International Conference

Big Data, Knowledge and Control Systems Engineering

BdKCSE'2018

Sofia, Bulgaria
21 November 2018

Institute of Information and Communication Technologies
- Bulgarian Academy of Sciences

John Atanasoff Society of Automatics and Infomatics

PROCEEDINGS

International Conference on

Big Data, Knowledge and Control Systems Engineering - BdKCSE'2018

21 November 2018
108 G. S. Rakovski Str., Hall 2, Floor 2, 1000 Sofia, Bulgaria

Institute of Information and Communication Technologies
of the Bulgarian Academy of Sciences
"John Atanasoff" Union on Automatics and Informatics, Bulgaria

Editor:
Rumen D. Andreev

Cover Designer: Ivan Panayotov
Prepress: Ivan Panayotov

Department of Communication Systems and Services
Institute of Information and Communication Technologies - Bulgarian Academy of Sciences
Bl. 2, Acad. G. Bonchev Str., 1113 Sofia, Bulgaria

Conference scope

The International Conference "Big Data, Knowledge and Control Systems Engineering" (BdKCSE'2018) aims to provide an open forum for the dissemination of the current research progress, innovative approaches and original research results on all aspects of Big Data Management, Technologies, and Applications. Organizer of the BdKCSE'2018 Conference is the Institute of Information and Communication Technologies of the Bulgarian Academy of Sciences, and co-organizer is the "John Atanasoff" Union of Automatics and Informatics, Bulgaria.

Big Data Management, Technologies, and Applications discuss the exponential growth of information size and the innovative methods for data capture, storage, sharing, and analysis. Modern technologies continue to become more complex as do the applications. The integration of technologies, complex relationships of applications and the accelerated technological changes are new challenges to technology management.

Topics such as product development, innovation management, and research and development management have become very popular. Big data spans dimensions as volume, variety, velocity, volatility and veracity, steered towards one critical destination – value. Following from these, the conference is devoted toward improving the understanding, systems engineering, human cognition and modeling, and data.

The conference will help the research community identify the novel important contributions and opportunities for recent research on the different intelligent methodologies and techniques in the field.

Table of contents

Organized by:

Institute of Information and Communication Technologies
- Bulgarian Academy of Sciences

"John Atanasoff" Union on Automatics and Informatics, Bulgaria

Our conference will take place at the
Federation of the Scientific Engineering Unions /**FNTS**/, Bulgaria

Program committee

Conference Chairs

▪ Chairman	Assoc. Prof. Rumen Andreev	Bulgarian Academy of Sciences
▪ Vice chairman	Assoc. Prof. Lyubka Doukovska	Bulgarian Academy of Sciences
▪ Vice chairman	Assoc. Prof. Yuri Pavlov	Bulgarian Academy of Sciences

Program Committee

▪ Abdel-Badeeh Salem	Ain Sham University	Egypt
▪ Chen Song Xi	Iowa State University	USA
▪ Dimiter Velev	University of National and World Economy	Bulgaria
▪ František Čapkovič	Slovak Academy of Sciences	Slovakia
▪ George Boustras	European University	Cyprus
▪ Georgi Mengov	University of Sofia	Bulgaria
▪ Jacques Richalet		France
▪ John Wang	Montclair State University	USA
▪ Kosta Boshnakov	University of Chemical Technology and Metallurgy	Bulgaria
▪ Krasen Stanchev	University of Sofia	Bulgaria
▪ Ljubomir Jacić	Technical College Požarevac	Serbia
▪ Ljupco Kocarev	Macedonian Academy of Sciences and Arts	Macedonia
▪ Milan Zorman	University of Maribor	Slovenia
▪ Neeli R. Prasad	Aalborg University, Princeton	USA
▪ Olexandr Kuzemin	Kharkov National University of Radio Electronics	Ukraine
	German Academic Exchange Service	Bonn
	North Rhine-Westphalia	Germany
▪ Peđa Milosavljević	University of Niš	Serbia
▪ Peter Kokol	University of Maribor	Slovenia
▪ Petko Ruskov	University of Sofia	Bulgaria
▪ Radoslav Pavlov	IMI, Bulgarian Academy of Sciences	Bulgaria
▪ Rumen Nikolov	UniBIT-Sofia	Bulgaria
▪ Song II-Yeol	Drexel University	USA
▪ Sotir Sotirov	University “Prof. Asen Zlatarov”	Bulgaria
▪ Tatiana V. Atanasova	IICT, Bulgarian Academy of Sciences	Bulgaria
▪ Tomoko Saiki	Tokyo Institute of Technology	Japan
▪ Uğur Avdan	Anadolu University	Turkey
▪ Valentina Terzieva	IICT, Bulgarian Academy of Sciences	Bulgaria
▪ Valeriy Perminov	National Research Tomsk Polytechnic University	Russia
▪ Vera Angelova	IICT, Bulgarian Academy of Sciences	Bulgaria
▪ Vyacheslav Lyashenko	Kharkov National University of Radio Electronics	Ukraine
▪ Wojciech Piotrowicz	University of Oxford	UK
▪ Zlatogor Minchev	IICT, Bulgarian Academy of Sciences	Bulgaria
▪ Zlatolilia Ilcheva	IICT, Bulgarian Academy of Sciences	Bulgaria

Publisher:
"John Atanasoff" Union on Automatics and Informatics, Bulgaria

108 G. S. Rakovski Str., 1000 Sofia, Bulgaria
Phone: +3592 987 6169
e-mail: sai.bg.office@gmail.com
www.sai.bg

ISSN: 2367 - 6450

Forest Ecological Management and Renewable Timber Production: Utility Approach

Yuri Pavlov Pavlov
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences,
1113 Sofia, Bulgaria
yupavlov15@isdip.bas.bg

Abstract: In the paper is demonstrated a system engineering value driven approach within modeling of forest resource management and timber renewable production. A multi-attribute utility function is evaluated and used as objective function in the frame of an optimal control timber renewable production problem. The optimal control solution is based on the individual consumers' preferences represented as utility objective function.

Keywords: Preferences, Utility, Optimal Control, Renewable Timber Production, Ecology

1 Introduction

The use of natural resources from a contemporary point of view requires a comprehensive assessment mathematical approach that includes factors such as economic efficiency, social effect, environmental impact, etc. In the paper we will look at the case of logging management in conditions of renewable natural resources, taking into account the economic efficiency and the social effect for the population (Clark, Clarke, Munro, 1979). This problem in terms of non-renewable resources has been analyzed in previous developments based on the model of Hotelling and the solution proposed by Clarke in the international conference BdKCSE-2017 (Hotelling, 1931; Clarke, 1983). A multi-attribute utility function has been built, taking into account three factors, economic efficiency as the quantity of timber harvested, the social effect as resources invested in production and ecological state of the forest (Lyubenova and all, 2015). We will use this utility function as an example for objective function in this study. Generally, for each problem a corresponding utility function have to be built (Keeney & Raiffa, 1993).

2 Utility Function Determination

In a previous investigation aiming planning of new forests in an environmentally sound way with minimal negative effects on the environment and economically effective resource management of forest timber was determined the following utility function (Lyubenova and all, 2015). The expert analysis and the structuring carried out in (Lyubenova and all, 2015) led to accepting the following sub-objectives and the appropriate criteria, which adequately describe the main objective and are real, physically measured quantities. The domain of variation of representing criteria is determined as follows:

- X_1-factor (material services as volume of timber per hectare- economic effect) [10-300 $m^3.ha^{-1}$];
- X_2- factor (regulating and supporting services - ecological effect) [1-200 number of species per hectare];
- X_3- factor (percentage of employed locals in the forestry sector - social effect) [1-30%].

Independence by utility was found by the decision maker (DM) between the following factors:

- X_2 from X_1; X_2 from X_3;
- X_3 from X_1; X_3 from X_2;

The DM's preferences for x2 at different values of x1 and x3 do not change, suggesting utility independence of x2 from the changes of other two factors. Whatever the reserves of wood in the forest ecosystem are the employment of the population in the forestry sector may be different, but in any case the preferences are directed to the presence of a large species richness of the forest, i.e. great variety of species of trees, grasses, moss, lichen etc. that form the ecosystem and ensure its greater stability.

The lottery preferences of DM for x3 at different values of x1 and x2 do not change, suggesting utility independence of x3 from the changes of other two factors. This means that whatever the reserves of wood are and at different species richness of the forest ecosystem, in any case, the preferences are aimed at the increasing of number of workers in the forestry sector. Using the theory for decomposition of multi-attribute utility to simpler functions (Keeney & Raiffa, 1993) the following multi-attribute utility structure is determined:

$$U(X_1; X_2; X_3) = k_1 U(X_1; X^o_2; X^o_3) +$$
$$+ f_2(X_1) \times [U(X^o_1; X_2; X^o_3)] + f_3(X_1) \times [U(X^o_1; X^o_2; X_3)] +$$
$$+ f_{23}(X_1) \times [U(X^o_1; X_2; X^o_3)] \times [U(X^o_1; X^o_2; X_3)],$$
$$\text{where } U(X^o_1; X^o_2; X^o_3) = 0 \text{ and } U(X^*_1; X^*_2; X^*_3) = 1.$$

In the formula above $X^o=(X^o_1; X^o_2; X^o_3)=(10,1,1)$ and $X^*=(X^*_1; X^*_2; X^*_3)=$ (300,200,30). The functions f_2, f_3 and f_{23} have the forms:

$$f_2(X_1) = U(X_1; X^*_2; X^o_3) - k_1 U(X_1; X^o_2; X^o_3),$$
$$f_3(X_1) = U(X_1; X^o_2; X^*_3) - k_1 U(X_1; X^o_2; X^o_3),$$
$$f_{23}(X_1) = U(X_1; X^*_2; X^*_3) - f_2(X_1) - U(X_1; X^o_2; X^*_3).$$

Each of these sixth functions was evaluated based on the DM's preferences (Pavlov, Andreev, 2013). The described value modeling is part of the Smart Forest Ecological Management System ForEco which is based on Open Data and FIWARE (Lyubenova et al, 2015). The graphical appearance of utility functions at fixed eco parameters (132 numbers of species per hectare) can be seen in the following figure:

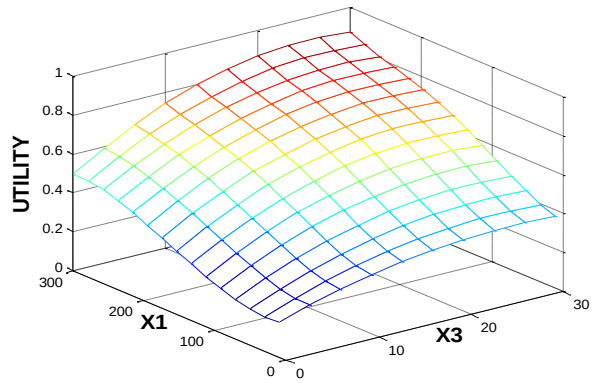

Figure 1. Utility function U(x_1,132,x_3)

This utility model could be perceived as a part of mathematical description of the social and ecological human expectations in the frame of the main objective of the complex production problem. The described model was included as a part of the Smart Forest Ecological Management System ForEco which is based on Open Data and FIWARE.

3 Dynamic Model and Effective Resource Management

In this study we will assume that the logging resources are renewable and are described by a mathematical model of the type:

$$\dot{y} = F(y) - qyK, \quad F(y) = r(y(\theta - y)).$$

In the above formula q, r, and θ are constants (q=0.7, r=1.1, θ=1). With (y) is noted the available renewable natural resource (timber per hectare-X_1), and with K we mention the funds invested in the logging (employed locals in the forestry sector-X_3). Such models are used in

other areas of activity where the resources are renewable (Clark, Munro, 1975). It is seen that equilibrium in the system occurs when the derived quantity of extracted wood (the derivative of y) is zero. In the survey (Clark, Munro, 1975; Clark, Clarke, Munro, 1979) the optimization criterion is the maximum profit. This does not, however, take into account the social impact of production on the population as well as the eco conditions at the moment. These factors could be reported by including in a integral (objective function) a multi-attribute utility function and aiming it at the maximum. The goal is to maximize the integral:

$$J = \int_0^T U(y,K)(py - c)Kdt \cdot$$

In this formula U(y,k) is the utility U(X_1; 132; X_3), p and c are constants (p=1.3, c=0.31), and T is the end of the production period. The task thus defined is a optimal control problem. A solution based on the Pontryagin's principle of maximum may be sought (Clarke, 1983; Gabasov, Kirilova 1981). The necessary condition for optimality is the following:

$$H(T,\psi,y,K) = \max_k [\psi(F(y) - qyK) + U(y,K)(py - c)K], 0 \le K \le K_{max} \cdot$$

For autonomous systems, as in this case, is performed:

$$\max_K H(T,\psi,y,K) = h, \quad h \ is \ cte.$$

The type of optimal control is determined by the following manifold:

$$\Phi(y,K,\psi) = (U(y,K)(py - c) - \psi qy)K] = 0.$$

The optimal control could be determined by the manifold as follows:

$$\Phi(y,K,\psi) = (U(y,K)(py - c) - \psi qy)K],$$
$$\Phi(y,K,\psi) > 0 \Rightarrow K = K_{max}, \quad \Phi(y,K,\psi) < 0 \Rightarrow K = 0.$$

Thus defined the optimal control problem defines "chattering control" or "bang-bang control". Obviously in the logging for a period of several years this is unacceptable. Such management could be acceptable in fisheries or in agriculture for some renewable resources. For idea a example of "chattering control" in biotechnology is shown in Figure 2.

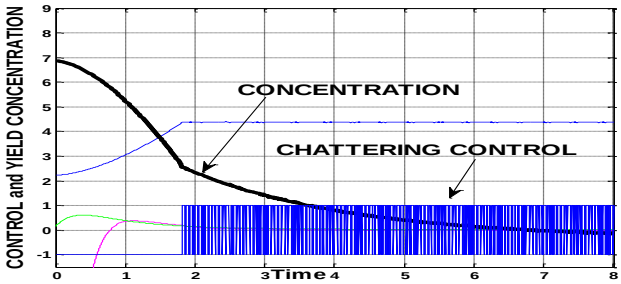

Figure 2. **Chattering control and Sliding manifold (concentration)**

The second possibility is to explore the manifold $\Phi(y,K,\psi)=0$ and determine the renewable natural resource y as function of the input K and hence the derivative of y. This control guarantees a sliding mode on the manifold. The sliding on the manifold will move the system to an optimal equilibrium point or an optimal periodic mode, depending on the type of the used utility function.

By the use of the control K as input fed-back control and by the use of the differential equation of the timber renewable production the vector state of the claused autonomous system will remain on the manifold. The beginning point will determine the moment, when the system across the manifold. After this moment the system stay on the manifold. In this investigation the control K is included in a utility function and this creates a problem. Such calculations could determine a huge formula for the input K.

However, the availability of a multi-attribute utility function allows another possibility, another "good" solution with simpler calculations. In the literary sources are discussed steady states that are optimal for economic reasons (Clark, Munro, 1975; Clark, Clarke, Munro, 1979). We can approach towards the problem by the same way and look for a state of equilibrium that sets, that determines the maximum of the integral criterion J. The maximum of the following function determines a equilibrium point as steady state possibility:

$\Phi(y)=U(y,F(y)/qy)(py-c)F(y)/qy$.

The K control is replaced with the value that is determined by the differential equation which described the renewable natural resource y (volume of timber per hectare), assuming the derivative to zero. The utility function and the optimal point (y*, K*) are shown in Figure 2.

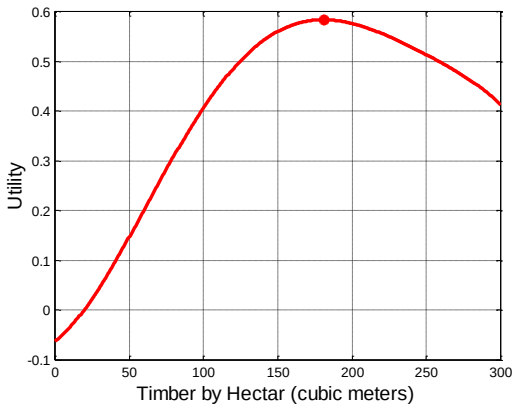

Figure 3. **Timber per hectare versus *U*(y, F(y)/qy)**

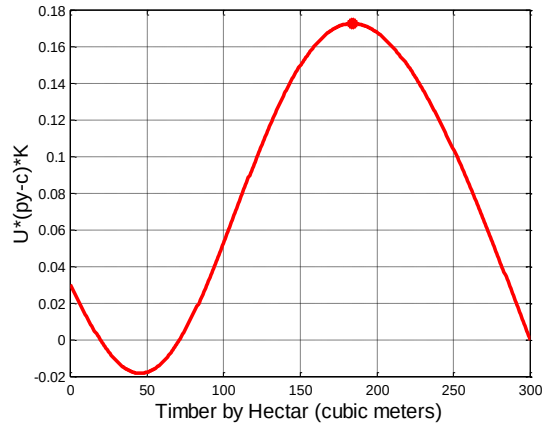

Figure 4. **Timber per hectare versus Φ(y)**

In the Figure 4 is shown the place of the equilibrium point (y*, K*) on the utility function. The location of this optimal point on the utility function is marked with stars.

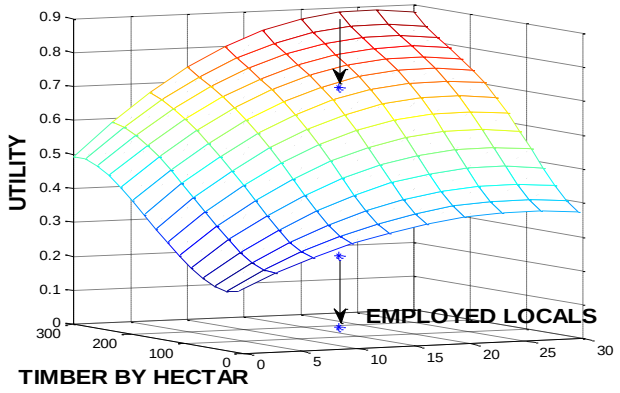

Figure 5. **Equilibrium point**

The formula that describe the dynamic of y determines the minimum time to attain the equilibrium, starting from any initial point (y_n,K_n). If y_n is less than y* ($y_n < y^*$), it is obviously necessary to let nature return to state y*. This determines K=0. Conversely, if y_n is greater than y* ($y_n > y^*$), then the control input is maximum $K = K_{max}$ until the equilibrium point (y*,K*) is reached.

In this way, good conditions for production can be defined, taking into account social and economic effects without relatively complex calculations. Of course, the optimal solution generally requires a mathematical study in details of the manifold $\Phi(y,K,\psi)=0$ described above.

4 Conclusions

In the paper is analyzed the optimal timber production in the conditions of renewable resources, taking into account the social and economic effects at the same time. A solution has been sought as a balance between natural resources and the economic benefit that provides maximum comfort for the decision-maker (the population). After building the decision maker's utility function, the determination of the optimal conditions is easy and straightforward. Here is a mathematical toolkit that allows the study of complex multi-factorial tasks in the production conditions with renewable resources. This toolbox would generally allow a mathematical description and find an optimal or sub-optimal solution.

The sequence of the study could be as follows. First it addresses the main purpose (the main objective) of the study and the factors that determine it. Second, designing utility functions as optimal criteria or as objective function or as a part of the model: determining a mathematical model and its parameters and finally optimizing and optimal solution at best.

References

[1] Keeney, R. & Raiffa H. (1993). *Decision with Multiple Objectives: Preferences and Value Trade-offs*, Cambridge & New York, USA: Cambridge University Press.

[2] Hotelling, H. (1931). The economics of exhaustible resources. *J. Political Economy*, 39, 137-175.

[3] Clark, C. W., and G. Munro. (1975). The Economics of Fishing and Modern Capital Theory: A Simplified Approach. *Journal of Environmental Economics and Management*, 2 (2): 92–106.

[4] Clark, C. W., F. H. Clarke, and G. Munro. (1979). The Optimal Management of Renewable Resource Stocks: Problems of Irreversible Investment. *Econometrica* 47(1):25–47.

[5] Clarke, F. (1983). *Optimization and Non-smooth Analysis*, Wiley, New York.

[6] Pavlov, Y. & Andreev, R. (2013). *Decision control, management, and support in adaptive and complex systems: Quantitative models*, Hershey, PA: IGI Global.

[7] *Gabasov*, R. & *Kirilova* F. M. (1981). Methods of *Optimization*, Nauka, Minsk,.

[8] Lyubenova M., Pavlov, Y., Chikalanov, A., Spassov, K.,· Novakova, G.,·Nikolov, R. (2015). Smart Forest Ecological Management System, *Journal of Balkan Ecology*, Vol. 18 (4), 363-375.

Activation Function Permutation for Multilayer Perceptron Training

Todor Balabanov, Tatiana Atanasova, Ivan Blagoev
Institute of Information and Communication Technologies - Bulgarian Academy of Sciences
Acad. Georgi Bonchev Str., block 2, office 514, 1113 Sofia, Bulgaria
todorb@iinf.bas.bg, atanasova@iit.bas.bg, i.blagoev@iit.bas.bg

Abstract: Artificial neural networks are very popular tool in the field of machine learning. Their common usage is in the fields of classification and forecasting. Artificial neural networks were developed as mathematical explanation of the biological neural networks, but their application is still far away from the live organisms. Artificial neural networks have two common working phases. They are first trained and after that used for solving of the specific task. Numerous studies have shown that a well-trained network can solve their tasks very efficiently. Training is a key stone, because it is time consuming even on very powerful computers. This research proposes permutation of the neurons activation function in order the convergence speed-up to be achieved.

Keywords: artificial neural networks, back propagation, machine learning.

1 Introduction

Artificial neural networks (ANN) are profoundly used in different scientific and daily life tasks. As a data handling tool they are very well known from decades [1, 2]. Generally artificial neural networks are weighted directed graph and there are many different configurations of this scheme. In the simplest case it is a multilayer perceptron, but there are much more exotic architectures as generalized nets [3], for example.

In the literature many different training algorithms are presented, but the back-propagation of errors to calculate a gradient that is needed to define weights in multilayer perceptron is the most popular one [4]. The back-propagation is an exact numerical method that is why it has a disadvantage of trapping in local optimums [5]. By using stochastic or hybrid training algorithms there are many attempts to fix this disadvantage [6, 7].

This research proposes activation function permutation as modification of back-propagation training algorithm, used for multilayer perceptron. By the proposed modification the train-

ing speed-up is achieved. The paper is organized as follows: Section 1 gives a brief introduction of the problem. Section 2 explains the proposed modification; Section 3 presents some experiments and results. Section 4 discusses the achieved results and proposes directions for some further investigations.

2 Neurons Activation Function Permutation

Permutation of the neurons in the hidden layer of a multilayer perceptron was proposed by Zankinski [8]. The efficiency of such neurons permutation is arguable, but it is perfect creativity direction for proposal of other permutation based ideas.

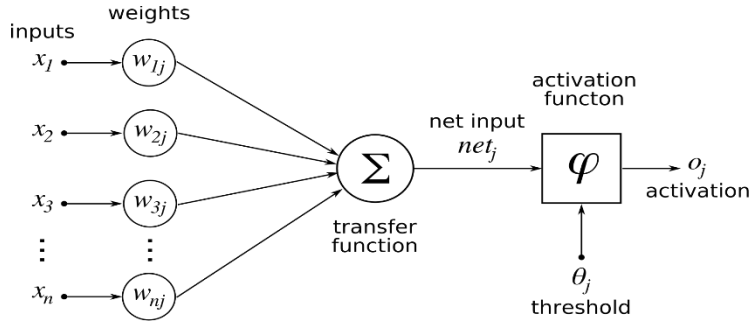

Figure 1. Artificial Neuron Components

Each neuron in an artificial neural network has weights attached to it, transfer function (in most cases a sum of the weighted input signals) and activation function (Fig. 1).

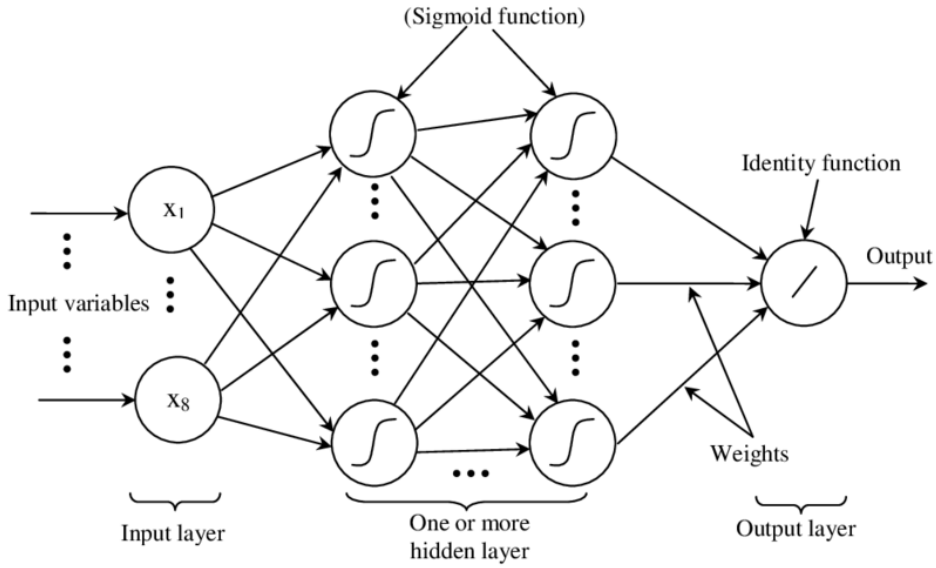

Figure 2. Multilayer Perceptron

The neurons within the topology of multilayer perceptron network are organized in layers (Fig. 2). The input layer has only feeding duties, but neurons in the hidden layer have to transform the signals inside the network. The transfer function collects input signals with particular weighting multiplication. Summation is the most used transfer function. When signals are collected then normalization should be applied in order a single signal to be emitted by the neuron. For this normalization an activation function is used. Different mathematical functions can be used for the activation [4], but the most widespread are hyperbolic tangent and sigmoid. The role of activation function is very important, because the amount of input links to a single neuron can vary from few to hundreds.

$$z = f(x \cdot w) = f\left(\sum_{i=1}^{n} x_i w_i\right)$$

$$x \in d_{1\times n},\ w \in d_{n\times 1},\ z \in d_{1\times 1}$$

Figure 3. Artificial Neuron Transfer Function

The weights themselves are among another important factor for neural network training. In classical multilayer perceptron weights are unconstrained real numbers. It means that big negative and positive values can be involved in the regular artificial neural network working mode (Fig. 3). Summation of multiplications between input signals and weights in large range is problematic, because different neurons will not participate in the network on an equal basis. All these problems are solved by normalization with proper activation function.

In this research a network with initial topology of three layers and hyperbolic tangent as activation function for each neuron is used. Several neurons are selected on a random basis and their activation functions are changed from hyperbolic tangent to sigmoid function for a single cycle of training. Thus permutation of the activation functions is achieved.

`Encog` Machine Learning Framework [https://www.heatonresearch.com/encog/] is used for experimenting and changing of the activation function during training phase of artificial neural network operation.

3 Experiments & Results

Experiments are done with custom made software solution, based on `Encog` programming library. The source code is under open source software license and it can be found at [9]. Digits charters recognition was selected as benchmark dataset. Model of multilayer perceptron with 256-266-10 neurons on input-hidden-output layers was programmed in Java. Hidden

layer was selected to be bigger in order better differentiation to be achieved. Dataset of ten digits is presented as `CSV` file.

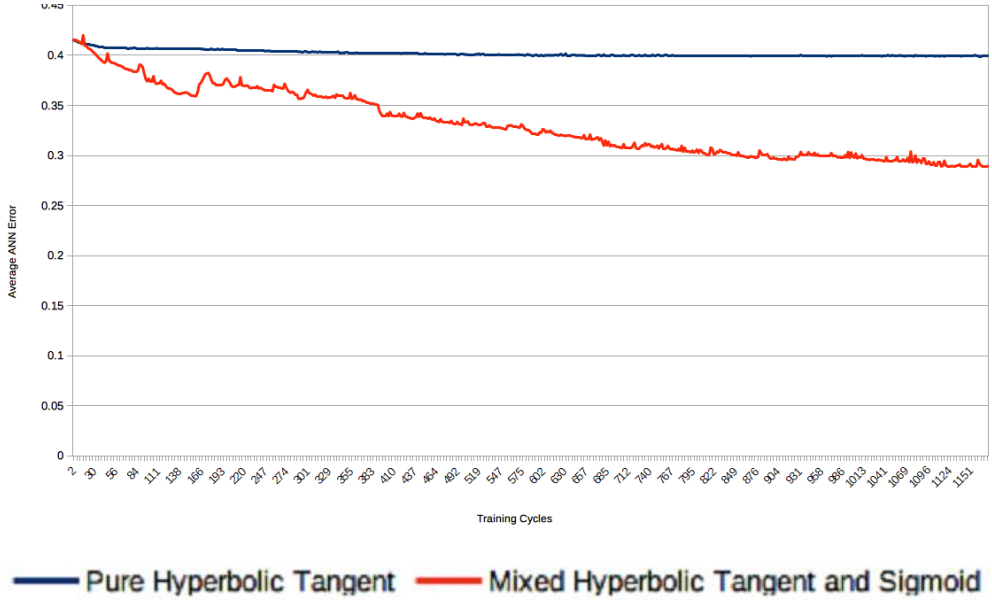

Figure 4. Training Convergence Comparison

Two artificial neural networks are tested for comparison. One network has only pure hyperbolic tangent activation functions and other network has mix of both two kinds (hyperbolic tangent and sigmoid). Experiments are conducted 30 times and their average results are presented in Fig. 4, where an ANN average total error is shown during training cycles. According the received results the mix of activation functions gives better convergence results. Swaps of the activation function give stairs like shape of the convergence curve.

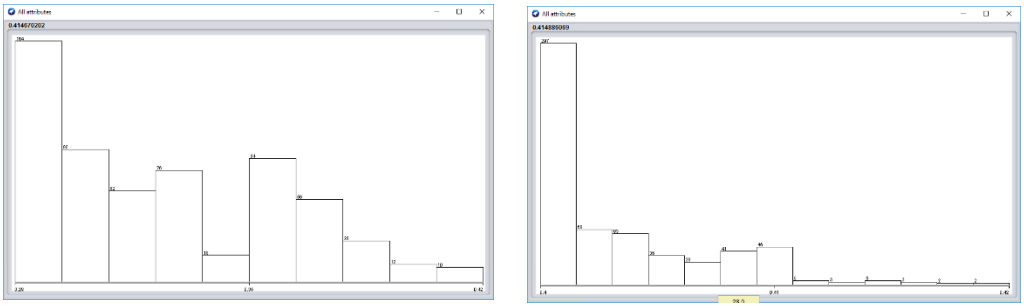

Figure 5. The error distribution for two kinds of ANN:
left – mixed activation functions; right - pure hyperbolic tangent

4 Conclusions

The paper proposes permutation of ANN activation functions as modification of the back-propagation training algorithm. It is clearly shown by the presented results that the proposal is efficient enough in artificial neural network training speed-up.

As further work it would be interesting to extend the proposed modification in the direction of investigation of other neurons activation functions [4]. Training of artificial neural networks in distributed environment [6, 7, 8] would be also interesting if it is combined with the proposed modification.

Acknowledgements

This work was supported by private funding of Velbazhd Software LLC.

References

[1] Atanasova, T., Barova, M., Balabanov, T., Using neural models to analyze time series in large volumes of data, Int. Conf. NVU "Vasil Levski", 193--198, ISSN:1314-1937, (2016)

[2] Tomov, P., Monov, V., Artificial Neural Networks and Differential Evolution Used for Time Series Forecasting in Distributed Environment, Proceedings of International conference Automatics and Informatics, ISSN 1313-1850, 129--132, Sofia, Bulgaria, (2016)

[3] Tashev, T., Hristov, H., Modeling of synthesis of information processes with generalized nets. In: Drinov, M. (ed.) Cybernetics and Information Technologies, vol. 2, Academic Publishing House, Sofia, 92–104, (2003)

[4] Zankinski, I., Tomov, P., Balabanov, T., Alternative Activation Function Derivative in Artificial Neural Networks, Proceedings of XXV International Symposium Management of Energy, Industrial and Environmental Systems, ISSN 1313-2237, Bankya, Bulgaria, 79--81, (2017)

[5] Balabanov, T., Avoiding Local Optimums in Distributed Population based Heuristic Algorithms (in Bulgarian), Proceedings of XXIII International Symposium Management of energy, industrial and environmental systems, John Atanasoff Union of Automation and Informatics, Sofia, Bulgaria, 83–86, (2015)

[6] Balabanov, T., Zankinski, I., Dobrinkova, D., Time Series Prediction by Artificial Neural Networks and Differential Evolution in Distributed Environment, International Conference on Large-Scale Scientific Computing, Sozopol, Bulgaria, Lecture Notes in Computer Science, vol. 7116, no. 1, 198–205, (2011)

[7] Balabanov, T., Zankinski, I., Barova, M., Strategy for Individuals Distribution by Incident Nodes Participation in Star Topology of Distributed Evolutionary Algorithms, Cybernetics and Information Technologies, Sofia, Bulgaria, vol. 16, no. 1, 80–88, (2016)

[8] Zankinski, I., Stoilov, T., Effect of the Neuron Permutation Problem on Training Artificial Neural Networks with Genetic Algorithms in Distributed Computing, Proceedings of XXIV International Symposium Management of Energy, Industrial and Environmental Systems, ISSN 1313-2237, Bankya, Bulgaria, 53--55, (2016)

[9] Balabanov, T., Encog digits classification resilient training example with regulated sin as activation functions, https://github.com/TodorBalabanov/EncogDigitsResilient, (2018)

Q-Learning Based Model of Node Transmission Power Management in WSN

A. Alexandrov, V. Monov
Institute of Information and Communication Technologies
Acad. G.Bonchev bl. 2, 1113, Sofia, Bulgaria
akalexandrov@iit.bas.bg, vmonov@iit.bas.bg

Abstract: The scope of this work is to investigate how machine learning may be used to help the WSN's nodes to manage their transmission power and to improve the internode wireless communications. The optimized transmission power has benefits in terms of WSN energy consumption and RF interlink interference.

We propose an adaptive model of a wireless sensor node based on Multi-Layer Perceptron (MLP) network representation and machine learning. The presented in the paper approach uses the Q-Learning which as a form of reinforcement machine learning for node Transmission Power Management (TPM) process. This inspires many practical solutions that maximize resource utilization and prolong the shelf life of the battery-powered wireless sensor networks.

Keywords: ANN, neural networks, wireless sensor network; transmission power control; Q-Learning; energy efficiency; quality of service

1 Introduction

The internode communication is usually the most energy consuming event in Wireless Sensor Networks (WSNs). One way to significantly reduce energy consumption is by applying an adaptive transmission power management techniques. This approach dynamically adjusts the transmission power in which depends on factors as wireless link Quality of Service (QoS) and the wireless node Received Signal Strength (RSS) value.

As is illustrated on the Figure 1 the reliable connection between sensor nodes depends on the distance between nodes, received signal strength and the level of the existing RF noise. In a real environment, the deviation between the needed transmission power for reliable communication between WSN nodes at one and the same distance can reach dramatically high values because of the mentioned above factors.

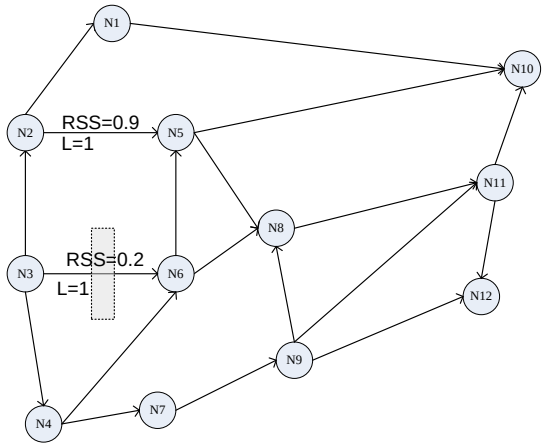

Figure 1. Wireless sensor network with RF barrier between Note 3 and Node 6

The task of the WSN power management (WSN-PM) stays more complex when the propagation of the RF transmission signal is influenced by the factors which are time changeable as cyclic sources of RF noise, interference RF sources with variable sizes and etc.

Therefore, one of the possible ways to solve the complexity problem is to use an adaptive model of a wireless sensor node based on a self-learning Artificial Neural Network (ANN). The term machine learning is a process to improve the performance of a system based on its past experiences. The method takes place when the problem is too complicated to be solved in real time, or in case that is not impossible the problem to be solved in a classical way. One of the machine learning methods is reinforcement learning. The method of reinforcement learning uses an agent executor environment desired by different methods and is based on the concept of reward.

Q-Learning is a form of reinforcement learning and it was proposed in 1989 by C. J. Watkins [1], [2]. This method is based on three main functions, an evaluation function, a function of strengthening and reinforcing function.

Q-learning involves two main entities: an agent and the environment (Figure 2). The agent plays as a learner and decision-maker at the same time, while the environment is an unpredictable and unknown which influences the agent's performance.

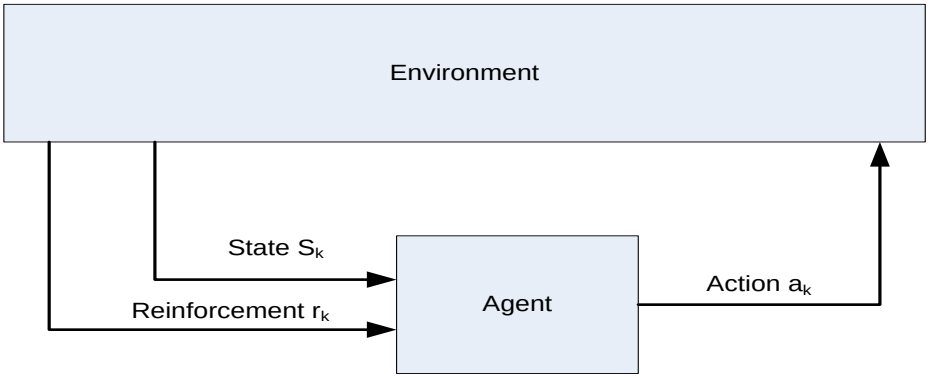

Figure 2. The interaction between agent and environment

Where:

S_k - represents the status of the environment

a_k - actions – decisions of the Agent

r_k - feedback called reward which evaluates the effect of the actions a_k

2 Related work

To overcome the disadvantages of the proactive and reactive techniques, machine learning represents an attractive solution [3] to reach a defined goal by learning the dynamics of the WSNs [4,5], predicting and adapting the transmission power values in different conditions. The objective is making WSNs autonomous without the intervention of developers and users to set the transmission power. To the best of our knowledge, only a few contributions have applied machine learning in TPC, mainly Reinforcement Learning (Q-LEARNING) and fuzzy logic [6].

Q-LEARNING in WSNs has been used in the literature but mainly for path selection in routing protocols and sleeping techniques, maintaining constant learning factors [7–11]. Static values would either bring the system slowly to convergence or make the system too reactive if the learning factor is constantly low or high respectively.

3 Proposed Q-Learning model and design

In the current paper, we represent the WSN as MLP type of network. Every wireless node can be represented as perceptron consisting of four components, i.e. inputs, weights, activation function and output. The generalized architecture of an MLP perceptron is given in Figure 3.

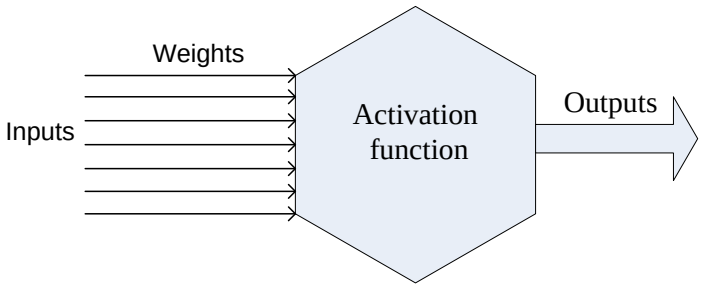

Figure 3. Generalized MLP perceptron

In figure 3, different inputs of different weights are captured, and if some weights match threshold values, it may activate the function to generate an output. The MLP architecture consists of three types of layers such as input, hidden and output [12]. Our approach is utilizing unsupervised back propagation based learning in which threshold values are used as the activation function. Back propagation methodology is used. If output values do not matches, it can be adjusted. The proposed model operates in three phases:

Data collection – in this stage, every wireless sensor node keeps track of the packets received from neighbors, packets forwarded to neighbors and average energy consumption in a specific period of time. This phase may take time to get few optimal values of the system usage over a specific period of time.

Learning stage – in this stage, MLP is trained to identify different communication sources which are located in its environment. In this stage, the sensor node learns the parameters of packets which the neighbor wireless node receives and sending over a specific time and the related energy consumption. The learning process is based on the Q-learning method as a form of reinforcement learning.

Results – In this stage, the accuracy of the learning process is reviewed. The system is capable to calculate the needed transmission energy amount depend on the current topology of the neighbor nodes.

The proposed MLP model of the wireless sensor node is shown in Figure 4.

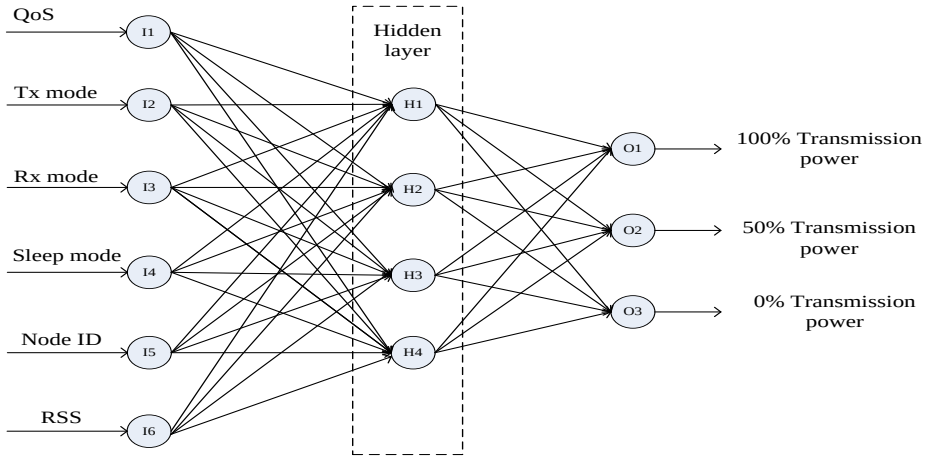

Figure 4. MLP model with a single hidden layer of the wireless sensor node

The amount of time a node remains inactive, or sleep mode. In the active mode, the node fall in the following active patterns: 100% transmit power, 50% transmit power and 0% transmit power. The training phase takes a time which depends on the needed accuracy of the system.

The proposed mechanism takes as input the mode as well as the energy consumption in a specific interval of time. The captured weights are multiplied and processed in the hidden layer. The three outputs are 100% Power transmission, 50% Power transmission and No transmission (0% Power transmission). The working mechanism of the proposed model is shown in Figure 5.

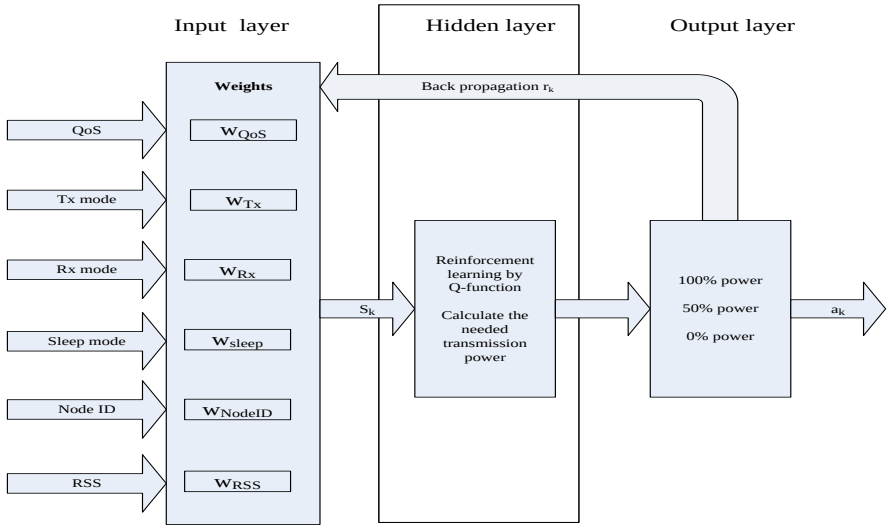

Figure 5. The working mechanism of the proposed model

Based on the Q-learning function definition [2], we have:

$Q_{k} = r_k (s_k, a_k) + K \max Q(S_k, a_{k,} w_k)$ $Q_k \in (0, 0.5, 1)$ (1)

Where:

r_k – reinforcement value

a_{k-} action state

K - discount factor $0 \leq K \leq 1$

w_k – weights in stage k

s_k - state of the environment in stage k

$\max Q(s_k, a_{k,} w)$ - is the function value state/action

In our case the learning process starts with preliminary fixed combination of weights (w_k), s_k=0 and a_k=0, forming the initial function value state/action $Q(s_k, a_{k,} w)$. During the iterations stages of the learning process Q-function generates a corrective signal and send it to the input of the system. The reinforcement value r_k provides for every new state S a signal that can be reward or punisment. In our case r_k takes values -1, 0 or 1. The discount factor K role is to mark the importance of the future reward. In our case K takes value 0 or 1.

The learning phase has a typical length of 100-150 hours and depends of the needed accuracy of the wireless system. The captured weights (w_{QoS}, w_{Tx}, w_{Rx}, w_{sleep}, w_{NodeID} and w_{RSS}) are multiplied with the input values and processed in the hidden layer for Q-function generation. The tree possible outputs generated by the Q function are 100% power transmission, 50% Power transmission and 0% (No) transmission.

The proposed mechanism in its learning stage observes many important features such as:

- the amount of time a node remains in active (transmit or receive) or sleep mode;
- energy consumption in sleep mode, transmit mode and receive mode;
- the level of the RSS (Received Signal Strength) and the real QoS values;
- the neighbor node IDs with their transmission pattern and energy consumption.

2 Conclusions

We have proposed an adaptive model of a wireless sensor node based on Multi-Layer Perceptron (MLP) network representation and machine learning. The approach presented in the paper uses the Reinforcement Learning which as a form of machine learning for node Transmission Power Management (TPM) optimization. This inspires many practical solutions that can maximize resource utilization and prolong the shelf life of the battery-powered wireless sensor networks.

References

[1] Watkins J. C. H, Learning from Delayed Rewards, Ph.D. Thesis, University of Cambridge, England. 1989.

[2] C. J. Watkins and P. Dayan. Q-learning. Machine Learning, 8(3):279–292, 1992

[3] Torrey, L.; Shavlik, J. Transfer learning. In Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques; IGI Global: Hershey, PA, USA, 2009; pp. 242-264.

[4] Gummeson, J.; Ganesan, D.; Corner, M.D.; Shenoy, P. An adaptive link layer for heterogeneous multi-radio mobile sensor networks. IEEE J. Sel. Areas Commun. 2010, 28, 1094–1104.Balmelli

[5] Lin, Z.; van der Schaar, M. Autonomic and Distributed Joint Routing and Power Control for Delay-Sensitive Applications in Multi-Hop Wireless Networks. IEEE Trans.Wirel. Commun. 2011, 10, 102–113.

[6] Wiering, M.; van Otterlo, M. (Eds.) Reinforcement Learning. In Adaptation, Learning, and Optimization; Springer: Berlin/Heidelberg, Germany, 2012; Volume 12.

[7] Sung, Y.; Ahn, E.; Cho, K. Q-learning Reward Propagation Method for Reducing the Transmission Power of Sensor Nodes in Wireless Sensor Networks. Wirel. Pers. Commun. 2013, 73, 257–273.

[8] Udenze, A.; McDonald-Maier, K. Direct Reinforcement Learning for Autonomous Power Configuration and Control in Wireless Networks. In Proceedings of the NASA/ESA Conference on Adaptive Hardware and Systems, AHS 2009, San Francisco, CA, USA, 29 July–1 August 2009; pp. 289–296.

[9] Hu, J.; Wellman, M.P. Nash Q-learning for general-sum stochastic games. J. Mach. Learn. Res. 2003, 4, 1039–1069.

[10] Y. Chu, P. Mitchell, and D. Grace, “ALOHA and q-learning based medium access control for wireless sensor networks,” in International Symposium on Wireless Communication Systems, 2012, pp. 511–515.

[11] A. Forster, “Machine learning techniques applied to wireless ad-hoc networks: Guide and survey,” in 3rd International Conference on Intelligent Sensors, Sensor Networks and Information. IEEE, 2007, pp. 365–370.

[12] Galzarano, S. Liotta, A.; Fortino, G. QL-MAC: A Q-learning based MAC for wireless sensor networks. In Algorithms and Architectures for Parallel Processing; ICA3PP 2013; Lecture Notes

Weights Permutation in Multilayer Perceptron

Todor Balabanov, Iliyan Zankinski, Rumen Ketipov
Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
acad. Georgi Bonchev Str., block 2, office 514, 1113 Sofia, Bulgaria
todorb@iinf.bas.bg

Abstract: Artificial neural networks are very popular in the field of artificial intelligence. In the last three decades many researchers are looking for different ways for artificial neural networks training speed-up. It is well known that once trained artificial neural networks are very efficient, but the training itself is generally slow and time consuming. This research addresses the possibilities for artificial neural networks training speed-up by permutation of the weights in the hidden layers.

Keywords: artificial neural networks, machine learning, back-propagation.

1 Introduction

Artificial neural networks are well known tool in the field of the artificial intelligence [1,2]. In most cases they are used for classification or forecasting problems. Training of the artificial neural networks is challenging and heavily researched problem. There are many different types of artificial neural networks, some of which are more exotic as generalized nets [3] than the others. The most popular is the multilayer perceptron for which back-propagation training algorithm is the most used one [4]. The training sometimes is sensitive in the direction of blocking in local optimums [5]. There are many alternatives for local optimums escapes by stochastic based trainings [6,7], but this research proposes stochastic weights permutation instead.

The paper is organized as follows: Section 1 introduces the problem; Section 2 describe the proposed training algorithm; Section 2 presents some experiments and results and Section 4 gives some conclusions and ideas for further research.

2 Training Algorithm Proposal

Proposed by Znakinski [8], algorithm for neurons permutation is in the base of the idea for artificial neural network weights permutation. Permutation of neurons is bigger than permutation of two weights, that is why this approach looks better. Which weights to be swapped is randomly selected (Listing 1). The training algorithm does not differs in any other way from the classic back-propagation training. The selection of neurons which weights to be swapped is done randomly. Selection is done in such way that neurons to be different in the layer in order self selection to be escaped. The permutation itself is not constantly applied, but it has probabilistic rate given as a constant.

```
/** Permutate two randomly selected weights. */
private void permutateRandomWeights() {
    /* Select random layer except the first one. */
    swapLayerIndex = 1 + PRNG.nextInt(layers.size() - 1);

    /* Select two random different neurons from the previous layer. */
    do {
        swapNuronIndex1 = PRNG.nextInt(layers.get(swapLayerIndex).size());
        swapNuronIndex2 = PRNG.nextInt(layers.get(swapLayerIndex).size());
    } while (swapNuronIndex1 == swapNuronIndex2);

    float weights1[] = layers.get(swapLayerIndex).getWeights(swapNuronIndex1);
    float weights2[] = layers.get(swapLayerIndex).getWeights(swapNuronIndex2);

    /* Swap randomly selected weights. */
    swapWeightIndex = 0;
    float difference = Math.abs(weights1[0] - weights2[0]);
    int length = Math.min(weights1.length, weights2.length);
    for (int i = 1; i < length; i++) {
        if (Math.abs(weights1[i] - weights2[i]) < difference) {
            swapWeightIndex = i;
            difference = Math.abs(weights1[i] - weights2[i]);
        }
    }

    float buffer = weights1[swapWeightIndex];
    weights1[swapWeightIndex] = weights2[swapWeightIndex];
    weights2[swapWeightIndex] = buffer;
}
```

Listing 1. Random selection of weights to swap.

Weights are swapped for a single cycle of training and after that they are returned on their original places (Listing 2). Weights swap is done only for short period of time, because if it is done for longer period it will block back-propagation procedure and it will not converge at all. Such weights permutation can be accepted as analogy of a training noise. The other very im-

portant modification is that weights with minimum difference should be swapped. If the difference of the swapped weights is bigger the algorithm is more damaging in the training process than helping.

```
/**
 * Swap weights back after random swap.
 */
private void swapBackWeights() {
    if (swapLayerIndex == -1) {
        return;
    }
    if (swapNuronIndex1 == -1) {
        return;
    }
    if (swapNuronIndex2 == -1) {
        return;
    }
    if (swapWeightIndex == -1) {
        return;
    }

    float weights1[] = layers.get(swapLayerIndex).getWeights(swapNuronIndex1);
    float weights2[] = layers.get(swapLayerIndex).getWeights(swapNuronIndex2);
    float buffer = weights1[swapWeightIndex];
    weights1[swapWeightIndex] = weights2[swapWeightIndex];
    weights2[swapWeightIndex] = buffer;

    swapLayerIndex = -1;
    swapNuronIndex1 = -1;
    swapNuronIndex2 = -1;
    swapWeightIndex = -1;
}
```

Listing 1. Weights back swap.

3 Experiments and Results

All experiments are done with custom software application written as open source software project, publicly available [9]. Java programming language is used for the implementation of multi layer perceptron for learning XOR function. The topology of the network is 2-30-1 and it was selected specially with bigger hidden layer in order better training comparison to be done. The artificial neural network structure is as one-dimensional arrays for the layers and two-dimensional arrays for the weights. All training examples are shuffled before each training cycle. This is done in order the network to learn the examples not the order in which they are supplied.

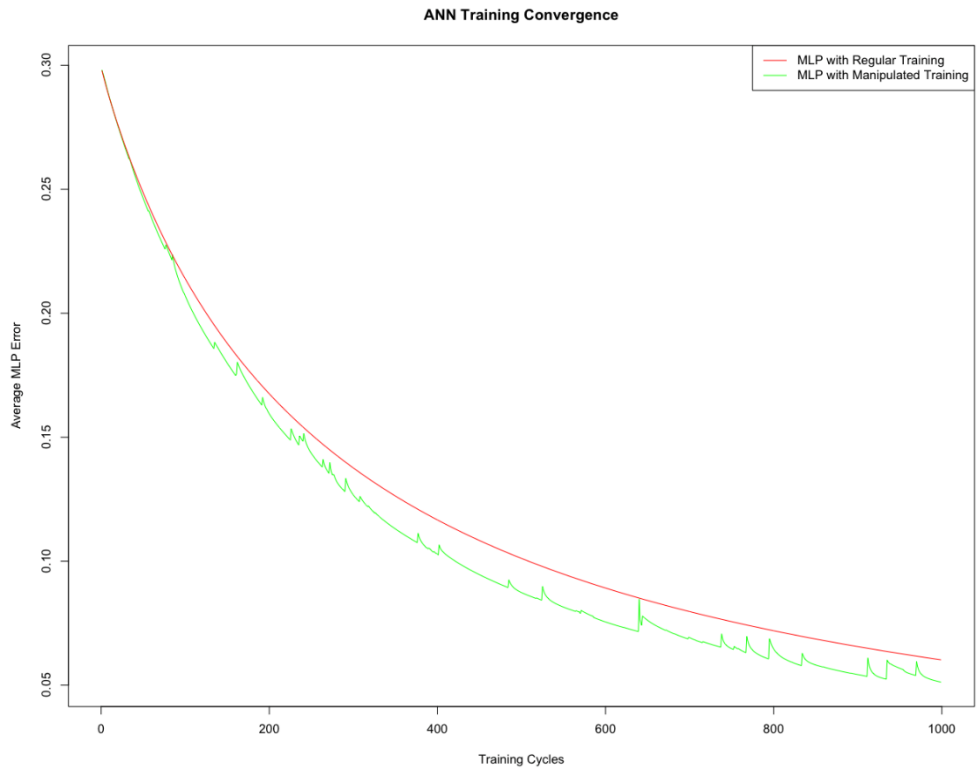

Figure 1. Classic back-propagation compared with permuted weights algorithm.

Classic back-propagation algorithm is compared with permuted weights algorithm. Both trainings are done 30 times and the average performance was presented in Fig. 1. The results shows that permuted weights algorithm is less stable, but outperforms slightly classic back-propagation. The graph shows also a kick-back effect of weights permutation. It looks like training has limited restart, but convergence speeds-up.

4 Conclusions

Experiments clearly shows that the proposed training modification leads to training speed-up, even that the process is not so stable. As further work it will be interesting weights permutation to be combined with activation function permutation.

Acknowledgements

This work was supported by private funding of Velbazhd Software LLC.

References

[1] Atanasova, T., Barova, M., Balabanov, T.: Use of neural patterns to analyze time lines in large amounts of data. Publishing complex Vasil Levski, 193--198, ISSN:1314-1937, (2016)

[2] Tomov, P., Monov, V., Artificial Neural Networks and Differential Evolution Used for Time Series Forecasting in Distributed Environment, Proceedings of International conference Automatics and Informatics, ISSN 1313-1850, 129--132, Sofia, Bulgaria, (2016)

[3] Tashev, T., Hristov, H., Modeling of synthesis of information processes with generalized nets. In: Drinov, M. (ed.) Cybernetics and Information Technologies, vol. 2, Academic Publishing House, Sofia, 92–104, (2003)

[4] Zankinski, I., Tomov, P., Balabanov, T., Alternative Activation Function Derivative in Artificial Neural Networks, Proceedings of XXV International Symposium Management of Energy, Industrial and Environmental Systems, ISSN 1313-2237, Bankya, Bulgaria, 79--81, (2017)

[5] Balabanov, T., Avoiding Local Optimums in Distributed Population based Heuristic Algorithms (in Bulgarian), Proceedings of XXIII International Symposium Management of energy, industrial and environmental systems, John Atanasoff Union of Automation and Informatics, Sofia, Bulgaria, 83–86, (2015)

[6] Balabanov, T., Zankinski, I., Dobrinkova, D., Time Series Prediction by Artificial Neural Networks and Differential Evolution in Distributed Environment, International Conference on Large-Scale Scientific Computing, Sozopol, Bulgaria, Lecture Notes in Computer Science, vol. 7116, no. 1, 198–205, (2011)

[7] Balabanov, T., Zankinski, I., Barova, M., Strategy for Individuals Distribution by Incident Nodes Participation in Star Topology of Distributed Evolutionary Algorithms, Cybernetics and Information Technologies, Sofia, Bulgaria, vol. 16, no. 1, 80–88, (2016)

[8] Zankinski, I., Stoilov, T., Effect of the Neuron Permutation Problem on Training Artificial Neural Networks with Genetic Algorithms in Distributed Computing, Proceedings of XXIV International Symposium Management of Energy, Industrial and Environmental Systems, ISSN 1313-2237, Bankya, Bulgaria, 53--55, (2016)

[9] Balabanov, T., Multi Layer Perceptron implemented in Java, https://github.com/Coding-Sunday-Sofia/multi-layer-perceptron, (2018)

Use of Information Resources of Organizational Systems to Support Managerial Decisions

Bakanova N. B.*, Atanasova T. V.**
*Keldysh Institute of Applied Mathematics - RAN,
Miusskaya sq., 4, Moscow, 125047, Russia
**Institute of Information and Communication Technologies – BAS,
"Akad. G. Bonchev" St., bl. 2, Sofia - 1113, Bulgaria
nina@keldysh.ru, atanasova@iit.bas.bg

Abstract: The article considers the possibility to develop information support functions for management activities based on the internal resources of the information environment of the organization. The advantage of such functions is providing real-time, up-to-date and reliable data that is derived from information sources within the organization or obtained from external sources, tested in applied subsystems.

Keywords: information resources, documentation support, management activity support.

1 Introduction

Making informed and justified management decisions is the most significant function in the work of the management organization. Leading employees in daily activities must constantly accept and reject hundreds of different alternatives for various decisions aimed at fulfilling the strategic and tactical tasks of the organization. In the decision-making process, it is necessary to take into account: the current state of the organization, relationships with counterparties, development prospects and many other problems affecting the effectiveness of the organization. The activities of leading decision-making bodies of large organizations are complicated by a number of features, including a) wide variety of simultaneously performed activities at different levels of the management hierarchy; b) high level of responsibility; c) non-algorithmic, creative character of the activity carried out in the conditions of the most diverse limitations; d) a number of other related to the specifics of the management function.

To provide information support to the management activities, research and development in a number of related scientific areas is ongoing [1-3]. One of the research directions is the creation of knowledge bases in the field of organizational management in specific areas. This is an interesting, but resource-intensive task. The use of expert knowledge in the activities of

organizational structures is also a promising study, which unfortunately has, at present, significant limitations on efficiency.

Actual directions of developments in this area are research on organizational management processes, identification of tasks, support of which can be implemented on the basis of resources stored in the databases of the company's software landscapes, creation of methods and algorithms that allow using these data for information support of management activities.

In connection with ubiquitous processes of informatization of organizational structures, research on the creation of information services for decision support based on the use of available information resources has growing significance [4-6]. The introduction of information technologies in the activities of organizational structures, along with the fulfillment of the basic tasks in the automation of managerial activities, ensures the accumulation of the information resource of the organization, creates opportunities for developing new methods for supporting decision-making based on the accumulated information resources.

This paper considers the possibility to develop special services to support the making of management decisions based on information resources in the computing landscape of the organization.

4 Problem description

On the servers of the existing information resources of the organizational management, a lot of databases containing information on the activities of the organization are formed, as:

- data on the fulfillment of tasks, orders, projects;
- data on the performance of business processes in the organization;
- information on relationships with counterparties;
- strategic data on directions and conditions of prospective studies;
- information about human resources;
- information on financial activities;
- other data determined by the specifics of the organization.

All these data are accumulated in the databases of the local subsystems in the computing landscapes and represent a powerful information resource of the organization that can be used to prepare new management solutions that take into account the experience of implementing of previous tasks. The main part of the data is the result of the work of the listed subsystems. The other part of the data is additional information that is formed during the process of providing the functions of the system's life cycle, for example, the time of information reception/transmission, data about the waiting processes, data about retransmissions, the time of

information distribution, and the time for the preparation of responses to the requests. The data, in its majority is technical information, at the same time it indirectly characterizes the business processes of the organization. Specialized joint processing of the specified data allows its usage for development of services of information support of administrative processes of the organization [6].

In addition, a promising direction is the integration of information resources for the implementation of tasks to support the making of managerial decisions. In the existing information environment of organizational management, the support functions are usually provided only by means of local subsystems, they have not a systemic nature, and the possibilities of multipurpose usage of information resources of various subsystems are not realized. The databases of software landscapes are used for solving only local problems.

To implement management decision support services (based on the existing information environment), it is required to consider the structure of the computing landscapes of organizational management, the reflection of the organization's business processes in information systems, the opportunities for the integrated use of data from various subsystems for implementing methods and software support tools.

5 Structure of the information environment of organizational management systems

From the point of view of organizing the information processes, the complex computing landscape of a large management organization can be represented in several levels. Each level is designed to solve its group of tasks, while ensuring interaction with other levels of the information environment. The main components include:

– level of the operating environment of the complex computing landscape;
– level of maintenance of lifecycle processes of applied systems;
– level of applied processes;
– level of support for making management decisions.

The set of the applied subsystems required in the management activities of the organization is determined during the design of the information system. Each of the subsystems provides information and technological support for the work performed, the accumulation of information about the processes and results of work.

The subsystems traditionally incorporated in the complex computing landscape for solving organizational management tasks include: (i) documentation support of management (document circulation of the organization); (ii) support of financial activities; (iii) management of

personnel services; (iv) material and technical support; (v) planning tasks for different target areas, etc.

Schematic representation of the complex landscape of applied systems of organizational management is presented in Figure 1.

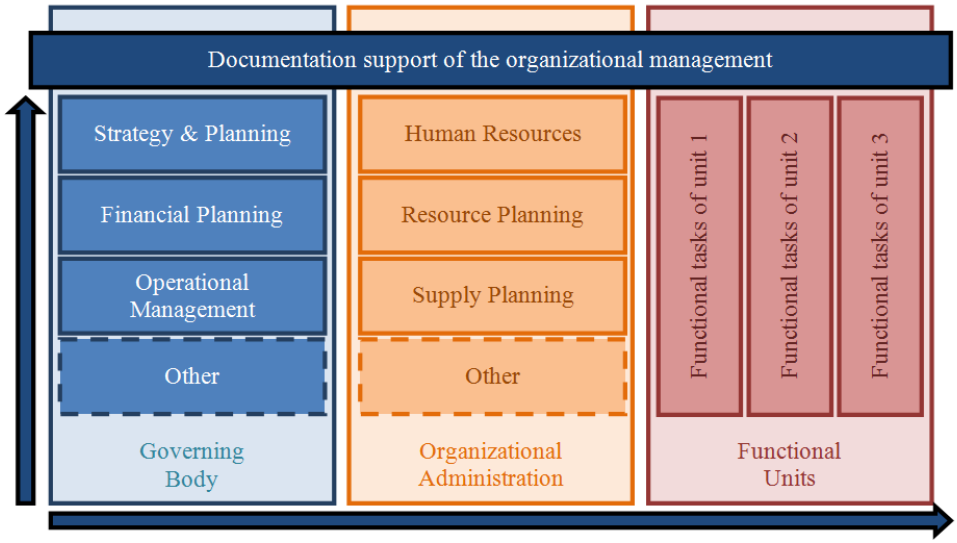

Figure 1. Information landscape of organizational management

The applied subsystems process the data and form summary reports, references, reminders, sheets, specifications, which are used in the operational work of the organization.

In the process of functioning of these subsystems, information is accumulating (descriptive, quantitative, qualitative) about the processes and features of the performance of specific management tasks, about the experience of solving problems, about the participants in the processes.

In aggregate, such databases represent an information resource of the organization. On their basis, the level of support for managerial decision-making can be formed, including data warehouses and specialized processing algorithms.

Retrospective arrays accumulated in information stores of the organization are usually used for obtaining references and summaries within the application systems that provided the formation of this data. Since data is created, controlled and stored in databases of application subsystems, their status of internal information resources is determined, which validates them as reliable and objective information. The undoubted advantages of using internal data are the following:

– approved sources and mechanisms of information flow in the organization;

– proven processes of reception, sorting, unification and control for duplication of incoming information;
– established data structure simplifies the sampling processes for use in the support algorithms;
– availability of mechanisms that monitor incoming information.

Taking into account the importance of the tasks of improving management activities, the problem of comprehensive use of the accumulated resource for creating information support tools on various aspects of the organization's activity is of practical interest.

6 Reflection of the organization's business processes in information systems

As it can be seen in Figure 1, the subsystem "documentation support of management" covers all business processes of the organization. Most of the indicators of improving management processes, increasing efficiency, identifying bottlenecks and shortcomings in organizational activities can be identified on the basis of investigation on processing, passing and executing business documents [7]. The key issues of such analysis are sources of information, performance evaluation criteria, methods for obtaining estimates, data processing methods for comparing the indicators of various functional units in the organization.

Expert analysis of documents and indicators of their processing makes it possible to assess results of performance activities. However, expert analysis is a long, expensive and resource-intensive procedure. For the purpose of performing the analysis without using expert procedures, the dataflow has been analyzed. The object of this study has been the details of the document, data on the performing processes, as well as the normative code of the rules, according to which the documents are passed and processed in the management organization [8].

As a result of the conducted studies of a) document management processes, b) data monitoring of the functioning and interaction of information systems, c) analysis of the regulation rules, d) consultations with experts of the applied field, the following provisions were formulated which have formed the basis for this development:

– on the base of the information received at the registration stage, one can determine the characteristics of the actuality and significance of the document;
– the number of established control terms indicates the attention of the governing bodies to the implementation of this document;
– the quality of the document processing performance is demonstrated by the promptness of the preparation and harmonization of the documents and the results, as well as the

characteristics of the processes of preparing resolutions at all levels of the organizational hierarchy;

– when evaluating the work of the functional organizational unit, data on the totality of subordinate units are taken into account;
– a number of other provisions, based on monitoring data and interaction of information systems.

7 Analysis of the information resource

On the base of available information resources a method has been developed. The method includes:

– evaluation criteria and scales for the graduation of performance evaluations;
– algorithm for selecting information needed for analysis;
– algorithm for obtaining estimates based on selected criteria;
– algorithm for analyzing activities based on obtained estimates.

The method allows to calculate the evaluation criteria for information processes, to perform a comparison of the activities of participants in management processes and to identify problem areas in the organization's activities [9]. The developed criteria for evaluating the information processes are oriented to algorithmic processing and do not require special expertise. The data are extracted during the operation of information systems. Figure 2 shows the scheme for grouping the criteria.

The upper level of the hierarchy in each group represents a certain indicator of the organization's information processes. In the analysis, the indicators can be used separately or integrally.

The first group forms the "the importance of the document" indicator, which is built on the identification characteristics of the document and allows evaluating the composition of the input document flow. The estimates of this group do not change during the processing of the document and are used to highlight the most significant documents for the purpose of further research of information processes. The indicator of significance includes two components: "importance" and "links (relationships) with activities". The criteria defining "importance" assess the type of the document, the correspondent, the availability of directive terms /deadlines and other data. Attitude to the highest priority activities is determined depending on the relationship of the document with the directory of thematic areas and standards of the organization and on the priorities of documents processing adopted by the organization. For example, the response to a request from a higher-level organization on the problem that the organization

oversees is more important than the interdisciplinary query on cross-industry topics, etc. In the process of setting specific tasks, the priorities of the rating scale can be changed, and additional restrictions are included. The estimates on the criteria of the first group can be used to study the input flow, the formation of data samples (by the dates of receipt, projects, the importance of documents, etc.) to analyze the implementation of information processes.

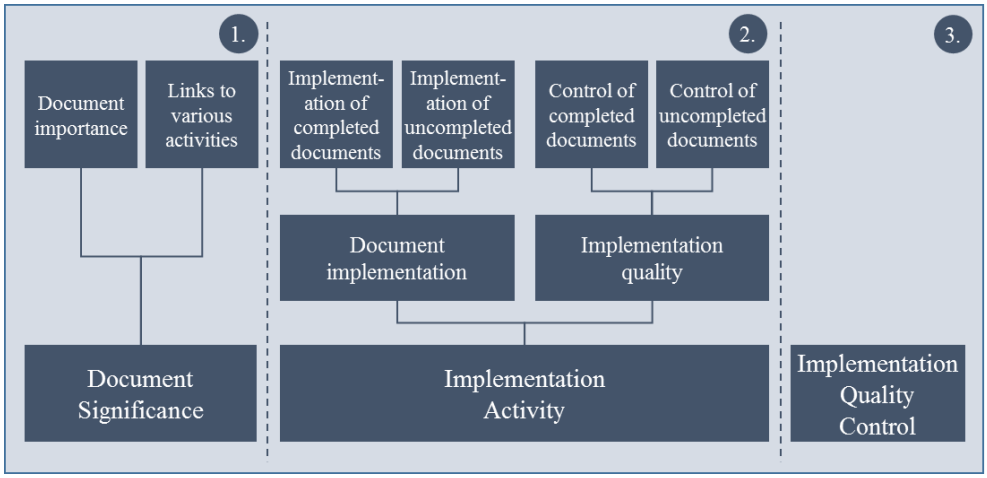

Figure 2. Groups of criteria

8 Conclusions

At present, there are no developments that suggest methods for studying the data composition from applied subsystems for purposes of summarizing management problems, identifying analogue situations, and providing estimates based on the results of the assignments. This is caused by the scatteredness of applied research directions, and the fact that the applied subsystems are focused on the data that is obtained from their own functioning.

To solve the problem, an approach has been developed that makes it possible to expand the information content on the level of support for making managerial decisions by creating an integrated information environment that accumulates data from all applied subsystems to implement management support functions without using labor-intensive expert information processing.

At the center of the developed approach lies the idea of using the information resource of software landscapes functioning in organizations to improve the efficiency and information processing in management activities.

Acknowledgements

This work was partly supported by a grant РФФИ №18-07-00833А

References

[1] Ferilli S. (2011) Automatic Digital Document Processing and Management, *Advances in Pattern Recognition*, Springer-Verlag London, 196 p.

[2] Executive's Guide to Information Technology (EGIT) (2007) Ed. J. Baschab, J. Piot, N. G. Carr. – J. Wiley & Sons, Inc., 672 p.

[3] Gibson, J. (2011) Organizations: Behavior, Structure, Processes, New York: McGraw Hill, 640 p.

[4] Bakanova N., Atanasova T. (2013) Data Sources Determination for Information Management in OMIS, Proc. Int. Conf. Robotics, Automation and Mechatronics RAM 2013, pp. 25-28.

[5] Atanasova T., Bakanova N. (2014) Information support for decision making in organization management, Proc. Int. Conf. Robotics, Automation and Mechatronics RAM'2014, pp. 72-76.

[6] Bakanova N. (2015) Analysis of information process in a management organization for realization of decision making support, *Электросвязь*, No 5, pp. 59–62.

[7] Bakanov A.S., Volchkov D.V., Bakanova N.B., (2015) Model of human interaction with the intellectual information system, (in Russian), *Электросвязь*, No 4, pp. 47–49.

[8] Bakanova N.B., (2013) Informational approach to the development of management support systems, *University Bulletin (State University of Management)* Moscow, (in Russian), № 16, pp. 193-200.

[9] Bakanova N., Atanasova T. (2018) Analysis of Information Resources of Organizational IS for Development of Algorithms for Support of Decision Making (in Russian), NVU "Vasil Levski", Tom 9, pp. 101-105.

A Study on Behavioral Models in a Smart Home Environment

Victor Danev, Victoria Muchanova, Alexandre Chikalanov, Roumen Nikolov
University of Library Studies and Information Technologies
Bul. Tzarigradsko Shosse 119, Sofia, Bulgaria
victor_danev@abv.bg, viktoriya.muchanova@gmail.com,
a.chikalanov@unibit.bg, r.nikolov@unibit.bg

Abstract: The main goal of the paper is to describe and analyse the initial steps in a study on behavioral models in SmartPlace environments (apartments, houses, buildings). The paper steps on a smart home research environment, which consists of integrated solutions for smart management of resources and comfort: using variety of smart sensors; appliances and other devices; registered useful models for lighting and heating; development of the technology infrastructure for data analytics, mobile applications, and user interfaces (Android/iOS); a cloud infrastructure and services for Big Data analytics and support for providing a Smart Place as a Service (SPaaS) functionality. The paper analyses the challenges, related to the energy-related occupants' behavior studies and behavioral monitoring in smart-home environments for health-care applications. Study methodologies (in-situ studies, laboratory studies, surveys, mixed methods studies). Measures for research to respect high level ethical standards, including protecting the rights and benefits of participants have been taken. Some challenges and research topics related to occupant behavior for future multidisciplinary research have been identified.

Keywords: Internet of Things (IoT), Cloud Computing, Fog Computing, Smart Place as a Service, Smart Home, Smart City, Active and Healthy Ageing, Activities of Daily Living, Big Data, Data Analytics

1 Introduction

Society is in a process of digital transition and as a result, organizations are producing and storing vast amounts of data. Managing and gaining insights from the produced data is a challenge and a key to competitive advantage. Analytics solution as data mining can discover new patterns from large data sets. "Big data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse [1]. Big Data refers also to various forms of large information sets that require special computational platforms in order to be analysed. Big data and the technologies associated with it can bring significant benefits to the business. But the tremendous uses of these technologies make difficult for an

organization to strongly control these vast and heterogeneous collections of data to get further analysed and investigated.

The construction sector becomes more and more data intensive. A smart home may include: devices with a range of sensory capabilities such as programmable communicating thermostats, lighting and smart power strips; communications systems that facilitate two-way information flow between devices and the occupant, and possibly between devices and the utility or other third-party firms such as security system providers; and, monitoring and control systems that allow occupants to track energy usage and change the operations and functions of devices within the home [2]. The devices included in a smart home can vary tremendously in their design and function, but, in general, smart devices provide customers with the following two functions: ability to monitor energy use in real-time or near real-time for the whole house and/or by device; and, ability to remotely control systems or appliances in a home. The advent of smart homes may ensure smart technologies become a commonplace feature of people's lives, whether they are wanted or not [3].

Nine research themes related to smart home environments are identified and analysed [4]. They are classified in three groups: (1) views of the smart home—functional, instrumental, socio-technical; (2) users and the use of the smart home—prospective users, interactions and decisions, using technologies in the home; and (3) challenges for implementation of the smart home - hardware and software, design and home-keeping. The smart home lets users to properly manage the inhabited place environment and resources and to improve the living experience through new functionalities, such as - remote control and automation of appliances, heating and lighting management, comfort management, security management, etc. In addition, smart homes are becoming nodes of a smart energy system that allows utilities to respond to real-time flows of information on energy demand fed back by smart meters from millions of homes [5].

A fast-growing research area is behavioral monitoring in smart home environments [6]. Globally, in 2010, the buildings sector accounted for more than one-fifth of total worldwide consumption of delivered energy, with an increasing projection rate among all sectors [7]. Presently, 73% of electricity and 55% of natural gas in the United States is consumed in buildings, with other countries encountering similar consumption challenges. Energy-related occupant behavior in buildings is a key issue for building design optimization, energy diagnosis, performance evaluation, and building energy simulation[8]. Occupant actions can have a significant impact on the real energy use and occupant comfort in buildings. A deeper understanding of occupant behavior and improving capability to quantify its impact on the use of

building technologies and building performance with modeling and simulation tools are crucial to the design and operation of low-energy buildings, where human-building interactions are key aspects of concern. However, the influence of occupant behavior is under-recognized or over-simplified in the design, construction, operation, and retrofit of buildings. Occupant behavior is complex and requires an inter disciplinary approach to be fully understood. In addition, occupant behavior is influenced by external factors such as culture, economy, and climate, as well as internal factors such as individual comfort preference, physiology, and psychology.

This paper provides an initial study on behavioral models in a smart home environment (Smart Place) implemented in the context of the D-PLACE project "Digital Solutions for Intelligent Management of Buildings and Places". The main competitive advantage of the solution is its holistic approach – a combination of both conventional construction technologies and solutions for energy efficiency using ICT (FIWARE, cloud and mobile technologies). The system is based on an innovative model: Smart Place as a Service [9] which is implemented by usage of open source software and hardware, open standards and open data and thus avoiding vendor dependency. It will use also a combination of Big Data processing and simulation software and thus - providing additional opportunities for offering flexible solutions ('what-if' functionality) for smart places, driven by: energy efficiency, budget, security and comfort. Smart Place is opening a new possibility - to continuously learn and anticipate the needs and preferences of the inhabitants in terms of temperature, light and comfort. In addition, the solution aims at following the revised EU Energy Performance of Buildings Directive 2018/844 which entered into force on 9 July 2018, and the emerging Smart Readiness Indicator for Buildings (smartreadinessindicator.eu).

2 The Smart Place Concept

The general objective of the Smart Place project is to develop a model and a Future Internet based platform and services for energy efficiency and user comfort monitoring and management in a Smart Building environment by using a combination of Big Data processing and simulation software [9, 10]. Smart Place is an integrated system with great potential for market penetration. The Smart Place as a Service concept counts on an intelligent self-learning platform, which is not just storing Big Data on the cloud, but also – analysing behavior patterns and harnessing sophisticated algorithms for self-learning and optimization. The architecture is built according to a high level multi-tier architecture paradigm (Figure 1).

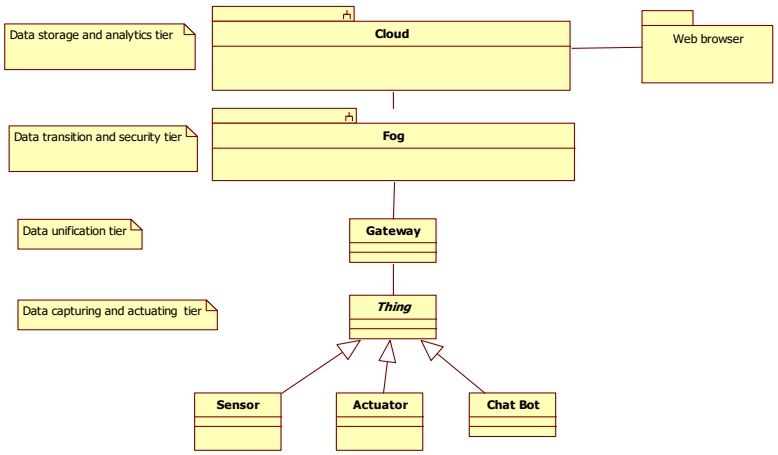

Figure 1. High level multi-tier architecture of proposed application

Smart Place includes development of an integrated environment, which consists of three components:

- integrated solutions for smart management of resources and comfort: using variety of smart sensors; appliances and other devices; registered useful models for lighting and heating; development of the technology infrastructure for data analytics, mobile applications, and user interfaces (Android/iOS);
- platform for sharing of: projects, algorithms, a constructor with graphical interface for smart solutions dedicated to semi-professional/professional end-users for management of smart places;
- cloud infrastructure and services for Big Data analytics and support for providing a Smart Place as a Service (SPaaS) functionality. The envisaged products and services have innovative character in a global reach.

The integrated solutions for smart management of resources and comfort of inhabited places consist of a variety of intelligent sensors and actuators, appliances; a set of appropriate models for lighting systems and heating; appropriate technology infrastructure and tools for Big Data analytics, mobile applications and user interfaces (Android/iOS). The model of Fog Computing is being used for analysing and acting on IoT data. This allows to analyse the most time-sensitive data at the network edge, close to where it is generated instead of sending vast amounts of IoT data to the cloud. The system acts on IoT data in milliseconds and sends selected data to the cloud for historical analysis and longer-term storage [11].

The main outcome of the first phase will be a well calibrated and evaluated (in a real environment with real users) prototype of a smart place (apartment), which includes a system for

monitoring of energy consumption and comfort. This will lead to improved indicators for energy consumption and level of comfort and wellbeing. This prototype will be in the core of service for an integrated solution for a smart place, which (depending on the concrete needs and technology competence of the user) will include a complete product (software, sensors, actuators, controllers) with basic functionality, or – a service (design and implementation of an integrated solution, including a sensor network, appliances, etc.). The implemented smartphone application will allow monitoring the indicators of energy efficiency and comfort, providing control of the smart cyber-physical system, as well as to control different components. In addition, by using the concepts of Fog Computing and Cloud Computing, some functionality of data analytics and predictions will be implemented, e.g. behavior monitoring and predictions, creation of user profiles, etc. This will allow better personalisation of the application based on the evaluation of the user experience. The application will use also instruments for learning, e.g. through serious games and gamification scenarios. Different award schemes will be promoted through specialized or general social networks aiming to stimulate users to reduce energy consumption and CO_2 emissions. The prototype is based on open source components and will provide services to individual users. These services will be available through a specialized Smart Place platform and dedicated to different types of inhabited places, such as apartments and homes. Remote management of the resources and comfort will be achieved by dedicated smartphone applications and specialized data analysis and recommendations for behavior change.

The Smart Place as a Service concept counts on an intelligent self-learning platform, which is not just storing Big Data on the cloud, but also - analysing behavior patterns and harnessing sofisticated algorithms for self-learning and optimization. (Fig. 2 & Fig 3).

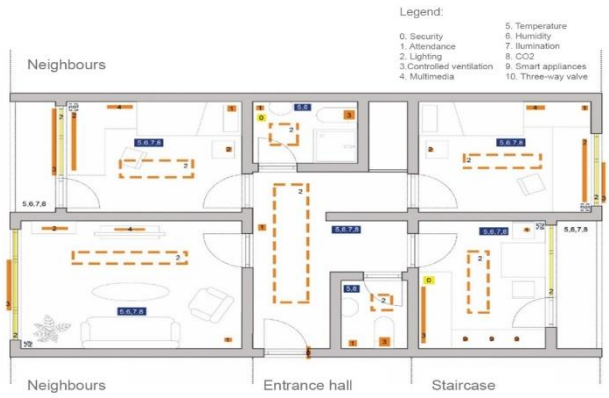

Figure 2. Scheme of a typical apartment with sensors and appliances

Figure 3. A prototype of a smartphone user interface

3 Energy-Related Occupants' Behavior Studies

Energy efficiency is a key strategy for the sustainable growth of our planet, thus EU set a clear objective: in 2020, the consumption of traditional energy and greenhouse emissions should be reduced by 20% [12]. The EU determined around 650 action measures in the residential sector. Some of these measures included the use of smart meters, information campaigns or the encouragement to change user behavior [13]. One of the main goals of the Smart Place study is to develop a proper methodological framework to study the occupants' behavior related to energy consumption by using data collection and analytics, simulation, modeling and evaluation. The methodological framework will be based on the outcomes of the report of the International Energy Association (IEA), Energy in Buildings and Communities (EBC) Annex 66 "Definition and simulation of occupant behavior in buildings" embracing the efforts of more than 100 researchers from 20 countries working together from November 2013 to May 2018 [8]. It was outlined that:

- Occupant behavior has significant impacts on energy use and occupant comfort. There is a need of data, methods, and models to understand and reduce the gap between simulated and measured building energy performance by representing occupant behavior in a standardized ontology and XML schema and developing occupant behavior modelling software;
- Data collection is of crucial importance for occupant behavior modeling. Most data collection procedures are conducted in a typical working or living environment rather

than a laboratory. Researchers are required to have a good understanding of the available data collection methods and apply them to the most appropriate situation;

- Choice of occupant behavior simulation models depends on the building context. The evaluation of occupant behavior models should have explicit metrics that come from the application scenarios to quantify their performance. Often the occupant behavior models are integrated with building performance simulation programs, such as EnergyPlus, ESP-r and DeST. However, user-friendly interfaces need to be further developed to enable occupant behavior simulation for practical applications.
- The representation of occupant behavior diversity in simulation programs is critical. Occupant behavior diversity should be addressed with different approaches, such as case measurements and questionnaire surveys.
- Interdisciplinary research across the building, social, behavioral, data and computer sciences is needed to understand, represent, model and quantity the impact of human behavior on building energy use, occupant comfort and health.

The Smart Place study will adopt the CAFCLA Approach for Activity Monitoring and Energy Efficiency [14] which is based on contextual information, localization and social computing. CAFCLA serves as a basis for the development of an intelligent recommendation system that encourages responsible energy consumption in households. The system integrates combination of the following functionalities within the same structure:

- It obtains contextual information through the implementation of wireless sensor-networks which collect data from multiple sources to define the environment. This includes sensors that collect environmental parameters (temperature, humidity or lighting) or data on device usage (the switching of lights, the status of blinds and windows).
- It implements a real-time localization system that allows users to be identified and tracked at all times. Users' positions permit identifying patterns of behavior that help to describe good or bad energy consumption habits.
- It integrates a social machine that provides users with recommendations. The social machine uses virtual organizations of agents to provide the system with intelligence. Monitoring of all contextual parameters, localization and tracking of users, management of communications and data, as well as the generation of recommendations for users encourage the efficient use of energy.

Smart Place monitors temperature, lighting, electric consumption and users' movement. Some of the recommendations that have been predefined in the Smart Place study are as follows: turn off heating if the temperature is over 18◦C; turn off lights if lighting is over 200 lux.; turn off heating and lights when the last person leaves the house; turn off lights if there is no movement in certain areas; Turn off room lights with no occupancy; turn off heating if no movement is detected at night and the temperature is above 18◦C; optimize the use of the heating schedule by identifying the times in which there are people at home; etc. The development of a recommendation system such as the one presented in this work requires a set of functionalities which are difficult to merge, as shown in the following. The main objective of the recommendation system is to achieve a substantial reduction in energy consumptioninhouseholdsandimprovethedemandresponsethroughthepromotionandacquisition of energy-efficient habits.

4 Behavioral Monitoring in Smart-Home Environments for Health-Care Applications

The aging population is a global challenge. The EC has begun to stimulate cooperation between countries on the care of the elderly with such supranational platforms as the European Innovation Partnership on Active and Healthy Ageing (ec.europa.eu/eip/ageing). As part of the EC initiative to promote the "Silver Economy" over the past three years, a number of macroeconomic analyzes have been carried out and recommendations have been made on how best to stimulate the emerging silver economy in Europe. The EC considers that the silver economy can lead to "systemic innovation", while unlocking the possibility of new jobs and economic growth. The need of care increases with age, so that sustainability of the current social and health-care systems is most severely endangered [15]. A shift from a cure to prevention paradigm is recommended, which also implies focusing on prevention places (homes, living environments), in addition to treatment places (hospitals, sheltered houses). Supporting prevention and enabling home-based assistance is among the main goals of Ambient Assisted Living (AAL) techniques. Many technologies contribute to the AAL vision, predecessors of which can be found in home automation and telemedicine technologies. For instance, a pilot on the implementation of a Smart Health Cardio-Belt designated to facilitate the medication of people with cardio-vascular diseases was implemented in the frames of the FP7 ELLIOT Project [16]. This technology supports the doctors' efforts to observe patients and undertake the most appropriate treatment. The technology allows the observation to be carried out at a distance and

permanently for the period in which the patient is equipped with a cardio-belt while in the same time, the patient is free to conduct a normal style of life. Many AAL systems rely on smart-home technologies, in which a wide variety of sensors are distributed in the home environment to infer behavioral patterns. This may effectively complement classical telemedicine services. Indirect behavioral monitoring, can be carried out through home sensors with no burden on the end-user.

In order to support an AAL related study the Smart Place environment was equipped with some sensors, such as: passive infrared sensors for motion detection, suitable for tracing room occupancy; magnetic contact sensors, useful for monitoring open/close states of different objects, e.g. interactions with doors, drawers and cabinets can be easily detected with such sensors; toilet presence sensor; fridge sensor, to detect openings of the fridge door; power meter, to monitor home appliances use, such as microwave oven, air conditioning, TV, etc. A bed occupancy sensor, useful in tracing sleeping patterns, and chair occupancy sensor, to gather information on how much time and when a user sits on a chair are envisaged as well.

5 Study methodologies

The Smart Place study uses the following study methods: in-situ studies, laboratory studies, surveys and mixed approaches [8]. In-situ studies involve monitoring occupants in their natural environment through on-going data collection. Data are normally generated via sensors that are integrated into the living environments. The sensors detect different type of data, such as occupants' presence, energy use, indoor environmental quality [17] However, the use of existing occupied places limits the flexibility of experiments and the visits to the place can be invasive for occupants. Moreover, this might reduce the accuracy of measurements and may introduce errors. Ethics, participant recruitment, and informed consent remain fundamental challenges for this approach [18].

Laboratory studies require participants to spend time and interact within a specially equipped environment that is intended for scientific studies. Most of the labs look like real home environments, but they are equipped with a large number of sensors and allow greater control over the environment and provide very good opportunities for different scenarios to be simulated according to the experimental design.

Surveys rely on the self-reporting, either by filling out questionnaires or through interviews and focus groups. They are an effective tool for improving the understanding of occupant behavior, and can be used to support the in-situ studies. This method is useful in its ability to reveal the logic and rationale behind habits and behaviors. Surveys are a cost-effective means

of achieving a large sample size and can measure phenomena that would be difficult or impossible to measure with sensors (e.g. comfort level). However, lack of understanding or the misinterpretation of questions might cause occupants to report things incorrectly.

Mixed methods studies can be designed in a number of ways, all with the common feature of combining multiple methods (qualitative, quantitative, or both) in a single study. They usually refer to the type of data being collected for analysis. These can be either quantitative or qualitative. Parallel research designs, which conduct qualitative and quantitative analysis in parallel followed by a comparison for final interpretation, allow researchers to quantify occupant actions and obtain a better understanding of cause and effect while measuring behavior in-situ. Explanatory sequential mixed method designs are appropriate for situations where the quantitative data that are collected cannot be fully explained by the data alone and qualitative methods may offer more insight.

The research must respect high level ethical standards, including protecting the rights and benefits of participants. Any research involving vulnerable participant groups (e.g., children, prisoners, institutionalized individuals) is subject of special attention. Participants' privacy and confidentiality must be maintained and guaranteed with regard to any personal data. In addition, informed consent must be obtained to ensure prospective participants understand the nature of the research, that they can voluntarily decide whether to participate, and that they can cease participation at any point.

6 Conclusions

Occupant behavior is a complex and interdisciplinary research topic, and there are many challenges for future multidisciplinary research. For example [8]: Definition of reliable and affordable ways to collect large-scale occupant behavior data; Development and application of occupant behavior models; Representation of inter-occupant behavior diversity; Consideration of interaction of multiple occupants; Model fidelity for specific application context; Methods and datasets for model evaluation, verification, and validation; Standard approaches to integrating occupant behavior models or tools with the existing building performance simulation programs.

The research would lead to advanced solutions and applications, such as: Guideline to integrate occupant behavior sensing, analytics, modeling, and simulation with the building lifecycle, including planning, design, construction, commissioning, operation, controls, and retrofit; technology development and evaluation, considering different scenarios of occupant

behavior (including in an AAL context with telemedicine functionality) in the modeling, simulation, and evaluation of building technologies to understand the variation of performance; etc.

Acknowledgement

The research undertaken in this paper was funded under the National Research Program on "ICT for common digital market, education and security". The research infrastructure for conducting the research was funded by the project D-PLACE (№ BG16RFOP002-1.002-0798) under the Operational Program "Innovation and Competitiveness"

References

[1] Manyika, J., et al (2011), Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

[2] Saul-Rinaldi, K, LeBaron, R and Caracino, J (2014), Making Sense of the Smart Home. Applications of Smart Grid and Smart Home Technologies for the Home Performance Industry, National Home Performance Council, UK.

[3] Haines V, et al (2007) Probing user values in the home environment within a technology driven smart home project. Personal and Ubiquitous Computing 11, pp 349–359

[4] Wilson, C., Hargreaves, T., Hauxwell-Baldwin, R. (2015), Smart homes and their users: a systematic analysis and key challenges, Personal and Ubiquitous Computing, February 2015, Volume 19, Issue 2, pp 463–476

[5] Darby S (2010) Smart metering: what potential for householder engagement? Building Research and Information, 38(5), pp 442–457

[6] Delzendeh, E., Wu, S., Lee, A., Zhou, Y. (2017), The impact of occupants' behaviors on building energy analysis: A research review, Elsevier, Renewable and Sustainable Energy Reviews, Volume 80, December 2017, Pages 1061-1071

[7] USEIA (2014), Energy Information Administration, International Energy Outlook, Buildings sector energy consumption, Analysis of world buildings sector delivered energy consumption 2010 to 2040, www.eia.gov

[8] IEA (2018), EBC Annex 66, Definition and Simulation of Occupant Behavior in Buildings, The Regents of the University of California (through Lawrence Berkeley National Laboratory), https://annex66.org/sites/default/files/2018FinalReport/Annex%2066%20Final%20Report%20-%20read.pdf

[9] Nikolov; R,. Chikalanov, A; Shoikova, E, Paskalev, D; Rashevski, M (2016), Smart Place as a Service: A Model for Providing Big Data Solutions for Smart and Energy Efficient Buildings and

Places, Proceedings of the International Conference BdKCSE'2016 – Big Data, Knowledge and Control Systems Engineering, Sofia, 1-2 December

[10] Chikalanov, A., Nikolov, R., Petkov, J. & Petrov, Z. (2017), "A use-case driven open reference architecture for cross-domain IoT open source component based applications", International Conference "Big Data, Knowledge and Control Systems Engineering – BdKCSE'2017", 7 December 2017, Sofia (Bulgaria), Open Access Proceedings, ISSN:2367-6450, pages: 9-25

[11] Cisco (2015), Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are, White Paper, https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf

[12] Gynther, L.; Lapillone, B; Pollier, K. (2015), Monitoring of energy efficiency trends and policies in the EU. An Analysis Based on the ODYSSEE and MURE Databases; E.U. ODYSSEE-MURE Project; Followme Communication: Aigues Vives, France, September

[13] Barbu, A.D.; Griffiths, N.; Morton, G. (2013), Achieving Energy Efficiency through Behavior Change: What Does It Take? In EEA TechnicalReport, No 5/2013; European Environment Agency: Copenhagen, Denmark, 2013; ISSN 1725-2237. 2.

[14] García, Ó., Prieto, J., Alonso, R., Corchado, J. (2017), A Framework to Improve Energy Efficient Behaviour at Home through Activity and Context Monitoring, Sensors Journal (Basel), Aug; 17(8): 1749.

[15] Mora, N., Matrella, G., Ciampolini, P. (2018), Cloud-Based Behavioral Monitoring in Smart Homes, Sensors, doi:10.3390/s18061951

[16] Kovatcheva E., Nikolov R, Madjarova M., Chikalanov A (2013) Internet of Things for Wellbeing: Pilot Case of a Smart Health Cardio Belt, in Proceeding MEDICON 2013: XIII Mediterranean Conference on Medical and Biological Engineering and Computing, September 25-28, Seville, Spain

[17] Haldi, F., et al (2017). Modelling diversity in building occupant behavior: a novel statistical approach, Journal of Building Performance Simulation, 10, 527-544

[18] Gilani, S. & O'Brien, W. (2017), Review of current methods, opportunities, and challenges for in-situ monitoring to support occupant modelling in office spaces. Journal of Building Performance Simulation, 10, 444-470

Scientific Data Processing from Remote Objects

Dichko Bachvarov[1], Ani Boneva[1], Veronika Ivanova[2], Yordanka Boneva[1],
Boyan Kirov[3], Katya Georgieva[3]

[1] Institute of Information and Communication Technologies – Bulgarian Academy of Sciences,
Acad. G. Bonchev Str., Bl. 2, Sofia 1113, Bulgaria,
dichko1952@abv.bg, a_boneva1964@abv.bg, boneva.yordanka@gmail.com

[2] Institute of Robotics - Bulgarian Academy of Sciences,
Acad G. Bonchev Str., bl.2, Sofia 1113, Bulgaria
iwanowa.w@abv.bg

[3]Space Research and Technology Institute - Bulgarian Academy of Sciences,
Acad G. Bonchev Str. Bl. 1, 1113 Sofia, Bulgaria,
bkirov@space.bas.bg; kgeorg@bas.bg

Abstract: The article presents some methods and tools used for data processing of scientific data from remote objects. “Remote objects” will be named all observed objects for which direct contact of the researcher with them is impossible due to remoteness, unfavorable conditions for a person's life or technological limitations. For such objects it is typically that the researcher uses specialized intelligent electronic devices, controlled remotely and designed for data collection. Examples are Space scientific instruments, Autonomous research systems exploring the world's oceans, Nuclear reactor monitors, Medical robotic systems, and other scientific equipment operating autonomously and intended for control of activities of great responsibility. Several basic stages (for the acquisition and analysis of scientific data from remote objects) can be pointed out, each of which involves the realization of specific information processing. This article discusses scientific experiment infrastructure, communication methods, specific noise proofing protocols, and relevant data structures that ensure the execution of these stages.

Keywords: Remote Objects, Scientific Data Processing, Noise proofing protocols, Data structures

Introduction

By working with remote objects the researcher has no possibility of direct contact with the physical processes being studied and uses as intermediary instruments remotely controlled intelligent electronic devices, specially designed for data collection. Examples of such devices are Space scientific instruments, Autonomous research systems exploring the world's oceans, Nuclear reactor monitors, Medical robotic sensors systems and other scientific equipment operating autonomously and designed to control activities of great responsibility. Several stages can be identified for the acquisition and analysis of scientific

data from remote objects, each of which involves the realization of specific information processing: 1) Measurement with specialized equipment (Sensor units) of the parameters of the physical processes; 2) Communication between the Sensor Units and the Concentrator and corresponded Data structures, used for archiving of the measurements; 3) Communication of the concentrators with an external memory device, ensuring the data integration into file structures; 4) Preparation of communication records, which upon request from the remote Operation station are sent by radio or other channels to it; 5) The Operation station restores the integrity of the received information and stores it in a database maintained by it. It also includes a set of software tools that allow the visualization of scientific data, trend tracking, and reporting of extraordinary events. The researcher has the capabilities for remote control of the specialized measuring instruments, as well as the individual monitoring of their work.

As 3, 4 and 5 was presented from the authors earlier in [1] and [3], this article focuses on stages 1 and 2 only. It includes next parts:

1) Autonomous measurement units (Sensor units) and Concentrators (Acquisition control unit -AcU) architectures;

2) Data structures used in the Sensor units for archiving of the measuring;

3) Transfer protocol data structures;

4) Low level data transfer protocol;

5) High level data transfer protocol;

6) Data structures used for archiving in AcU;

7) Examples.

1. Autonomous measurement units (Sensor units) and Concentrators (Acquisition control unit -AcU) architectures.

By definition, the remote objects aren't accessabled for the researcher because of which the research tasks are performed by a group of autonomous microcontroller systems, having different specialized equipments for monitoring and studying the observed processes. These systems, named Sensor units, are sophisticated devices, that have different hardware and software implementations. As they are exploring in different aspects the properties of the remote object for the researcher it is important to receive integrated information of their work. In this way, he/she has the opportunity to detect the interrelations between the events taking place in the studied object. The data representation and its semantic can be different for each one of the Sensor units. For the integration of the data streams from Sensor units associated with observed

remote object special control block is used, named AcU (acquisition control unit). It funcks as Concentrator of data of the group of Sensor units. Other its functions are unified communication with the Sensor units and powering of them. AcU does that by using wires connectons with each of the Sensor unit. For communication between AcU and Sensor units noise protected, opto isolated physical interface (typically RS 422) and coresponded logical protocol are used. An example of a block diagram of Sensor unit, used as Langmuir's probe in the Space plasma research is shown on Fig 1.

On Fig. 2. an example of a block diagram of AcU is shown, servicing group of Sensor's units. For special cases, AcU offers processing of the Sensor unit analogue signals by the own 16-bits optoisolated ADC. Typically, the communication between AcU and Sensor units is implemented with RS422. AcU is connected to the remote BSM (Base Storage Module) using wireless interface (radio, infra sound or others). The power Supply of AcU is autonomous (battery, Solar Panel or others own sources).

The comunication between Su (named „experiment" and AcU is realesed using noise protected logical protocol, named Data transfer protocol. This protocol is build as two levels messages between AcU and Su:1) Low level Data Protocol; 2) High level Data protocol.

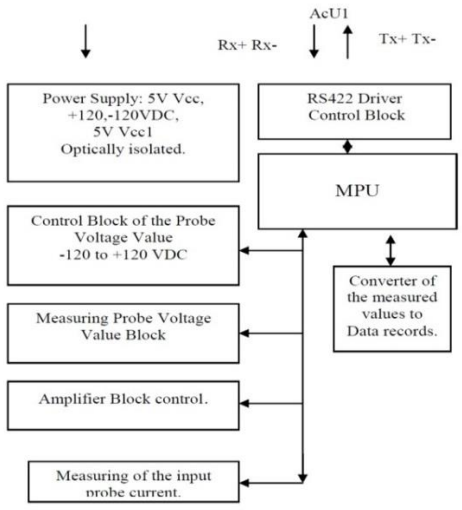

Figure 1. Example of Su block diagram.

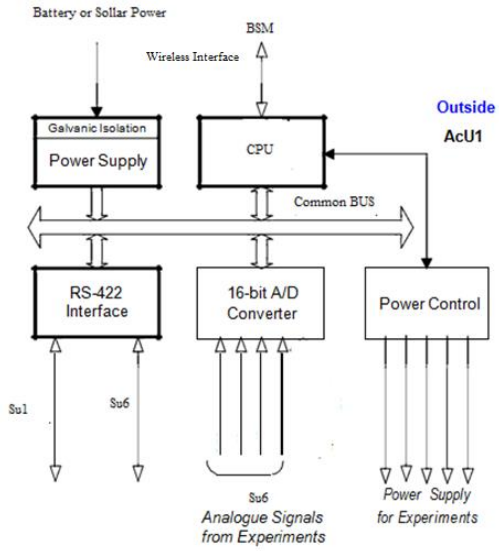

Figure 2. An example of AcU block diagram.

2. Data structures used in the Sensor units for archiving of the measuring.

After being powered and received a command from AcU each Sensor unit is starting own experiment and producing data as its results. The data is recorded in circular buffer with two pointers: "In" and "Out". The "In" pointer points to the place where the results have to be written and "Out" points to the place from where the old results have to be read and sent to AcU. After each action each of two pointers are incremented. If the value of "Out" is equal to this of "In" the buffer is empty and the communication terminates. When the value of “In" overlaps the value of "Out" the recording of data continues. Old data are lost.

3. Transfer protocol data structures [4].

The communication between the Sensor units and AcU is performed by exchange of messages with different length. The message structure is shown in Table 1.

Table 1. Messages structure

Byte Index	Description	Value In Hex
1	Start of Header **SOH**	01h
2	Number of Data Bytes – 1 (k)	k=0…255
3	Message (Packet) Type	
4	Start of Text **STX**	02h
5	1st Byte of Data	High byte in word case
6	2nd Byte of Data	Low byte in word case
.		
.		
.		
N-1	Last byte of Data	N=k+5
N	End of Text **ETX**	03h

After receiving and evaluating the last byte of data transfer (ETX) the receiver acknowledges the block with sending one acknowledge (ACK) byte. A negative acknowledge (NAK) byte forces to repeat the transfer from the user side. A NAK will be sent after a timeout 1 sec when a receiver error (parity or overflow) occurs, or if the number of data bytes between STX and ETX doesn't correspond to the byte-count in the header, or when the transfer is not completed in 1 second. To bring a broken communication alive, the transfer will be repeated if none acknowledgement arrives in 2 seconds.

Table 2. Message Acknowledge bytes

	Signal	Value in hex
	Acknowledge **ACK**	06h
	Negative Acknowledge **NAK**	15h

The Message Type byte structure is shown in Table 3. The TBD is „not protocol defined"-free for defining of the application code.

Table 3. The message type bytes

Sender	Code	Function	Data
AcU	01h	Request for housekeeping	Time and date
Experiment	01h	Housekeeping, no science data	Housekeeping data
Experiment	81h	There is science data to send	Housekeeping data
AcU	02h	Request for science data	None (one 00h)
Experiment	02h	Last science data	Science data
Experiment	82h	There is more science data	Science data
AcU	04h	Set operation mode	TBD
AcU	08h	Send parameters	TBD
AcU	…	…	TBD
Experiment	…	…	TBD

4. Low-level data transfer protocol

The simple Binary Synchronous Communication like protocol ensures the transparent transfer of Experiment binary data. It is named Low level protocol. It strictly follows the execution of the requirements for the messages structure and timing conditions described in Table 1. It is true syntax but not a semantic protocol. Using messages type bytes with codes 04h and 08h (TBD) it allows for AcU to control operations individually from Sensor unit, building a base of realization of High-Level data transfer protocol with it.

5. High-level data transfer protocol

High-level data transfer protocol is semantic depended protocol, using the framework of the Low-Level protocol. The messages' field "Data" consist of command code, parameters or House Keeping values, when the sender is AcU. When the message is sent from the Sensor unit, in this field the answer of the unit (or part of it) or House keeping data sent to AcU is placed. Typically in the embedded text the length of the answer and CRC8 control byte are included. The initial scenario of High-Level protocol is:as follows

•After powering up an experiment, the AcU waits some minutes ~300 sec. After this delay it requests for the housekeeping data. It sends the time and date in this request message. The time format is GMT.

•The experiment answers with a housekeeping data message. This message informs the AcU, if science data are prepared to be sent. In this case the AcU requests the science data. The science data message informs the AcU, whether the Experiment has more data to send, or not.

•The AcU repeats this communication sequence in every 10 seconds. The most significant bit in the message type byte is set, whenever the Experiment has more science data to send. Further message types can be defined. For example: download a new program code, execute calibration, etc.

In the following example the AcU requests for housekeeping data. Because of a parity error a transmission is repeated. There is one transmitted science packet.

Table 4. Example of High -Level Transfer Data Protocol.

Sender	Function	Length	Data in Hex
AcU	Request for housekeeping	N	01 N-6 01 02…time and date … 03
Exp.	Acknowledge	1	06
Exp.	Housekeeping, science exists	N	01 N-6 81 02…housekeeping…03
	After 1 s delay		
AcU	Negative acknowledge	1	15
Exp.	Housekeeping, science data	N	01 N-6 81 02…housekeeping…03
AcU	Acknowledge	1	06
AcU	Request for science data	6	01 00 02 02 00 03
Exp.	Last science data	N	01 N-6 02 02…science data … 03
AcU	Acknowledge	1	06
	After 10 s		
AcU	Request for housekeeping	N	01 N-6 01 02…time and date … 03

6. Data structures used for archiving in AcU.

AcU is working as intermediater between the remote BSM and “experiments” of the controlled group, using for wireless communication with BSM a data buffer having a structure, shown on Figure 3.

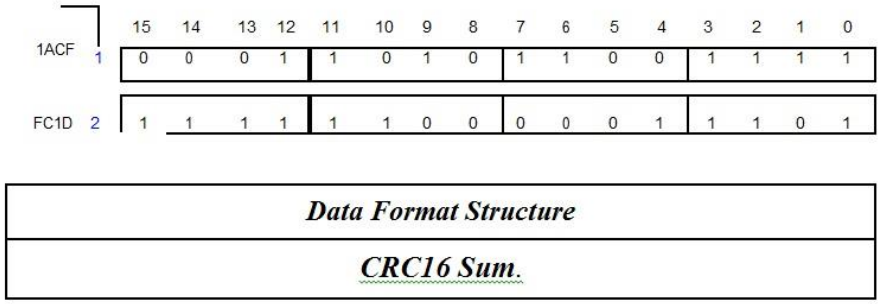

Figure 3. Data Structure used for wireless communication between AcU and BSM.

Table 5. Data Format Structure

Name	Length in bytes	Description	Remark
PID	2	Process ID	(experiments number * 16) + 0x880C or 0x980C
Seq	2	sequence count	increasing value | 0xc000
Len	2	data length	length of data bytes – 1, max value is 260
Sec	1	seconds	seconds in time stamp: value = 0 – 59
Min	1	minuets	value = 0 – 59
Hour	1	hour	value = 0 – 23
Day	1	day	value = 1 – 31
Month	1	month	value = 1 – 12
Year	1	year	value = 0 corresponds to year 1900
Type	4	packet type	always = 0x30000020
HK/Sc	1	housekeeping	housekeeping = 0. science data = 0x20
SID	1	structure identification.	values are described by the experiments
Data	261	Payload	number of used bytes = len + 1

Each field of the structure from Table 5 is filling from the „experiment” answers received from AcU or from the received message sent from BSM to AcU (to start new command). The messages between AcU and BSM includes at the begining four synchronizing bytes 1ah, cfh, fch and 1dh. Each message is ending with control sum CRC16 of the fields from Table 5.

7. Examples

The above presented architecture, data structures, protocols and science data processing are experimented in Space projects, but all of them could be used successfuly as project solutions in other activities, where the researcher's access to the observed processes is technologicaly impossible. One example of such applicatons are medical robots, used in the laparoscopy [2]. These devices includes a system for positioning of a number of laparascopy instruments. As these instruments are working into troacars, after they are placed in their initial positions they are functioning autonomously, controlled from own controllers under a coordination of common main computer, connected to Operator station. Different instrumens can do differnt tasks: diagnostic, therapeutic, special functions. The operator is programming them with a corresponded program, using the Operator station,from where he/she has the abilities for monitoring and control of their work.

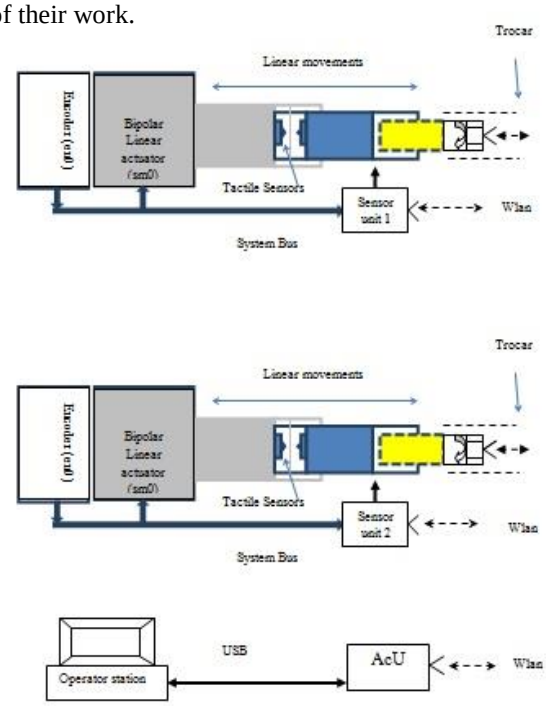

Figure 4. An example of an architecture of Laparoscopy system.

On Figure 4 a laparascopy system is shown, which includs two laparoscopycs instruments, controlled by Sensor unit 1 and Sensor unit 2. Each of the instruments has two types embeded sensors: tactile and force sensors. In the construction of the instruments electromechanical devices for realizig of their programmed movements are included. In the system a controller AcU, coordinating of the executions of the tasks of each instrument, and connected to the Operator station (last is functioning as BSF) are included. The communication between

Sensor unit 1, Sensor unit 2 and AcU is wireless, using the WLAN created from these devices. The communication between AcU and Operator station is with wires, (by using USB). For realizing of the control and monitoring of Laparoscopy system the described above protocols and data structures are used.

Conclusions

Scientific exploring of remote objects creates serious challenges for researchers. The need to work with autonomous devices physically close to the processes studied, the coordination and synchronization and management of the work with groups of such devices, the transmission of the collected integrated information to remote archiving environments and the subsequent processing of the primary data implies the use of noise-proof logical protocols , as well as robust communication channels that provide data credibility. The methods developed for successful telemetry in Space projects, can be successfully used in applications where the researcher's access to the studied processes is technologically limited.

References:

[1] Bachvarov Dichko, Ani Boneva, Bojan Kirov, Yordanka Boneva, Georgi Stanev, Nesim Baruh, Primary information preprocessing system for LP, DP devices -project "Obstanovka", Internacional Conference, BdKCSE'2014 – Big Data, Knowledge and Control Systems Engineering , November 5, 2014, Sofia, Bulgaria, *Proceeding* , pp. 65- 74, http://css.iict.bas.bg/eLbrary/BdKCSE'2014_Proseedings.pdf

[2] Ivanova Veronika, Dichko Bachvarov, Ani Boneva, Smart Surgical Instruments for Robots, International Conference"Robotics & Mechatronics and Social Implementations"2018, 28.08. – 01.09.2018, Varna, *j. Scientific Announcements*, ISSN 1310-3946/Year XXVI, Volume 4, Issue 225, Of The Scientific – Technical Union on Machine - Bilding, August 2018, Sofia, pp. 98 – 103

[3] Batchvarov D., B. Kirov, A. Boneva, R. Krasteva, S. Klimov, K. Georgieva, Software Package for Primary Processing of Telemetric Information, Tenth Jubilee National Conference wich International Participation Dedicated to the 70th Anniversary of Acad. Dimitar Mishev, Contemporary Problems of Solar-Terrestrial Influences, *Proc.* ISBN 954-91424-1-8, Editor: Acad. Dr. Stoycho Panchev, 6. Session "Space Instrumentation and Technologies"-SIT, pp. 202-205, STIL-BAS, Sofia, (20-21 November 2003)

[4] Nagy János, Sándor Szalai, Kálmán Balajthy, István Horváth, Gábor Endrőczi, OBSTANOVKA CONTROL and DATA ACQUISITION UNIT, *Reference: BSTM-DACU-05*, Issue: 2 Rev: 4, 8/29/2013, KFKI-RMKI, pp. 1-191

GDPR and Regulatory-based Approaches for Protecting Organization's Information.

Gaidarski, I., Kutinchev, P.,
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences
i.gaidarski@isdip.bas.bg, kutinchev@isdip.bas.bg

Abstract: The paper present our approach for regulatory based protection of the sensitive information in organizations. The General Data Protection Regulation (GDPR), also known as EU Regulation 2016/679, is designed to strengthen and unify data protection for the personal information of all individuals ('data subjects') within the European Union. GDPR provides the context, guiding principles, and governance framework for collecting and processing personal data of data subjects within the EU. To reach compliance with the regulation, our approach is based on creating of component-based architecture framework for ISS, conceptual models for data protection and implementation with COTS IT security products as Data Leak Prevention (DLP) solutions. Our approach is data centric, which is holistic by its nature to protect the meaningful data of the organization.

Keywords: Data protection, Sensitive data, Personal Information, Data governance, Holistic Approach, Data Leak Prevention, DLP, GDPR

1 Introduction

Today, the information is one of the most valuable assets. The information provides competitive advantages for the organizations and is one of the crucial factors for good performance on the market. It includes customer data, legal and financial records, know-how, patents, day-to-day correspondence – they are from vital importance.

Recent years, organizations from different sectors and sizes has become victims of data breaches and loss of sensitive data. These incidents cost millions and can damage the brand and reputation of the organizations, even take them out from the market.

The data loss incidents are different – leakage of trade information and customer lists, sale of customer data to external parties, loss of USB sticks, laptops or mobile devices. Part of them are caused from outside hacker attacks, but the majority of them are results from internal users or trusted third parties actions [1].

Security deals with the protection of assets, taking associated risk into account. Each asset must be considered in the context of its value and the associated risk of its loss. The main goal of Information Security is the protection of information in all its forms. The protection of data must be considered as one of the highest priorities for every organization. To accomplish this, a number of IT Security controls must be implemented, combining strategic, operational and tactical measures.

There are some basic principles of security, as the three Ds and the CIA triad.

Three Ds concern three universal security aspects - Defense, Detection, and Deterrence:

- Defense reduces the possibility of successful asset compromising;
- Detection is the process of identifying the threats to the assets;
- Deterrence concerns all measures that are helpful for reducing the frequency of compromises of security.

The CIA triad consists from Confidentiality, Integrity, and Availability:

- Confidentiality restricts access to the information only to authorized systems, devices, application or persons;
- Integrity assures that the data in the system has not been altered in an unauthorized way;
- Availability refers to the assurance that the relevant service will be available always when it's needed [1].

The CIA triad represents an open model and additional aspects can be added - Safety, Authenticity, Accuracy and others [2].

Information security (IS) activities concern compliance with certain legislative regulations and industry standards, necessary for its normal operation. Some of them are IS standards such as ISO 27000 [6][7], ISACA's COBIT [8][9], NIST "800 series" [10], sector-specific regulations - the Gramm-Leach-Bliley Act (GLBA) [11] for the financial sector, Sarbanes-Oxley Act (SOX) [12][13] for US public companies, Health Insurance Portability and Accountability Act (HIPAA) [14] and Payment Security Industry (PCI) Data Security Standard (DSS) [15] for credit card operators.

These standards and regulations incorporate the most important aspects of IS, but they are rather set of static good practices and don't correspond to the dynamic of the processes in the organization and changes in the environment, including the requirements of the new regulations - for example recently adopted regulation EU GDPR [16] for protection and processing of personal data of EU Citizens.

The paper outlines a regulatory based approach for protecting the information in the organization, based on creating of component-based architecture framework for ISS, conceptual models for data protection and implementation with COTS IT security products as Data Leak Prevention (DLP) solutions. Our approach is data centric, which is holistic by its nature to protect the meaningful data of the organization.

2 EU GDPR and data protection

The General Data Protection Regulation (GDPR), also known as EU Regulation 2016/679, is designed to strengthen and unify data protection for the personal information of all individuals (‘data subjects’) within the EU. Organizations with over 250 employees (or all organizations wherein processing of personal data is not occasional or includes particular types of sensitive personal data) and that store personal data of those individuals within EU states, must comply with the GDPR, even if the organization is located or operates outside the EU. When it went into effect on May 25, 2018, it replaced the Data Protection Directive (Directive 95/46/EC) of 1995.

GDPR provides the context, guiding principles, and governance framework for collecting and processing personal data of data subjects within the EU. A key focus of the regulation is on the data controllers and processors who manage and execute the processing of personal data. The GDPR highlights expectations of the data controllers and processors to implement appropriate technical and organizational measures to maintain the confidentiality, integrity, and availability of personal data.

The GDPR defines several articles, which concerns the protection of the organization’s data:

- Article 5: Principles relating to personal data processing;
- Article 17: Right to erasure ('right to be forgotten');
- Article 24: Responsibility of the controller;
- Article 25: Data protection by design and by default;
- Article 31: Cooperation with the supervisory authority;
- Article 32: Security of processing;
- Article 33: Notification of a personal data breach to the supervisory authority;
- Article 34: Communication of a personal data breach to the data subject;
- Article 39: Tasks of the data protection officer.

Some of the articles concern legislative procedures, other are people related, and others concern technical solutions. To ensure compliance with GDPR regulation, must be implemented appropriate technical and organizational measures in the organization.
To be able to control the personal data, every organization must take control of all of the data it uses, processes, stores and communicates. The organization must protect the data and to demonstrate that all of the data processing, storing and communication is performed in accordance with the Regulation. [14][22][23]

For effective data protection, we suggest a holistic approach in which we must consider all the possible data - both inside and outside the organization. Holistic approach will ensure that the organization's most sensitive data is identified and effectively protected with combination of security controls and measures. The goal is to protect the organization's data keeping regulatory compliance and continuity of the business process. [20]

3 Conceptual modeling of Information Security Systems

The main purpose of the ISS is to protect and secure the organization's information assets. An important requirement to contemporary methodologies for development of systems including ISS is to ensure the achievement of such system's properties as interoperability, reusability of their components, easy deployment and scalability. The engineering-oriented architecting of systems is not able to guarantee their achievement. Since the top-down approach to systems development is very suitable for this objective, we suggest the usage of model-driven architecting of Information Security Systems. On this basis it is necessary to create a reference methodology for system development and implementation. [19] [20]
Our approach for conceptual modelling of ISS is based on description of the system's architecture. This approach gives us the opportunity for achieving important goals as: interoperability, reusability, easy deployment and scalability, and producing final results – reference templates for building ISS. The approach is based on the framework of architectural description of systems defined in the standards IEEE 1471[3] and IEEE 42010 [4].
According to the standards, the basic concepts which defines the framework of architecture description of a system (Figure 1) are:

- Environment – determined by all influences upon a system that are categorized as concerns. It's the base that contains all domains, in which the system is considered.
- Viewpoint – captures the conventions for constructing, interpreting and modelling a type of view that is in relation to a specific Model Kind.

- View – representation of the system from the perspective of a related set of concerns: each view corresponds to exactly one viewpoint and is addressed to identify system stakeholders and answers their identified concerns. A view is comprised of architecture models.
- Stakeholders – individuals, groups and organizations with interests to the system;
- Concerns – A concern space is formed from the union of all stakeholder concerns [5], [6];

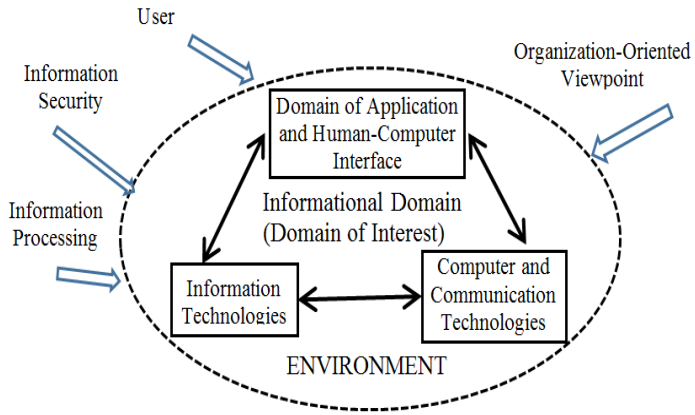

Figure 1. A framework for architectural description of ISS.

To achieve effective data protection, we suggest a multi-layered conceptual model of ISS organized around the viewpoints "Information Security" – first layer, and "Information Processing" – second layer. The stakeholders related to the "Information Security" viewpoint are developers and integrators. Their concerns are conceptual integrity, deployment, scalability, reusability, structure and system properties. The "Information Processing" viewpoint focuses on information processing, semantic of information and relationships between information objects [19]

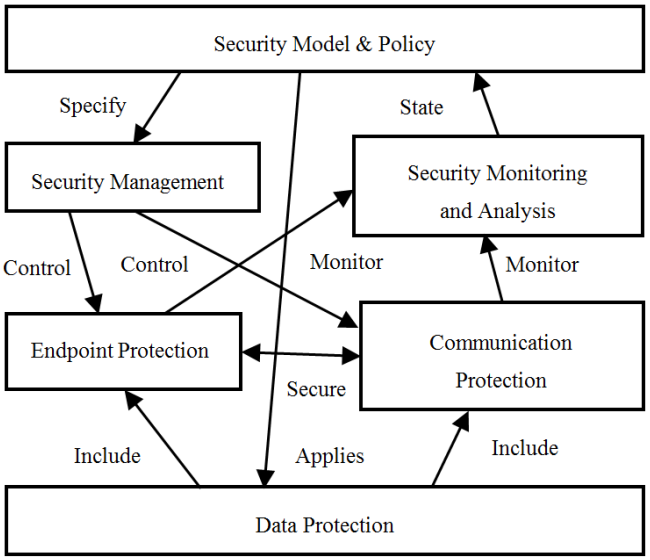

Figure 2. Meta-model of "Information Security" viewpoint.

The main questions to the Information Security Systems are "What", "How" and "Why" to protect.

The suggested meta-model of "Information Security" consists of six components (Figure 2), each of them provides different functionalities [19]:

The second meta-model "Information Security" represents the three types of data states that result from the work of available systems and devices (Figure 3) [19]:

- Data-at-Rest – Inactive data, stored within the IT infrastructure or on media – databases, servers, intranet sites, workstations, laptops, mobile devices, portable storage, removable media, cloud storage.
- Data-in-Use – Active data that is being printed, copied, accessed or used by a system and processed in application.
- Data-in-Motion – Data transmitted by the networks or endpoints.

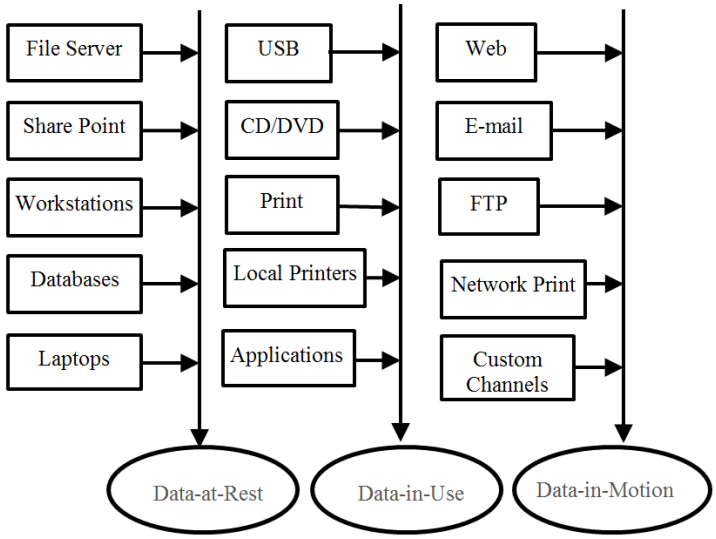

Figure 3. Meta-model of "Information Processing" viewpoint

In order to protect the specific types of data in a system, it is also necessary to implement specific Information Security Techniques (IST) in the basic components from the "Information Security" meta-model. The data have to be protected against loss, theft, unauthorized access and changes by applying IST such as confidentiality controls, integrity controls, access control, isolation and replication [1].

Figure 4 shows a multi-layered conceptual model of ISS, which contains the meta-models representing "Information Security", "Information Processing" viewpoints and the relations between them:

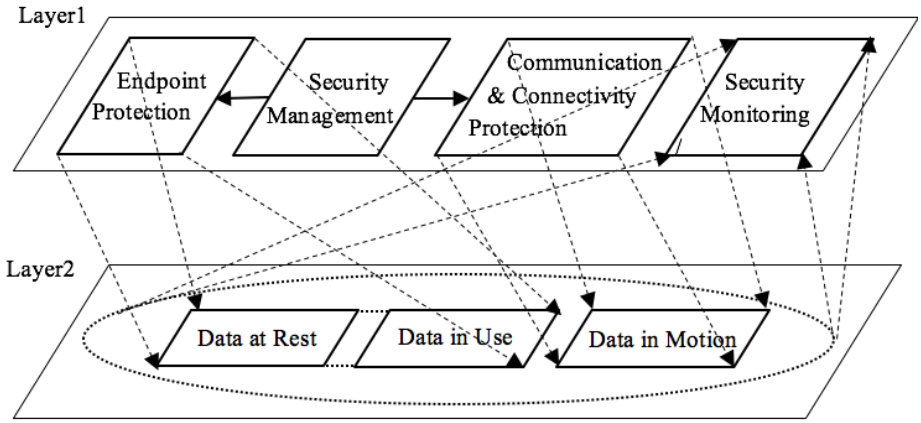

Figure 4. Two-Layered conceptual model of ISS.

- Endpoint Protection component protects “Data-at-Rest” and “Data-in-Use” of the endpoints through relevant IST, as access control, passwords, antivirus, audit trails, physical security measures, etc.;
- Communications Protection component protects “Data-in-Motion”, using cryptographic techniques, network segmentation, perimeter protection, gateways, firewalls, intrusion detection, network access control, deep packet inspection and network log analysis;
- Security Management component protects all configuration, monitoring, and operational data, using cryptography.
- Security Monitoring & Analysis component is responsible for the protection of the data for current system state, the monitoring of key system parameters and indicators. The typical ISTs in this component are cryptographic techniques.

4 Data protection tools as part of the ISS

Although security solutions like firewalls, unified threat management, intrusion detection/prevention systems can detect and/or check any threat to an organization, these tools are not effective when it comes to data-specific approaches. This is where Data Leak Prevention (DLP) solutions for dedicated data protection comes into the picture.

Data Leak Prevention systems are designed to prevent attempts to steal, modify, prohibit, destroy or obtain unauthorized access and usage of the data, without interrupting normal business processes. DLP can significantly reduce the leak of sensitive information, resulting from internal threats, as human errors, intentional actions or outside breach. The main goal of DLP is to stop the data before it leaves protected environment of the organization. DLP solutions can provide very useful information for process of protection of the information:

- Identification of the violations, threats, risks and vulnerabilities to the data,
- Violations of security policies and procedures,
- Discovering and identifying all sensitive information in the organization.

DLP systems can be different types, depending on their focus area:

- With focus on servers, global communications and data channels of the organization. The DLP can control email servers, file transfers from file servers, and Internet traffic filtering;
- Focused on endpoints and local data channels – workstations, laptops, mobile devices - tablets and phones. Controlled channels include all possible physical ports, personal

emails, file transfer to cloud services and more – as DeviceLock DLP Suite, shown on Figure 5 [19][20][23].

The implementation of DLP in organization can effectively protect the sensitive information from inside and outside threats and additionally brings the following results – Figure 6 [19]:

- Reducing the sensitive information leak incidents;
- Limiting data leak channels;
- Increasing the visibility of sensitive information, by the discovery function of the DLP (Data-in-Rest);
- Improving compliance with the internal security policies, legal regulations and privacy directives;

Figure 5. DeviceLock Endpoint DLP Suite.

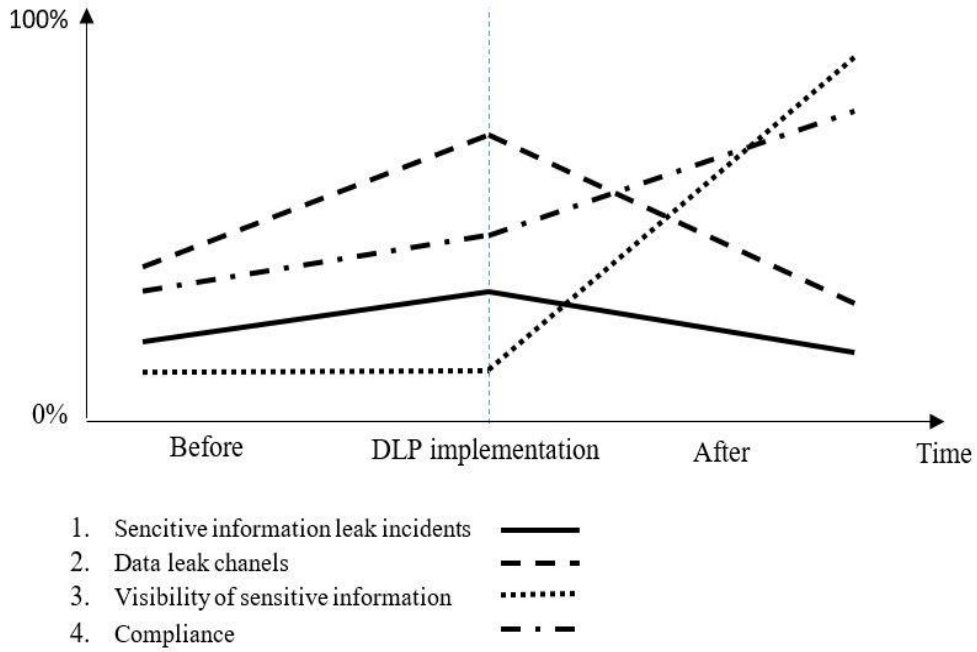

Figure 6. Generalized results from DLP implementation.

5 Conclusion

In this paper we presenting an architectural description of Information Security Systems (ISS) with conceptual models for protecting the data in the system. The goal of architecture model construction is to ensure scalability, usability, interoperability, easy deployment and ease-of-use templates of ISS across organizations of different types and sectors of work. Using this approach allows us to achieve model independence when changing external or internal conditions. As the data protection is one of the major components of ISS, we review specific data-centric protection tools, such as Data Leak Protection (DLP) systems and their deployment as part of the ISS. One of the biggest advantages of the DLP systems is their capabilities to protect data according to the behaviour of different users and to achieve compatibility with multiple security standards, internal rules and regulations, such as EU GDPR in their data protection and notification for data breaches part.

The rules and limitations, set by the EU GDPR can be set as policy in the proposed conceptual model and the corresponding implementation, resulting in compliance with the regulation, protection from data leakage and real-time alerts/notifications for data breaches.

6 Acknowledgement

This study is partially supported by a research project grant “Modelling the Architecture of Information Security Systems in Organizations”, Ref. No: 72-00-40-230/10.05.2017, ICT Sciences & Technologies Panel, Program for Young Scientists and PhD Students Support – 2017, Bulgarian Academy of Sciences.

References

[1] Rhodes M., (2013). Information Security The Complete Reference, 2nd Edition, The McGraw-Hill.

[2] Ernst and Young (October 2011). Data Loss Prevention - Insights on governance, risk and compliance, DLP conceptual model_EY_Data_Loss_Prevention.pdf

[3] Ernst and Young (June 2011).The evolving IT risk landscape - The why and how of IT Risk Management today.

[4] Hintzbergen, J, Hintzbergen K. (2010) . Foundations of Information Security Based on ISO27001 and ISO27002. Van Haren.

[5] ISO 27001 official page: https://www.iso.org/isoiec-27001-information-security.html, May 2017

[6] IT Governance Institute. COBIT Security Baseline: An Information Survival Kit, 2nd ed. IT Governance Institute, 2007

[7] COBIT resources: http://www.isaca.org/COBIT/Pages/default.aspx, May 2017

[8] NIST Special Publications (800Series): http://csrc.nist.gov/publications/PubsSPs.html, May 2017

[9] Gramm-Leach-Bliley Act (GLBA) resources: https://www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act, May 2017

[10] Anand, S. Sarbanes-Oxley Guide for Finance and Information Technology Professionals, Wiley, 2006

[11] Sarbanes-Oxley Act (SOX resources: https://www.sec.gov/about/laws/soa2002.pdf

[12] Beaver, K, Herold R.. The Practical Guide to HIPAA Privacy and Security Compliance, 2nd ed. Auerbach, 2011

[13] PCI Security Standards: https://www.pcisecuritystandards.org/pci_security/, May 2017

[14] EU General Data Protection Regulation official page: http://ec.europa.eu/justice/data-protection/reform/index_en.htm, May 2017

[15] IEEE Std 1471, IEEE Recommended Practice for Architectural Description of Software-Intensive Systems (2000).

[16] ISO/IEC/IEEE 42010:2011 - Systems and software engineering - Architecture description Iso.org. 2011-11-24. Retrieved 2013-08-06

[17] Hilliard R., Aspects, Concerns, Subjects, Views. OOPSLA'99 Workshop on Multi-Dimensional Separation of Concerns in Object-Oriented Systems

[18] Tudor, Killmeyer J., (2001) Information security architecture: an integrated approach to security in the organization, CRC Press LLC.

[19] Gaydarski I, Minchev Z., (2017) Conceptual modeling of information security system and its validation through DLP systems, 9th International Conference on Business Information Security (BISEC-2017), 18th October 2017, Belgrade, Serbia

[20] Gaydarski I, Kutinchev P., Andreev R., (2017) Holistic Approach to Data protection - identifying the weak points in the organization, International Conference on Big Data, Knowledge and Control Systems Engineering (BdKCSE'2017) 7-8 December 2017, Sofia, Bulgaria

[21] DeviceLock www.devicelock.com/products/, August 2017

[22] AlienVault www.alienvault.com/, November 2018

[23] Infinity Ltd, www.infinity.bg, November 2018

Open Source Data for Developing Desktop Application as Part of eOUTLAND Project

Stefan Stefanov
Institute of Information and Communication Technologies-Bulgarian Academy of Sciences
Acad. Georgi Bonchev bl. 2,1113 Sofia, Bulgaria
e-mail: stefans.stefanov303@gmail.com

Abstract: In the framework of the cross border project eOUTLAND as part of the Greece-Bulgaria Intterg Program, support activities have been funded for a new approach idea in favor of volunteer and firefighter teams. The scope is on the development of the field skills in both types of professional and non-professional individuals. As part of Bulgarian Academy of Sciences (BAS) activities is creation of field risk maps in cases of wild land fires and flood events. For a better presentation of the available field data about those dangerous natural hazards we have decided to build up a web-based open source system applicable in operational room as desktop visualization tool of the needed data. In our paper we will describe all free sources which were used for description of the test areas located on the territories of Zlatograd forestry department (covering Zlatograd, Madan and Nedelino municipalities) as wild fire prone areas and Svilengrad municipality as flood prone area. This information is important, because every involved team in suppression measures needs data from the field and they also send such data for the field. In this cases GIS tools can support in the best possible way their orientation and field activities.

Keywords: Open source tool, Open source data, eOUTLAND, GIS.

1 Introduction

We are using free data and open source software solutions for developing our application as part of eOUTLAND project for support of volunteer and firefighter groups in cases of wildland fires and floods. The free sources of data and the open source software solutions that we are using for developing our desktop application give us a lot of opportunities and possibilities to build application that can be used in the operational rooms in support of volunteer and firefighter teams in cases of wildland fires and floods. The application can visualize Points of Interest (POIs) which this people need to know like logistic centers for water supplies and

firefighting tools, which are usually well done instructions on a paper documents. We are going to use basic layers from these data sources that will visualize the needed information in a way that can be easily understood by users of the desktop application. We are using several sources from where we are getting the data. These sources of free data are described in brief in the following sections.

First source of data that we are using is the Natura 2000 [1] (Figure.1) official web site from where we used the free delivered layers with the boundaries of areas that we are going to test.

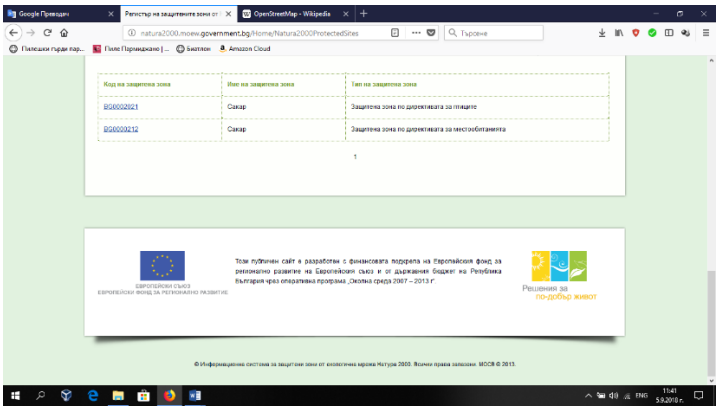

Figure 1. Natura2000 official web site

The second source of data that we are using is OpenStreetMap (OSM) [9] (Figure.2) which obtain highly detailed free GIS (Geographic Information System) data with different levels of accuracy and completeness. OpenStreetMap is open maps data source and is built by a community of mappers that contribute and maintain data about roads, trails, cafés, railway stations, and much more, all over the world. OpenStreetMap powers map data on thousands of web sites, mobile apps, and hardware devices. Some OpenStreetMap data is supplied by companies that choose to freely license either actual street data or satellite imagery sources from which OSM contributors can trace roads and features. OpenStreetMap uses a topological data structure, with four core elements (also known as data primitives): nodes, ways, relations and tags.

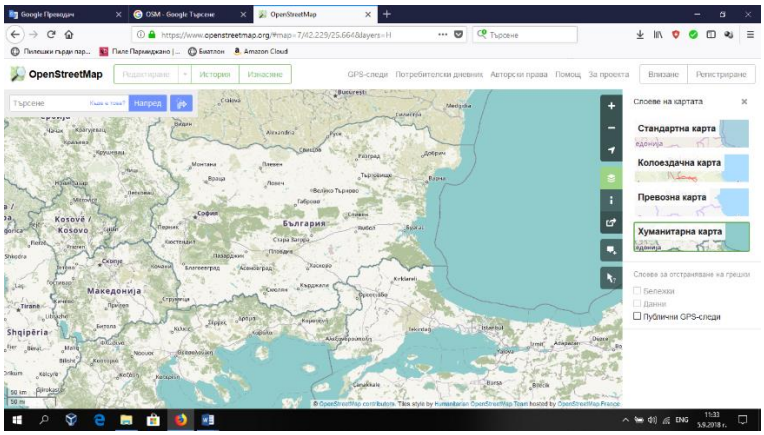

Figure 2. OpenStreetMaps (OSM) GUI (Graphical User Interface)

The third source that we are using is Esri Open Data [10] (Figure 3) which is very rich of free GIS data. It lets the users to explore over 105,628 open data sets from 5,680 organizations worldwide. The formats of data that can be downloaded are in spreadsheet, KML (Keyhole Markup Language), shapefile and API's are OGC WMS (Web Map Server), GeoJSON and GeoService. Esri Open Data is managed by the largest commercial GIS organization in the world.

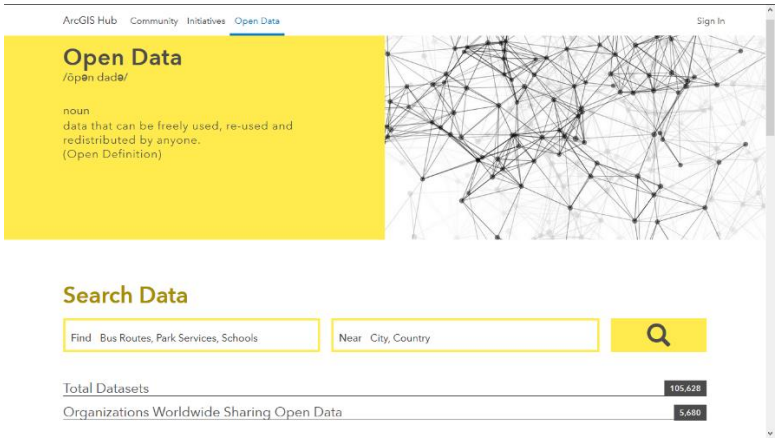

Figure 3. Esri Open Data main page

The Fourth source that we are using is Sentinel Satellite Data. The European Space Agency's Sentinel satellite is now the highest spatial resolution available to the public for free. Sentinels Scientific Data Hub (Figure 4) is a user interface that lets users to control the data easily. The Copernicus Open Access Hub [11] is a web-based system designed to provide

Earth Observation (EO) data users with distributed mirror archives and bulk dissemination capabilities for the Sentinels products.

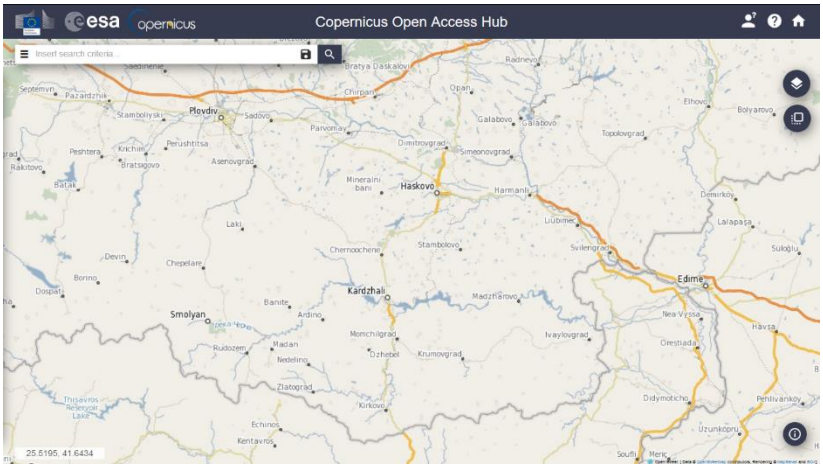

Figure 4. Sentinels Scientific Data Hub user interface

2 Software solutions for building the desktop application

In order to build our application for a Desktop usage we will use free and open source software for processing the data such as: Geoserver, Qgis, Web App Builder and Boundless WEBSDK.

QGIS is a GIS client where users can visualize, manage, edit, analyze data and compose printable maps. It includes analytical functionality through integration with GRASS (Geographic Resources Analysis Support System), SAGA (System for Automated Geoscientific Analyses), Orfeo Toolbox, GDAL/OGR (Geospatial Data Abstraction Library) and many other algorithm providers. It runs on Linux, Unix, Mac OSX and Windows and supports numerous vector, raster and database formats and functionalities.

Web App Builder [7] is a plugin for QGIS that allows easy creation of web applications based on layers, map compositions and bookmarks, as configured within a QGIS platform. The applications can also include additional web services, various controls, and other interactivity.

Boundless WEBSDK [8] provides tools for easy-to-build JavaScript-based web mapping applications. It makes use of the JavaScript React framework (A JavaScript library for building user interfaces) to provide modular components, which can be used to create complete web mapping applications quickly and easily.

Geoserver allows users to process maps and data from a variety of formats to standard clients such as web browsers and desktop GIS programs. Data is published via standard based

interfaces, such as WMS (Web Map Service), WFS (Web Feature Service), WCS (Web Coverage Service), WPS (Web Processing Service), Tile Caching and more. GeoServer comes with a browser-based management interface and connects to multiple data sources at the back end.

3 Test areas

The desktop application which we are developing and implementing is focused on specific test areas. They cover territories of Natura 2000 located in Zlatograd, Madan, Nedelino and Svilengrad municipalities.

Natura 2000 is a pan-European network of protected areas designed to ensure the long-term survival of the most valuable and endangered species and habitats for Europe in line with the major international environmental and biodiversity arrangements. All sites are within the ecological network and are defined in accordance with two of the European Union's key environmental directives.

Zlatograd Municipality [6] is located in southern Bulgaria, at the end of the Eastern Rhodopes, to the last Rhodope hill, after which the Aegean Plain begins. Zlatograd is hidden in the valley of the Varbitsa River. The relief of the municipality is low and medium mountainous and extends in the southwestern part of the Eastern Rhodopes. The area belongs to the continental -Mediterranean climate area, the South-Bulgarian climatic sub-region and the low-mountain climatic region of Eastern Rhodopes. The temperature is continental and prevails in the low mountainous regions, with a marked influence of the Aegean Sea, while in the mountainous areas there is a decrease in temperature, an increase in the amount of precipitation and the wind speed. The Municipality of Zlatograd is characterized by exceptionally preserved flora and founa biodiversity. In forests the most common are broad-leaved tree vegetation - beech, hornbeam, oak, birch and dogwood, and coniferous species - spruce, fir, white and black pine. The animal world in the area is related to the central alpine faunal species.

Madan municipality [5] is located in the eastern part of the Smolyan region. The relief of the municipality is low and medium mountainous and extends to the extreme southwestern part of the Eastern Rhodopes and to a small part of the Western Rhodopes. The Madan region falls in the Continental-Mediterranean area, the South-Bulgarian climatic sub-region, the Eastern Rhodopes and the low-climatic region, and the parts over 1000 m above sea level. in a mountainous climatic area of the same sub region. The Municipality of Madan has intact and hitherto natural habitats for a number of protected plants and animals. In forests the most common are broad-leaved tree vegetation - beech, hornbeam, oak, birch and dogwood, and

coniferous species - spruce, fir, white and black pine. The animal world in the area is related to the central alpine faunal species.

Nedelino Municipality [4] is located in the southeastern part of Smolyan District. The relief of the municipality is low and medium mountainous and extends in the southwestern part of the Eastern Rhodopes. The area belongs to the continental -Mediterranean climate area, the South-Bulgarian climatic sub-region and the low-mountain climatic region of Eastern Rhodopes. The temperate continental climate prevails in the low mountainous regions, with a marked influence of the Aegean Sea, while in the mountainous areas there is a decrease in temperature, an increase in the amount of precipitation and the wind speed. The Municipality of Nedelino is characterized by preserved biodiversity, intact and hitherto natural habitats for a number of protected plants and animals. In forests the most common are broad-leaved tree vegetation - beech, hornbeam, oak, birch and dogwood, and coniferous species - spruce, fir, white and black pine. The animal world in the area is related to the central alpine faunal species.

Svilengrad Municipality [3] is located in the southeastern part of the Haskovo region. At its northeastern side are the slopes of the Sakar Mountain, in the southwest - the Eastern Rhodopes, including parts of the Upper-Thracian Plain. Svilengrad Municipality is located in the Continental-Mediterranean climate area, which determines the continental - Mediterranean climate of the municipality. The most characteristic features of this type of climate are the warm summer and the mild winters, the relatively small annual temperature amplitude, the autumn-winter peak of precipitation and the lack of an annual sustainable snow cover. On the territory of the Municipality of Svilengrad are the protected area "Lozenski Path", which is a natural deposit of a marsh snowdrop and a natural park "Eastern Rhodopes", covering part of the lands of the village of Mezek and the village of Siva reka. In the territory of the municipality is part of Corine site "Sakar", having an international nature conservation status. About 70% of the territory of Southern Sakar is occupied by agricultural land in the place of oak and owl oak forests. Among them about 15% of the total area are dispersed xerothermic grass communities with predominance of whiteness, bulbous grass, savanna, etc. and more rarely mesoxerothermal vegetation. Svilengrad Municipality is inhabited by a large number of rare and protected species. The number of protected reptiles: the Mauremus rivulata, the Testudo graeca, the T. hermanni and the Ophisaurus apodus are among the highest in the Sakar Mountains. Of the birds, 76 species have been identified, 34 of which have an unfavorable conservation status in Europe.

The rich flora and fauna biodiversity of these four municipalities were the reason to be included as test areas in the eOUTLAND project. The desktop application which is built based on them is having its first trial versions and we are presenting the fire zones in municipalities of Zlatograd, Madan and Nedelino as predefined layers presented (figure.5). This part of the desktop application will have the ability to show any active fires on the test area. For this scenario on the picture we are using as base layer open source data from OpenStreetMap (OSM) [9]. However any user has as options also the possibility to choose different types of base layers [12] according to their needs. Any other data provided by the municipalities in GIS format can be also displayed on the application for the test area.

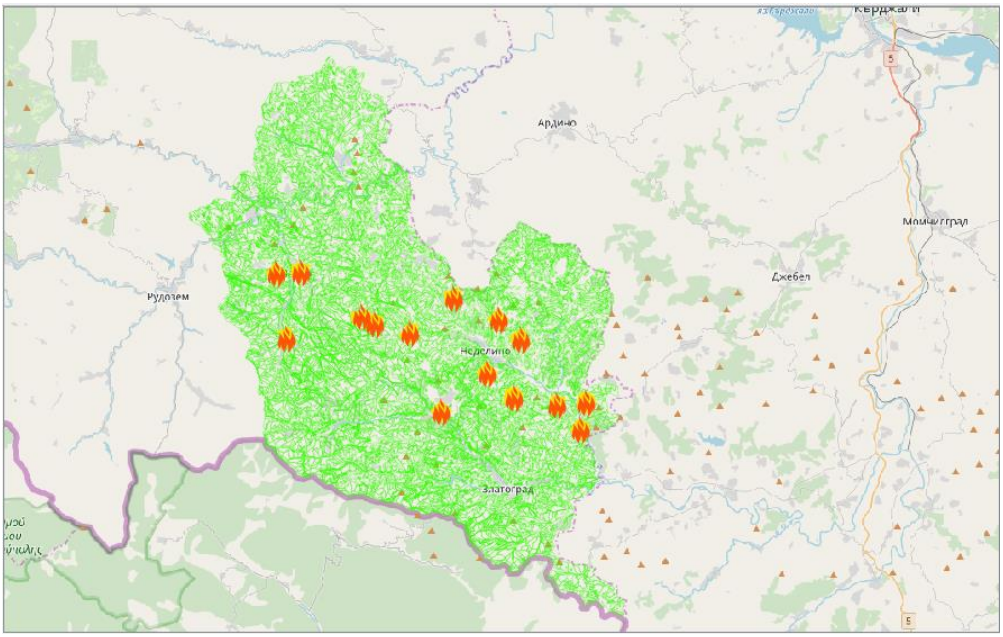

Figure 5. Fire zones in Zlatograd, Madan and Nedelino municipalities

The second part of our desktop application will have the ability to present flood endangered zones located at the area of Svilengrad municipality (figure.6). As base layer again we are using open source data from OpenStreetMap (OSM) [9]. The user needs and requirements can change this [12] according to their needs. Municipality of Svilengrad is a border zone thus any other important data for visualization can be displayed based on the local authorities preferences.

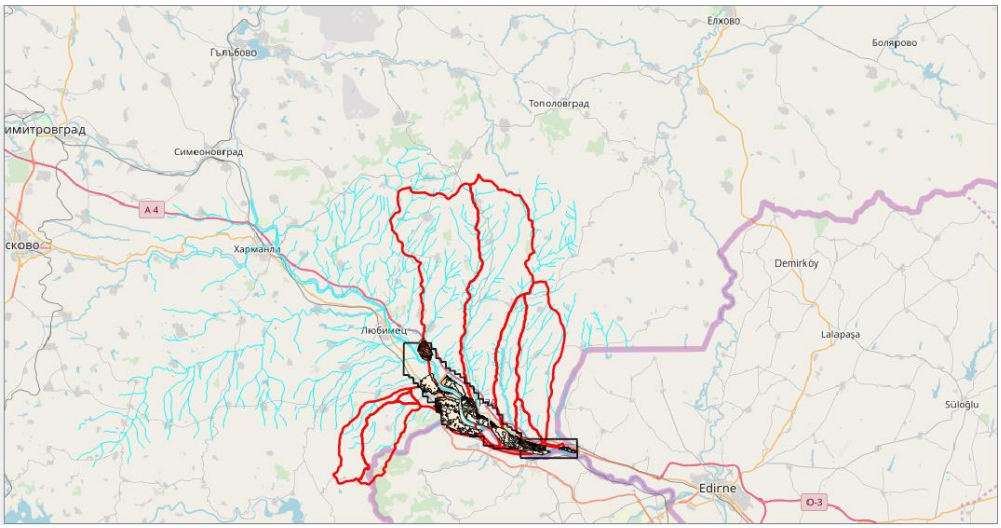

Figure 6. Floods danger zones at Svilengrad municipality

4 Conclusions

This article presents an initial attempt for building of a desktop web based application with the use of an open source data and software solutions only. Its potential users are volunteer groups and firefighting teams that need decision support tools in operational rooms. This kind of ICT solutions are necessary for the growing needs in the context of wildland fires and flood monitoring activities that are in place for Zlatograd, Madan, Nedelino and Svilengrad municipal areas. We are using services, based on the latest GIS technologies used for building this kind of platforms for management of large data sets. This kind of tools also gives opportunity to firefighters and volunteer teams, in fire departments and operational rooms to react faster in cases of wildland fires and floods. The main goal of the application is to deliver information about Points of Interest (POIs) which this groups need to know like logistic centers for water supplies and firefighting tools, without the need to have in their pockets paper instructions and paper maps to fulfill their field work. The desktop application will be tested during the life time of the project eOUTLAND.

Acknowledgement

This work is partially supported by the cross-border INTERREG program Greece-Bulgaria. Special thanks to the scientific adviser of my PhD studies - assoc. prof. dr. Nina Dobrinkova, for the critical notes and suggested improvements.

References:

[1] [Online] Available: *http://natura2000.moew.government.bg*

[2] [Online] Available: *https://scihub.copernicus.eu/dhus/#/home*

[3] [Online] Available: https://www.svilengrad.bg/

[4] [Online] Available: *http://nedelino.bg*

[5] [Online] Available: *http://www.madan.bg*

[6] [Online] Available: http://www.zlatograd.bg

[7] [Online] Available: http://boundlessgeo.github.io/qgis-plugins-documentation/webappbuilder/

[8] [Online] Available: https://sdk.boundlessgeo.com/docs/

[9] [Online] Available: https://www.openstreetmap.org

[10] [Online] Available: http://hub.arcgis.com/pages/open-data

[11] [Online] Available: https://scihub.copernicus.eu/

[12] H.Singh, D. Gambhir, "An Open Source Approach to Build a Web GIS Application", International Journal of Computer Science and Technology (IJCST), Volume 3, ISSN: 2229-4333, p110-p112,2012.

High-speed Semiconductor Laser Diode Driver with Analog Signal Modulation

Svetozar Ilchev *, Zlatoliliya Ilcheva **

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

Acad. G. Bonchev str., bl. 2, Sofia, Bulgaria

* svetozar@ilchev.net, ** zlat@isdip.bas.bg

Abstract: In this paper, we present a high-speed laser diode driver that has a very sensitive analog modulation input. It is designed to be part of the electronics of a laser projection system and it is responsible for the rendering of a color channel with accurate beam intensity during laser animations. The fast response time and the very good sensitivity result in excellent projection quality. Depending on the laser diodes, the laser driver can be adjusted to waste very little power in the form of heat, which enables its application in confined spaces and without an attached heatsink. Its compact size and the careful positioning of the electronic components facilitate the installation into the laser projector enclosure.

Keywords: Laser diode driver, Analog modulation, Laser projector, Laser animation.

1 Introduction

Semiconductor laser diodes are compact laser light sources with a typical optical power output ranging from several milliwatts up to several watts. Some of their application fields and use cases encompass multimedia entertainment (e.g. as part of laser projectors for laser animation shows), industrial production (e.g. as visual guidelines for machine equipment or quality control and human inspection and assembly), advertising (e.g. as projection devices for company or product logos), building automation and management (e.g. as information projection systems for building complexes), etc.

Semiconductor laser diode projection systems have several important advantages over traditional laser systems, e.g. systems based on a CO_2-laser. First, the laser light source is very compact - a laser diode with an optical power of several watts comes in a casing with a diameter of 9mm. Second, laser diodes are cheaper than traditional laser sources, which makes their application economically feasible for civil applications, some of which are mentioned above. Third, laser diodes can be installed with relative ease in target devices as long as good thermal conductivity is guaranteed.

In order for laser diodes to function properly, the following aspects should be observed:

1. Proper current regulation, preferably by means of a well-designed current feedback loop, is extremely important as laser diodes burn out very easily in cases of overcurrent.
2. Heat dissipation should be handled in a suitable manner - either by incorporating a passive heatsink of a suitable size or by combining a heatsink with an active cooling system (fans or Peltier elements).

The laser projection is typically implemented by means of moving mirrors mounted on galvanometer axes. The mirrors deflect the laser beam in horizontal and vertical directions - a principle similar to old TV sets with cathode ray tubes, so that a two-dimensional animation can be created if the mirror movement and the laser beam intensity are kept properly synchronized. The mirrors may support projections with speeds of about 20 to 40 thousand points per second and the laser diode driver must be capable of controlling the laser beam intensity with the same speed in order to provide good animation quality. The modulation of the beam intensity is passed to the laser driver in the form of an analog control signal (see [1]). As modulation sources, a computer or a specialized microcontroller may be used.

Usually, the laser beam is composed of red, green and blue (RGB) color components (channels) and one driver per color channel is used. All drivers are connected to a central ILDA distribution board, which also controls the laser scanners, the shutter and the interlock functionality of the projector.

This paper presents the design and implementation of a high-speed laser diode driver with analog signal modulation, which takes into consideration the important aspects, mentioned above, and implements fast and accurate current regulation while minimizing the heat dissipation. An additional challenge is the need for compact size and easy installation of the driver board inside the enclosure of the laser projector.

The next section reviews some related work. Then, the overall conceptual design of the driver is presented. In subsequent sections, the actual implementation is shown along with some pictures and some experimental results are presented. Finally, the conclusions summarize the insights we have gained and give some directions for future work.

2 Related work

This section presents some previous work related to laser diode drivers.

The author of [2] presents a high-speed laser diode driver for driving relatively low-power laser diodes with a current up to 80 mA. A more powerful laser diode driver is discussed in

[3]. It integrates a switching power converter to ensure high power efficiency and supports laser diode currents up to 2.5A. In [4], the authors present a system for optoacoustic imaging containing a laser diode driver design with a different modulation scheme which performs well for the given application.

The authors of [5] present a compact high-current laser diode driver with good heat dissipation design. In [6], a very high-power 150A laser diode driver is constructed making use of a commercial GaN FET transistor to achieve a laser input power of 4 kW. In [7], the same author presents some variations of the driver to meet the demands of 3D scanning applications requiring currents with maximum values greater than 35A. In [8], another relatively simple driver for laser diodes utilizing two reference voltages of 2.5V is presented.

In [9], a new technology developed by Quantel is presented, which enables the integration of several laser diodes in the infrared spectrum (wavelengths between 800-1000 nm) and their drivers on a single printed circuit board (PCB).

3 Overall conceptual design of the laser diode driver

This section presents a conceptual overview of the design of the laser diode driver. The laser driver is designed to deliver a maximum current up to 5A with sufficient speed to one or more laser diodes connected in series to the laser driver. The conceptual design is shown in Fig. 1.

The laser diode driver 100 consists of the following functional blocks: a protection block 110, a power supply block for the laser diode 120, a power supply block for the remaining electronic elements 130, an analog modulation block 140 and a block for electronic current regulation 150.

The protection block 110 accepts an input voltage ranging from 9V - 28V from an external power supply. The block contains a fuse 111 for overcurrent protection (from 2A to 10A), a transient-voltage-suppression (TVS) diode 112 that protects the driver from momentary voltage peaks of the external power supply and a MOSFET-transistor that is used to protect the driver from accidental polarity reversal of the input voltage, which would otherwise lead to serious damage of the driver and any connected laser diodes.

The power supply block for the laser diode 120 consists of a switching regulator 121 and the necessary supporting elements 122 - 128. The voltage divider 126 sets the output voltage that is passed to the laser diode(s). Ideally, the value of the output should be about 2V greater than the sum of the forward voltages of the connected laser diodes at maximum current. The

switching regulator transforms (steps down) the voltage in a highly-efficient manner - efficiencies of 85-90% are usual. In this way, the power dissipation of the current regulation block 150 (especially the transistor 152) is minimized and thus the overall wasted power in form of heat is greatly reduced. The circuit for emergency power off 128 is used to power off the laser diode in cases of overheating or other emergencies. It receives a control signal from an external microcontroller and it is opto-isolated from the other electronic components.

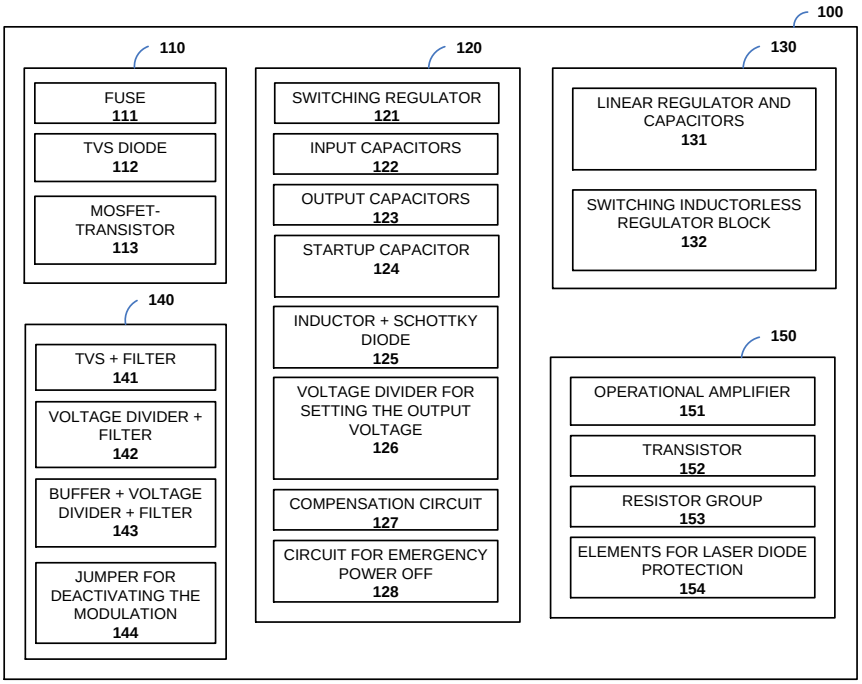

Figure 1. Conceptual design of the laser diode driver

The power supply block for the electronic components consists of a linear voltage regulator 131, and a switching inductorless regulator block 132. It is used to power the analog modulation block 140 and the current regulation block 150. As the power consumption of these two blocks is small (less than 5 mA of current in most cases), the use of the linear regulator does not lead to a significant increase in the working temperature.

The analog modulation block 140 accepts at its input an analog signal between 0V and 5V. First, the signal goes through a TVS diode and a filter 141 followed by a voltage divider and a second filter 142. After them, the signal passes through a buffer 143 made of an operational amplifier unit and another voltage divider and filter. The two voltage dividers are implemented by means of a variable resistor (trimmer or potentiometer). The first voltage divider (e.g. a trimmer soldered onto the laser driver board) is used to limit the maximum current allowed to pass through the laser diode. This is a factory/supplier setting which is dependent

on the model of the laser diode. The second voltage divider (e.g. a potentiometer mounted on the laser projector control panel) is used to provide a convenient way for the end-user of the laser projector to decrease the maximum current through the laser diode when the laser projection is set up or during the projection time.

The jumper for deactivating the modulation 144 connects a +5V DC voltage to the laser modulation input thus causing the maximum current set by the two voltage dividers to pass through the laser diode.

The block for electronic current regulation 150 consists of an operational amplifier unit 151 controling a MOSFET transistor 152. The amplifier makes use of negative feedback loop by measuring the laser diode current which passes through the resistor group 153. Different number of resistors and different resistor values may be used depending on the connected laser diode(s) and the maximum operating current. The elements for laser diode protection 154 may include a bleeding resistor and a TVS diode against electrostatic discharges.

In fig. 2, we show the integration of the laser diode driver 230 into a laser projector 200. The driver is powered over the wires 250 by a separate power supply 210 with a voltage ranging from 9V to 24V DC. The switch 220 is used to switch on the projector. The wires 260 connect the laser diode(s) 240 to the laser diode driver. There may be an individual semiconductor laser diode or a string of such diodes with a total forward voltage less than the voltage of the power supply minus two volts.

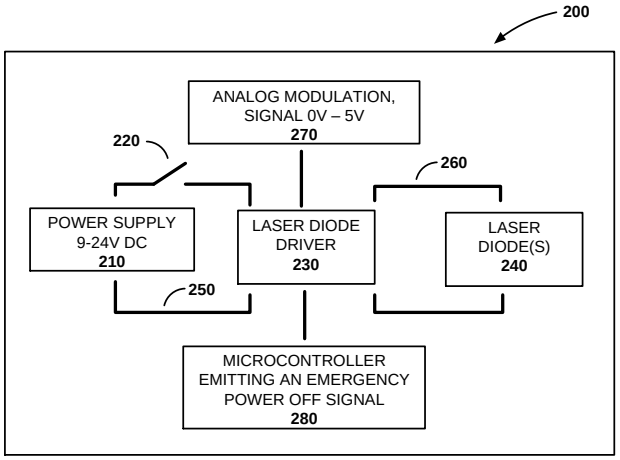

Figure 2. The laser diode driver as part of a laser projector

The analog modulation 270 is an external analog signal source delivering a signal ranging from 0V to 5V that controls the intensity of the laser diode(s). The maximum voltage of the signal source may be lower than 5V, e.g. 3.3V or 2.5V. The microcontroller 280 analyses the

operating conditions of the laser diode, e.g. its working temperature measured by a thermistor. If overheating or other emergency situations are detected, the microcontroller emits an emergency power off signal and shuts down the power to the laser diode.

4 Practical implementation of the laser diode driver

The laser diode driver is implemented in the form of a double-sided printed circuit board (PCB). In fig. 3, we show the front side of a PCB combining the laser diode driver (right half of the board) together with the microcontroller responsible for temperature measurement, cooling and emergency power off (left half of the board). An OLED or LCD display may show the working temperature(s) and important cooling parameters. If only the laser driver is needed for a given laser projector system, the two halves of the board can be physically separated by cutting the board in two in the middle. The two parallel traces visible near the middle bottom part of the board that connect the two board halves conduct the emergency power off signal from the microcontroller to the laser driver.

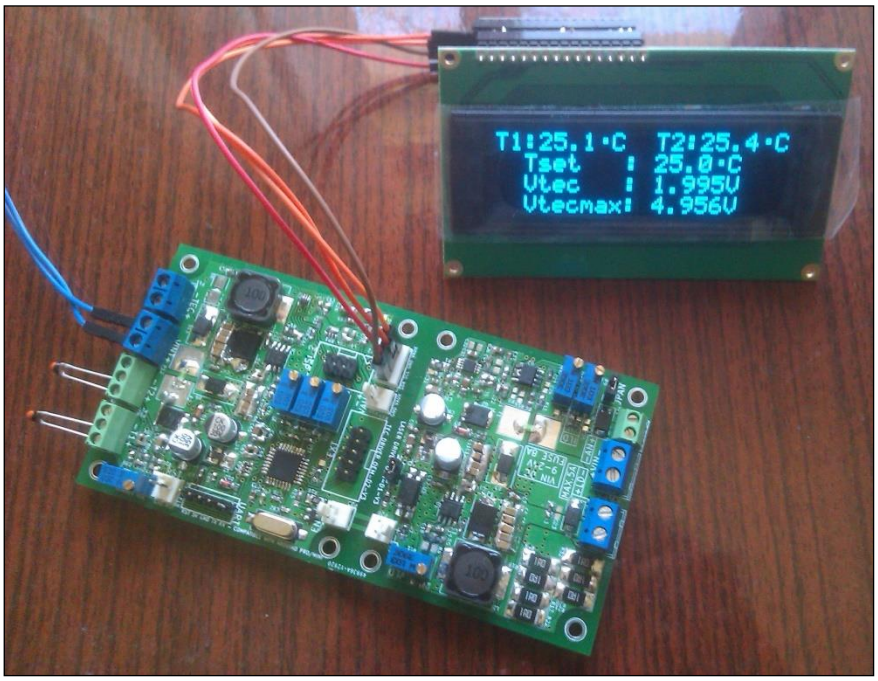

Figure 3. Driver implementation as one half of a double-sided printed circuit board

We use mostly SMD (surface-mount) components so that the assembly is quicker and the size of the board is minimized - the size of the laser driver half is 65x59mm. The blue terminals at the edge of the board are used for the high-current connections of the power supply and the laser diode(s). The green terminal next to them near the upper right corner of the board is used

for the analog modulation signal. The resistor group 153 (7 SMD resistors of size 2512) is visible near the lower right corner.

The back side of the board is shown in fig. 4. The right board half constitutes the laser driver and the left board half is the microcontroller. The electronic elements mounted on the back side of the board are linear voltage regulators and the transistor controlling the laser diode current. They are the components that generate most heat and benefit from the addition of a heatsink or an aluminum, brass or copper plate. The board is designed to be fixed to such a heatsink or plate by means of four to eight M3 screws located at the edges of each board half. The heatsink/plate makes physical contact with the top of the elements mounted on the back side of the board (fig. 4) and improves the power dissipation capabilities of the driver. The hardware design also uses some of the copper surfaces on the back and front sides of the board as "embedded" heatsinks, which allows the board to function well even without an external heatsink/plate provided that the output voltage of the switching regulator 121 is set properly.

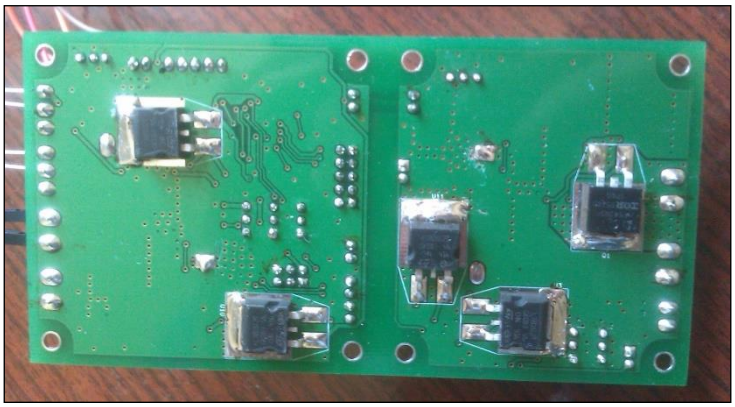

Figure 4. Backside of the laser diode driver

The laser diode driver uses a dual operational amplifier integrated circuit housed in the popular SOIC-8 package. There are many different operational amplifiers (e.g. the very popular LM358) which have the same case and the same pinout, so that, for each use case, we may tweak the current control speed and accuracy by changing the model of the amplifier. The same is valid for the MOSFET transistor and the linear regulators housed in a TO-263 package.

The board is designed to accept a maximum input voltage of 28 volts, which enables the use of laser diode strings. For example, a 2W blue Nichia diode usually has a forward voltage V_f of about 5V, so we may combine four such diodes in a single string obtaining a total optical output power of the blue color channel equal to 8W.

If need be, the blue trimmers soldered onto the board can be replaced by potentiometers connected by wires and mounted on the laser projector control panel to allow the end user to

change the laser driver working parameters. Mostly, this is convenient for changing the maximum current through the laser diode(s), which translates into a change of the maximum intensity of the color channel. In an RGB laser projector, the end user may tweak the balance of the white color by means of three such potentiometers - one for each color channel.

5 Experimental results

Preliminary experimental results show that the laser driver achieves a very accurate intensity modulation, which is both very fast and sensitive to the analog signal input. Fig. 5 shows the output at the laser diode of a square wave with a frequency of approx. 476 kHz generated in a quick-and-dirty fashion by the GPIO interface of a general-purpose microcontroller. This is one of the tougher cases to modulate and the initial results are very promising.

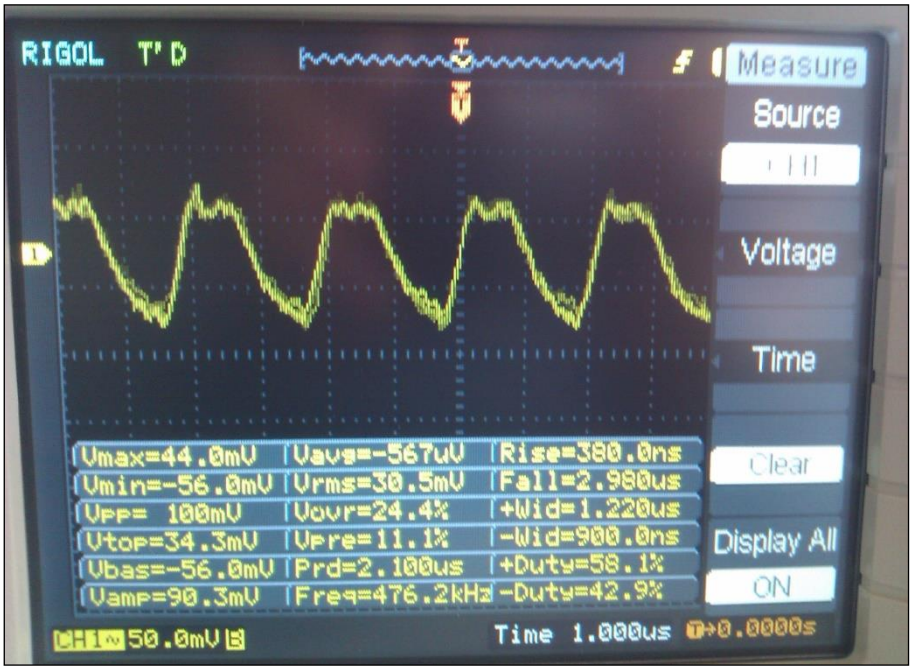

Figure 5. Intensity modulation of a square wave

This modulation result is achieved by using a relatively widespread operational amplifier from Microchip, so that further improvement is possible.

We have made several long-running tests with different currents and animations controlled by a computer. The results show that the laser driver works without any overheating issues or glitches for at least four hours of laser projection at 700mA, 1.5A and 3A with a properly set laser diode voltage. For safety, we have mounted the board on a heatsink, which is not strictly necessary. Of course, the laser diode forward voltage may fluctuate with time

and it usually decreases a little when the laser diode operating temperature increases. This means that over time the MOSFET transistor may have to dissipate a bit more power than initially measured at the start of the animation, so that the external heatsink serves as an additional precaution against these fluctuations or accidental driver misconfiguration.

We have also done tests on a string of 5 red Mitsubishi laser diodes and two blue Nichia diodes which performed to our satisfaction - the laser animations fed from our computer were replicated well and without any visible distortions or artifacts.

Some of our tests were made using 12V quality power supplies by Meanwell and other tests including the tests with laser diode strings were made using 24V power supplies by the same company. The laser diode driver functioned well with all power supplies and was capable to drive successfully all diode combinations.

6 Conclusions

The laser diode driver presented in this paper targets applications in single-color or multi-color laser projectors used for multimedia, industrial or advertising purposes. It is power-efficient, has a small size and can drive almost all commercially available semiconductor laser diodes with their recommended nominal currents. The analog modulation conforms to the ILDA specification [1], functions very well and provides a fine degree of control over the intensity of each color channel. The experimental results show that the laser driver with its high power and fast modulation speed is suitable for both commercial and scientific purposes.

The future work will encompass more practical tests of the driver and an optimization of the choice of electronic components so that the physical board size can be further reduced. In addition, a comparative study of the behavior of different combinations of operational amplifiers and MOSFET transistors may be beneficial to the optimization of the driver for different use cases.

Acknowledgement

This research was supported by the Bulgarian FNI fund through the project “Conceptual Modeling and Simulation of Internet of Things Ecosystems (KoMEIN).

References

[1] ILDA Technical Committee (1999). The ILDA Standard Projector, July 1999 (Retreived on 29.07.2018). URL: https://www.ilda.com/resources/StandardsDocs/ILDA_ISP99_rev002.pdf.

[2] Kabel, M. (2016). High Speed Laser Diode Driver. 10.13140/RG.2.2.30013.15841.

[3] Thompson M., Schlecht, M. (1997). High power laser diode driver based on power converter technology, IEEE Transactions on Power Electronics, vol. 12, no. 1, pp. 46-52, Jan. 1997, DOI: 10.1109/63.554168.

[4] Stylogiannis, A., Prade, L., Buehler, A., Aguirre, J., Sergiadis, G., Ntziachristos, V. (2018). Continuous wave laser diodes enable fast optoacoustic imaging, Photoacoustics, Volume 9, 2018, Pages 31-38, ISSN 2213-5979, DOI: https://doi.org/10.1016/j.pacs.2017.12.002.

[5] Fulkerson E., Lanning R., Telford, S. (2014). Compact high current, high efficiency laser diode driver. US Patent US9972969B2. URL: http://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&d=PALL&RefSrch=yes&Query=PN/9972969.

[6] Glaser, J. (2018). Kilowatt Laser Driver with 120 A, sub-10 nanosecond pulses in < 3 cm² using an GaN FET, PCIM Asia 2018; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management, Shanghai, China, 2018, pp. 1-6. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8470507&isnumber=8470501.

[7] Glaser, J. (2018). High Power Nanosecond Pulse Laser Driver Using an GaN FET, PCIM Europe 2018; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management, Nuremberg, Germany, 2018, pp. 1-8. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8402911&isnumber=8402799.

[8] Lee, D., Moore J. (2018). Circuit and Method of Operating a Laser Diode. US Patent US20180145482A1. URL: http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20180145482.PGNR.

[9] Canal, C., Laugustin, A., Kohl,A., Rabot, O. (2017). Disruptive laser diode source for embedded LIDAR sensors, Proc. SPIE LASE, 2017, San Francisco, California, United States, 2017, pp. 10086 - 10086 - 7. DOI: https://doi.org/10.1117/12.2250197.

Using R Programming Language for Processing of Large Data Sets

Ivan Blagoev

Institute of Information and Communication Technologies – Bulgarian Academy of Sciences

Acad. Georgi Bonchev Str., block 2, 1113 Sofia, Bulgaria

i.blagoev@iit.bas.bg

Abstract: In the paper some problems and limitations in working with big volumes of data are discussed. Several methods for data load optimization in R programming language are proposed. The developed methods can be combined with other approaches to resolving such issues and creating a higher-quality programming code. Accelerating the computing capabilities of R in computer system with GPU is considered.

Keywords: R, data load, memory optimization, GPU, NVIDIA CUDA.

1 Introduction

The Programming Language R is a free system for statistical computing and graphics; it is both a programming language and a software environment [1, 2]. R is compiled and runs on a wide variety of UNIX platforms, Windows and MacOS. It was created in 1996 by Ross Ihaka and Robert Gentleman. "R", which originates from the initial letters of the small names of its creators and, as its full name (programming language and development environment for statistical and mathematical data processing) means, is a powerful tool used by thousands of scientists, teachers and students around the world, from a variety of disciplines and directions.

In recent years, learning of R language has become a standard in the Departments of Statistics and Mathematics at most of the universities in Western Europe and the United States. Among its users are different companies like Google, Astra, Merck, AT & T Labs, Baxter Healthcare Corporation and many others. R is one of the 5 main languages of Data Science in 2017 together with Python, SQL, Spark, and Tensorflow [3] (Figure 1).

The programming language R offers a wide variety of statistical techniques (linear and nonlinear modelling, classical statistical tests, time series analysis, classification, clustering, etc.) and graphic techniques and is extensively extensible [4]. The environment is logically well arranged, and one quickly finds and uses the tools available to this "lab" for statistical

research. The functions available to the user are located in libraries located in the directory "`R_HOME/library`" ("`R_HOME`" is the directory where "R" is installed). This directory contains "packages" of features that are also structured into folders. The basic set of functions in "R" is called "base" and contains the functions for reading, processing and displaying information. Language includes options for working with different data types - both numeric, string and logical (boolean). Object structures are like vectors, factors, lists, matrices, and data frames [5]. R's capabilities in this regard are currently widely used in constructing mathematical simulations and models of different biological and natural systems.

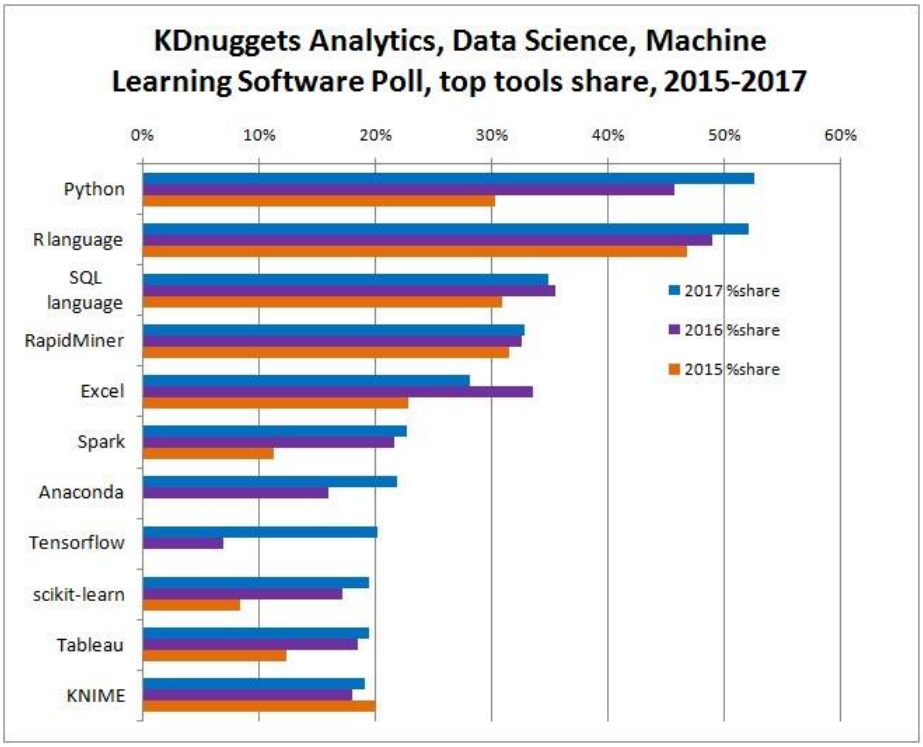

Figure 1. Main languages of Data Science Caption [3]

Exploring the programming language R and its interpretive environment allows performing a personal analysis of the data. The advantage of not being a professional programmer who knows the proper functioning of the computer system and program code implementation and yet can perform complex analysis of multiple-section data by only logically describing its task and receiving well-presented results is a great contribution to each area. But even if we are good programmers, we will again choose R when it comes to solving some types of tasks. R

tools for analysis, data calculation and result visualization would save time for many professional developers who, if using other software environments, would invest much more time and money to achieve this result.

2 Problems and Limitations in Working with Big Data

In addition to what we have pointed out as advantages here, it is good to address some of the features of the technology that can lead to constraints and problems in solving certain tasks. This is primarily the case with larger data sets (Figure 2).

Figure 2. Problems with handling large data sets

The reason of these constraints is that R calculates everything from data that is already loaded into the computer's RAM. If a large amount of data is needed, it could be invested in a more powerful computer system if computer resources are not enough. But this is not always possible and not always profitable. In this case, an effective solution is to combine R-language commands to optimize data loading in memory.

3 Methods for Reducing Data

3.1 Description of the problem

The problem is described as: a very large data file is available, and the available computer resources are not enough to load this data into memory. It is necessary to use some methods of data reduction by separating those that are not necessary for the implementation of the specific task.

3.2 Exclude rows with incorrect content.

In this case, columns without numeric values (`NA`) will accept them as such. Let the file we are loading is named "envdata_raw.csv" and has the following sample content:

```
"X","Y","Z"
1,2,3
1,NA,4
4,6,7
NA,NA,NA
```

```
4,8,NA
```

These rows can be removed when loading the file as follows:

```
>env_data=na.omit(read.csv("envdata_raw.csv", na.strings=c("",
"NA")))
> write.csv(env_data, file="envdata.csv")
> print(env_data)
        X       Y       Z
1       1       2       3
3       4       6       7
```

As a result, lines 2, 4 and 5 are not loaded in memory; the variable `env_data` contains only the rows with the valid numeric data for this case. A second line uses a command to save a processed content file called "`envdata.csv`". With the next need to work with this data, we can load the new file into R, and so we will avoid using computer resources to generate the same content.

3.3 List valid rows in a given data file without loading the entire file into a variable

```
>print(length(count.fields("envdata_raw.csv"))-1)
[1] 5
>print(length(count.fields("envdata.csv"))-1)
[1] 2
```

We remove the number "`1`" from the length result to avoid counting the column names as a data row.

3.4 Load a number of rows from the beginning of a file:

```
> env_data=read.csv("envdata_raw.csv", na.strings=c("","NA"),
nrows=2)
> print(env_data)
      "X"       "Y"       "Z"
1       1         2         3
2       1        NA         4
```

The `nrows = 2` parameter is added to the read command, which tells the command to read only the first two lines of the selected file. Thus, on the `env_data` object, we assign data only as we have determined and are needed for the specified instance.

3.5 Load any number of rows from the end of a file:

```
>env_data=tail(read.csv("envdata_raw.csv",2)
> print(env_data)
        X       Y       Z
4      NA       NA      NA
5       4        8      NA
```

With the "`tail`" command and a second parameter with a value of "`2`", it appears that only the last two lines of the contents of the specified file are loaded in the "`env_data`" object.

3.6 Deleting memory objects that we no longer use:

```
> x=tail(read.csv("envdata_raw.csv"),2)
>env_data=na.omit(read.csv("envdata_raw.csv", na.strings=c("","NA")))
> ls(all.names = TRUE)
[1] „env_data" „x"
>remove(x)
>ls(all.names = TRUE)
[1] „env_data"
>remove(env_data)
>ls(all.names = TRUE)
character(0)
```

When working with R, it is possible to load more than one object into the memory, such as a dataset, tables, matrices, etc. At one point, this would lead to the impossible loading of new data or the generation of new calculation results. For this, the `remove (.)` command immediately deletes the unnecessary memory object. We may have forgotten in the memory of the computer system objects we created, which we do not already use and are still alive, for which the `ls(all.names = TRUE)` command is added in the example. So we can see all objects created by scripts and commands and remove the unnecessary ones from them.

4 Acceleration of Processes through GPU

So far, we've looked at ways to improve performance in R through data processing methods. But we can also accelerate the computing capabilities of R if our computer system has a GPU. In the following diagram, you can see the benefit of reducing the calculation time by adding a GPU.

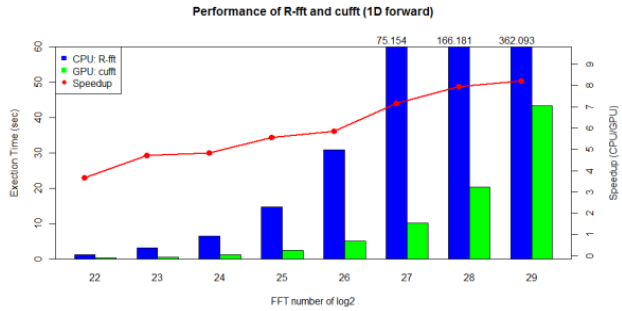

Figure 3. Speedup - CPU/GPU

For the comparative tests on the above scheme the result of 8 Intel® Xeon® CPU E5-2609 @ 2.40GHz processor and 64GB RAM has blue colour, and the result in green is the same with the NVIDIA GPU (Tesla K20Xm with 6GB device memory). On the scale in the upward direction is the time to perform the operation. The benefits of including GPU in the process are indisputable. But in order to be able to use such hardware when working with R,

we will first have to install a dedicated environment for that. Besides hardware like NVIDIA [8] from this example, it is possible to use ATI GPU and Intel XEON Phi Coprocessor.

Including each of the listed hardware in R, it is necessary to refer to the respective manufacturer's web portals for the technical documentation and to make the necessary installation. It is not possible to apply examples that deplete the possibilities of choosing hardware and operating systems on the topic. For the purpose of this material, we will take NVIDIA's hardware as one of the most common and commonly used for such purposes.

These GPUs are so widespread that they can meet on both laptops and workstations as well as in super computers. NVIDIA provides the specialized platform for massive CUDA calculations. With it, the super-parallel architecture of graphics processors remains hidden for the developer, and s/he only takes care of her/his tasks in the field of the application layer without having to think about the particularities of the GPU hardware and its integration with the specialized programming environment. In the following example, you can see how to install NVIDIA CUDA on an open source operating system Ubuntu:

1. To get started, we first provide the latest version of our GPU driver by downloading it from the NVIDIA portal and the corresponding version of CUDA;
2. We change the attributes of the downloaded files that denote it as executable:
   ```
   # chmod +x nvidia-linux-x86_64-390.87.run
   # chmod +x cuda_9.2.148*
   ```
3. Install several required libraries: `# sudo apt update`
   ```
   # sudo apt install freeglut3 freeglut3-dev libxi-dev libxmu-dev
   ```
4. Stop if the graphical environment X is running in the background, in our case:
   ```
   # sudo service lightdm stop
   ```
5. We first install the graphics driver for our GPU that we downloaded in step 1:
   ```
   # sudo ./nvidia-linux-x86_64-390.87.run
   ```
6. Installation of CUDA, in our case version 9.2:
   ```
   # sudo./cuda_9.2.148_396.37_linux.run
   ```
7. It is then necessary to define the basic parameters of the environment; using a text editor, we open the following file: `/etc/profile.d/cuda.sh` adding the following lines:
   ```
   export PATH=$PATH:/usr/local/cuda/bin
   export CUDADIR=/usr/local/cuda
   ```
   into a file: `/etc/ld.so.conf.d/cuda.conf`

   add the path to the library:

   ```
   /usr/local/cuda/lib64
   ```
8. Enter the paths to CUDA, in our custom shell:

```
# export PATH=$PATH:/usr/local/cuda-9.2/bin
# export CUDADIR=/usr/local/cuda-9.2
# export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/cuda-
9.2/lib64
```

9. After restarting the OS, the computer will be ready to use the GPU through NVIDIA CUDA.

5 Conclusions

The author's contribution is that this material helps solve problems with large data sets with the most optimal use of computer resources through the programming language R.

In conclusion, we can say that the examples presented here can not cover the whole topic of optimized data loading and the use of computing hardware when working with programming language R. Working with real data is always a challenge [7].

But the techniques presented here are among good practices and are often used; they could be combined with other approaches to solving problems in this area. Raising this issue also directs the R developer's thinking towards this problem, provoking a desire to seek approaches that lead to the creation of better-quality solutions through programming code.

References

[1] R: What is R? https://www.r-project.org/about.html, Accessed October 2018.

[2] Learn R Programming https://www.datamentor.io/r-programming/, Accessed October 2018.

[3] Top Analytics/Data Science Tools, https://www.kdnuggets.com/2017/05/poll-analytics-data-science-machine-learning-software-leaders.html, Accessed September 2018.

[4] Venables W.N., D.M. Smith and the R Development Core Team, „*An Introduction to R*“, 2016, Available at https://cran.r-project.org/doc/manuals/R-intro.html#Top

[5] 4 most used languages in big data projects: R, https://itnext.io/4-most-used-languages-in-big-data-projects-r-26c64027da69

[6] Working with Big Data in R, https://www.rstudio.com/resources/webinars/working-with-big-data-in-r/, Accessed September 2018.

[7] Nguyen Ch., Using R for Big Data Analytics, December 6, 2013, https://www.slideshare.net/adatao/adatao

[8] NVIDIA GPU, https://www.nvidia.com/object/doc_gpu_compute.html

Approach for Citizens Query Analysis in Large System of Document Processing

Bakanova N. B.

Keldysh Institute of Applied Mathematics - RAN, Miusskaya sq., 4, Moscow, 125047, Russia

nina@keldysh.ru

Abstract: The article considers the possibility of data analysis from the results of processing of one of the documentation stream in large organizations (documentary flows) which is known as "system of citizens' appeals". The information flow of the "appeal of citizens" is analyzed. Criteria for analyzing the results of processing citizens' appeals are developed. An approach that allows using data from workflow systems or other information systems to implement operational support for management activities by algorithmic methods without involving experts is proposed.

Keywords: information resources, document management organizations, "citizens' appeals", support for management activities.

1 Introduction

One of the current research areas in the field of management support is the development of applications based on the use of information accumulated in the databases of organizational systems. Experts consider the possibility of introducing some techniques used in intelligent systems into the systems of organizational management. Integration issues with ready-made expert systems applications are considered in [1, 2]. However, the conducted experiments revealed a number of features of management tasks, including: the specifics of the organization's activities, the active influence of external factors, the peculiarities of the intellectual activity of management subjects, the need to take into account the "human factor", which complicates the use of ready-made developments for organizational management systems.

Obviously, the systems of documentation support of management, as the systems for operational management of industry activities, human resources management, logistics and others information systems accumulate in their databases significant amounts of information that can be used to analyze the cooperation with higher and subordinate organizations, counterparties, to optimize internal organizational processes and other goals. This issue (as the problem

posed) is interesting for large management organizations, for government institutes, for businesses seeking to systematize and streamline their interaction with external correspondents.

This article discusses the possibility and proposes an approach for analyzing the data from the processing of one of the documentary streams received by large organizations and called “citizens' appeals” in the document management systems.

2 Research Methodology

Document circulation databases contain objective information about work with (i) appeals from citizens; (ii) decisions taken; (iii) measures to eliminate deficiencies; (iv) consideration of citizens' proposals on the work of this organization and (v) other issues. By directly processing these data, one can obtain statistical information on the number of requests in general, on the number of requests by type of correspondents (collective requests, individual citizens, public organizations), on the number of requests for the organization’s activities and other quantitative information. This information can be processed using special methods and algorithms and thus analytical studies, characterization and analysis of activities can be provided in areas of work of management organization.

In this case, the method developed by the author was used in the process of investigating the possibilities of implementing automated processes to support management activities [3-6]. The method is based on multi-criteria analysis of data obtained from the organization’s information systems.

The developed method makes it possible to calculate estimates according to selected criteria, to compare processes in various areas of activity reflected in information systems, to identify problem areas in the organization’s work. The method includes:

- evaluation criteria and grading scales;
- an algorithm for obtaining estimates for the selected criteria;
- an activity analysis algorithm based on the estimates obtained.

2.1 Characteristic of the information flow

The task set in the article is aimed at analyzing the results of processing citizens' appeals. To do this, we consider the composition of the information flow of the "appeal of citizens", presented in Figure 1.

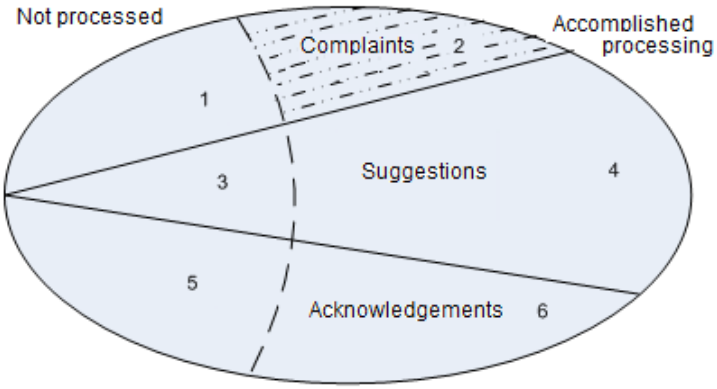

Figure 1. Information flow of the "appeals of citizens"

In the information flow of the “appeal of citizens”, there are two main groups:

- appeals of citizens who are in the processing stage (1,3,5);
- appeals from citizens whose processing has been completed (2,4,6).

If we summarize the types of documents, then the information flow can be divided into:

- complaints (1,2);
- suggestions (3,4);
- acknowledgements (5.6).

Development of criteria for analyzing the results of processing citizens' appeals was carried out for the shaded area (2).

2.2 Criteria Development

The developed criteria for analyzing the results of processing citizens' appeals are oriented to algorithmic methods and do not require special expertise. The data is retrieved during the operation of the information system and processed by special algorithms.

When developing the criteria, it was taken into account that the algorithms for selecting and evaluating data use qualitative criteria with verbal scales of assessment grades. Each criterion has an ordinal scale that allows you to set estimates based on the values of the attributes of the document card.

According to the semantic relation to the information processes, the developed criteria are combined into several groups. The grouping of criteria is carried out according to the semantic attitude to the details of the received document, links to previous requests, links to regulatory documents specified in the appeal or previous answers and other processing data.

Figure 2 and Figure 3 show the grouping of criteria according two indicators “Significance of the appeal” and “Effectiveness of result”. Tables show the grading scale for the developed criteria.

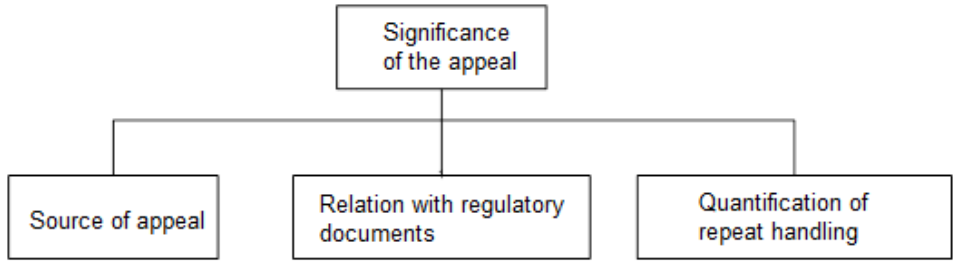

Figure 2. Grouping of criteria into the indicators “Significance of the appeal under handling”

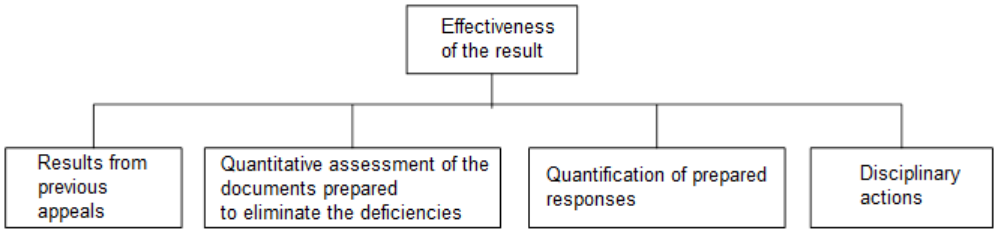

Figure 3. Grouping of criteria into the indicators “Effectiveness of Result”

Table 1. The evaluation criteria for the indicator "Significance of the appeal”

№	CRITERIA	ESTIMATES
The importance/ Significance of the document		
1	Q_{11}. Source of appeal	q_{11}^1 - government and higher organizations q_{11}^2 - subordinate organizations q_{11}^3 - public organizations q_{11}^4 - group of authors q_{11}^5 - private person individual
2	Q_{12}. Relationship with regulatory documents	q_{12}^1 - legislative references q_{12}^2 - links to government documents q_{12}^3 -references to departmental regulatory documents q_{12}^4 -references to other regulatory documents
3	Q_{13}. Directive execution dates of the document	q_{13}^1 - execute immediately q_{13}^2 - less than 15 days

		q_{13}^3 - more than 15 days q_{13}^4 - no deadlines
4	***Q14.*** Quantitative assessment of the repetition of the appeal handling	q_{14}^1 -repeat appeal > 7 q_{14}^2 -repeat appeal > 5 и ≤ 2 q_{14}^3 -primary appeal

Table 2. Evaluation criteria for the indicator "effectiveness of the result"

№	CRITERIA	ESTIMATES
Effectiveness of the result		
1	***Q21.*** Results of previous appeals	q_{21}^1 – elimination of defects q_{21}^2 – clarification given q_{21}^3 – transferred to a subordinate organization q_{21}^4 – transferred to another organization
2	***Q22.*** Qualitative assessment of the documents prepared for the elimination of deficiencies	q_{22}^1 – departmental order q_{22}^2 – indication of the subordinate organization q_{22}^3 – instruction q_{22}^4 – response to the author of the appeal
3	***Q23.*** Quantitative assessment of prepared replies	q_{23}^1 –> 7 q_{23}^2 – > 5 and ≤ 2 q_{23}^3 – 1
4	***Q24.*** Disciplinary impacts/ actions	q_{24}^1 – criminal case initiated q_{24}^2 – administrative penalty imposed q_{24}^3 – dismissal of the guilty q_{24}^4 – fine, censure q_{24}^5 – other

The number of indicators and evaluation criteria, as well as the estimates themselves, may vary to fit the specific research conditions. The scales of grades of estimates given in this article, as well as the grouping of criteria for processing citizens' appeals, are indicative and are intended to demonstrate the developed method that allows you to use data from workflow systems or other information systems to implement operational support for management activities.

3 Conclusions

The appropriate organization of data flows in large information systems that allows for constant monitoring of the work of the organizational units without additional resource costs can significantly expands the functionality of the information system. The proposed criterions'

grouping is the base for using of algorithmic processing of information without involving human experts. The multi-criteria analysis of data is used for the study of various processes in documentation support systems in large management organization.

References

[9] Ferilli S. (2011) Automatic Digital Document Processing and Management, *Advances in Pattern Recognition*, Springer-Verlag London, 196 p.

[10] Gibson, J. (2011) Organizations: Behavior, Structure, Processes, New York: McGraw Hill, 640 p.

[11] Bakanova N., Atanasova T. (2008) Modelling of integrated resources in distributed information networks. *Электросвязь*, No 3, pp. 41-44.

[12] Bakanova N., Atanasova T. (2013) Data Sources Determination for Information Management in OMIS, Proc. International Conference Robotics, Automation and Mechatronics RAM 2013, pp. 25-28.

[13] Atanasova T., Bakanova N. (2014) Information support for decision making in organization management, Proc. International Conference Robotics, Automation and Mechatronics RAM'2014, pp. 72-76.

[14] Bakanova N. (2015) Analysis of information process in a management organization for realization of decision making support, *Электросвязь*, No 5, pp. 59–62.

[15] Bakanova N., Atanasova T. (2018) Analysis of Information Resources of Organizational IS for Development of Algorithms for Support of Decision Making (in Russian), NVU "Vasil Levski", Tom 9, pp. 101-105.

ISSN: 2367 - 6450

Institute of Information and Communication Technologies - Bulgarian Academy of Sciences

"John Atanasoff" Union on Automatics and Informatics, Bulgaria