минска, руководитель многодневных горных и пеших походов (Тянь-Шань, Алтай, Прибайкалье, Беларусь), заядлый дачник, который на своей "фазенде" все сделал своими руками.

Вся жизнь Аркадия Дмитриевича была связана с наукой и была ей посвящена, он творил до последнего дня своей жизни: в 2013 г. в издательстве TUT Press им была опубликована монография на английском языке, готовилась к печати следующая монография, был подан доклад на конференцию…

А.Д. Закревский заслужил признание и авторитет не только как выдающийся исследователь и руководитель, но и как честный, справедливый и принципиальный человек. Он был наделен феноменальным талантом подводить теоретическую базу и находить элегантные решения для сложнейших задач дискретной математики, информатики и проектирования дискретных устройств. Если же говорить о гранях его характера, то главными из них являются трудолюбие; увлеченность научным поиском; стремление довести фундаментальные исследования до их воплощения в практике проектирования; твердость и бескомпромиссность. Талант и неустанный труд – истоки его достижений и величия как ученого и человека с большой буквы. В сердцах учеников и его последователей в области логического проектирования навсегда останется благодарная память об Аркадии Дмитриевиче – Учителе и Лидере.

П.Н. Бибило, Л.Д. Черемисинова

Список монографий Аркадия Дмитриевича Закревского

1. Алгоритмический язык ЛЯПАС и автоматизация синтеза дискретных автоматов. Томск,1966;
2. LYaPAS: A programming language for logic and coding algorithms. Academic Press, N.-Y., L., 1969 (with M. A. Gavrilov);
3. Алгоритмы синтеза дискретных автоматов. М.: Наука, 1971, 512 с.;
4. Логические уравнения. Мн.: Наука и техника, 1975, 96 с.;
5. Синтез асинхронных автоматов на ЭВМ. Мн.: Наука и техника, 1975, 184 с. (в соавторстве);
6. Система программирования ЛЯПАС-М. Мн.: Наука и техника, 1978, 220 с. (с Н.Р. Тороповым);
7. Логический синтез каскадных схем. М.: Наука, 1981, 416 с.;
8. Boolesche Gleichungen: Theorie, Anwendung, Algorithmen. VEB Verlag Technik, Berlin,1984, 304 s. (mit D. Bochmann und Ch. Posthoff);
9. Логика распознавания. Мн.: Наука и техника, 1988, 118 с.;
10. Параллельные алгоритмы логического управления. Мн.: Ин-т технической кибернетики АН Беларуси, 1999, 202 с.;
11. Полиномиальная реализация частичных булевых функций и систем. Мн: Ин-т технической кибернетики НАН Беларуси, 2001, 200 с. (с Н.Р. Тороповым);
12. Логика распознавания. Издание второе, дополненное. М., УРСС, 2003, 140 с.;
13. Параллельные алгоритмы логического управления. Издание второе, стереотипное. М.: УРСС, 2003, 200 с.;
14. Полиномиальная реализация частичных булевых функций и систем. Издание второе, стереотипное. М.: УРСС, 2003, 200 с. (с Н.Р. Тороповым);

15. Закревский А. Д. Логические уравнения. Издание второе, стереотипное. М.: УРСС, 2003, 95 с.

16. Основы логического проектирования. Книга 1. Комбинаторные алгоритмы дискретной математики. Мн: ОИПИ НАН Беларуси, 2004, 226 с. (с Ю.В. Поттосиным и Л.Д. Черемисиновой);

17. Основы логического проектирования. Книга 2. Оптимизация в булевом пространстве. Мн: ОИПИ НАН Беларуси, 2004, 240 с. (с Ю.В. Поттосиным и Л.Д. Черемисиновой);

18. Основы логического проектирования. Книга 3. Проектирование устройств логического управления. Минск: ОИПИ НАН Беларуси, 2006, 254 с. (с Ю.В. Поттосиным и Л.Д. Черемисиновой);

19. Логические основы проектирования дискретных устройств. М.: Физматлит, 2007, 589 с. (с Ю.В. Поттосиным и Л.Д. Черемисиновой);

20. Combinatorial algorithms of discrete mathematics. Tallinn: TUT Press, 2008, 192 p. (with Yu. Pottosin, L. Cheremisinova);

21. Optimization in Boolean space. Tallinn: TUT Press, 2009, 241 p. (with Yu. Pottosin, L. Cheremisinova);

22. Решение больших систем логических уравнений. Минск: ОИПИ НАН Беларуси, 2009, 96 с.;

23. Design of logical control devices. Tallinn: TUT Press, 2009, 304 p. (with Yu. Pottosin, L. Cheremisinova);

24. Вычисления в многомерном булевом пространстве. Минск: ОИПИ НАН Беларуси, 2011, 106 с.;

25. Combinatorial Calculations in Many-Dimensional Boolean Space. Tallinn: TUT Press, 2012, 121 p.;

26. Solving Large Systems Logical Equations. Tallinn: TUT Press, 2013, 114 p.

TABLE OF CONTENTS