ITHEA

International Journal

INFORMATION THEORIES & APPLICATIONS

2005 Volume 12 Number 1

International Journal
INFORMATION THEORIES & APPLICATIONS

ISSN 1310-0513

Volume 12 / 2005, Number 1

IJ ITA is official publisher of the scientific papers of the members of the Association of Developers and Users of Intellectualized Systems (ADUIS).

IJ ITA welcomes scientific papers connected with any information theory or its application. Original and non-standard ideas will be published with preferences.

IJ ITA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITA as well as the **subscription fees** are given on *www.foibg.com/ijita*.
The camera-ready copy of the paper should be received by e-mail: foi@nlcv.net.
Responsibility for papers published in IJ ITA belongs to authors.

General Sponsor of IJ ITA is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal "INFORMATION THEORIES & APPLICATIONS" Vol.12, Number 1, 2005
Printed in Bulgaria
Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria,
in collaboration with the V.M.Glushkov Institute of Cybernetics of NAS, Ukraine, and
the Institute of Mathematics and Informatics, BAS, Bulgaria.
Publisher: FOI-COMMERCE - Sofia, 1000, P.O.B. 775, Bulgaria. www.foibg.com, e-mail: foi@nlcv.net

ISSN 1310-0513

PREFACE

Verba volant, scripta manent !

The "**International Journal on Information Theory and Applications**" (**IJ ITA**) has been established in 1993 as independent scientific printed and electronic media. IJ ITA is edited by the *Institute of Information Theories and Applications FOI ITHEA* in collaboration with the leading researchers from the Institute of Cybernetics "V.M.Glushkov", NASU (Ukraine) and Institute of Mathematics and Informatics, BAS (Bulgaria).

During the years, IJ ITA became as well-known international journal. Till now, including this volume, more than **500** papers have been published. IJ ITA authors are widespread in **39** countries all over the world: **Armenia, *Belarus*, *Brazil*, Belgium, *Bulgaria*, *Canada*, Czech Republic, Denmark, Egypt, Estonia, Finland, France, *Germany*, Greece, Hungary, *Ireland*, *Israel*, Italy, Japan, Kirghizia, *Latvia*, Lithuania, Malta, Mexico, Moldavia, Netherlands, *Poland*, Portugal, Romania, *Russia*, Scotland, Senegal, Serbia and Montenegro, *Spain*, Sultanate of Oman, Turkey, *UK*, *Ukraine*, and *USA***.

Volume 12/2005 of the IJ ITA contains **53** papers written by **114** authors from **14** countries (*marked in italics above*), selected from several international conferences, seminars and workshops organized or supported by the Journal.

At the first place, the main source for selection were the **ITA 2005** Joint International Events on Information Theories and Applications, (June 20-30, 2005, Varna, Bulgaria):

- XI-th International Conference "Knowledge-Dialogue-Solution" (KDS 2005);
- Third International Conference "Information Research, Applications and Education" (i.TECH 2005);
- International Conference "Business Informatics" (Bi 2005);
- IV-th International Workshop on General Information Theory (GIT 2005);
- International INTAS-FET Strategic Workshop "Data Flow Systems: Algorithms and Complexity",
- Third International Workshop on Multimedia Semantics.

Several papers were selected from the pool of papers directly submitted to IJ ITA.

Main characteristic of ITA 2005 International Conferences was that the papers were combined into thematic sessions. Because of this, the selected papers are published in this volume following the thematic sessions' organisation.

Congratulations to Luis Fernando de Mingo López (Spain) and Milena Dobreva (Bulgaria) who were awarded by the International Prize "**ITHEA**" for the year 2005. The "ITHEA" Prize has been established in 1995. It is aimed to mark the achievements in the field of the information theories and applications.

More information about the IJ ITA rules for preparing and submitting the papers as well as how to take out a subscription to the Journal may be obtained from www.foibg.com/ijita.

Krassimir Markov
IJ ITA Founder and Editor in chief

International Prize "ITHEA"

Awarded Scientists till 2005:

1995	*Sandansky*	*K. Bankov, P. Barnev, G. Gargov, V. Gladun, R. Kirkova, S. Lazarov, S. Pironkov, V. Tomov*
1996	*Sofia*	*T. Hinova, K. Ivanova, I. Mitov, D. Shishkov, N. Vashchenko*
1997	*Yalta*	*Z. Rabinovich, V. Sgurev, A. Timofeev, A. Voloshin*
1998	*Sofia*	*V. Jotsov*
1999	*Sofia*	*L. Zainutdinova*
2000	*Varna*	*I. Arefiev, A. Palagin*
2001	*St.Peterburg*	*N. Ivanova, V. Koval*
2002	*Primorsko*	*A. Milani, M. Mintchev*
2003	*Varna*	*T. Gavrilova, A. Eskenazi, V. Lozovskiy, P. Stanchev*
2004	*Varna*	*B. Kokinov, T. Vamos*
2005	*Varna*	*L.F. de Mingo López, M. Dobreva*

IJ ITA major topics of interest include, but are not limited to:

INFORMATION THEORIES

Artificial Intelligence
- *Computer Intellectualisation*
- *Intelligent Networks and Agents*
- *Intelligent Technologies*
- *Knowledge Discovery and Engineering*
- *Knowledge Acquisition and Formation*
- *Distributed Artificial Intelligence*
- *Models of Plausible Reasoning*
- *AI Planning and Scheduling*

Bioinformatics
Business Informatics
Cognitive Science
Decision Making
Education Informatics
General Information Theory
Hyper Technologies
Information Models
Intellectualisation of Data Processing
Knowledge-based Society
Logical Inference
Natural language Processing
Neuroinformatics
Philosophy and Methodology of Informatics
Quality of the Programs
Software Engineering
Theory of Computation

APPLICATIONS

Business Information Systems
Communication Systems
Computer Art and Computer Music
Hyper Technologies
Intelligent Information Systems
Multimedia Systems
Programming Technologies
Program Systems with Artificial Intelligence
Pyramidal Information Systems
Very Large Information Spaces

FILTERED NETWORKS OF EVOLUTIONARY PROCESSORS*

Luis Fernando de Mingo López, Eugenio Santos Menéndez, Francisco Gisbert

Abstract: This paper presents some connectionist models that are widely used to solve NP-problems. Most well known numeric models are Neural Networks that are able to approximate any function or classify any pattern set provided numeric information is injected into the net. Neural Nets usually have a supervised or unsupervised learning stage in order to perform desired response. Concerning symbolic information new research area has been developed, inspired by George Paun, called Membrane Systems. A step forward, in a similar Neural Network architecture, was done to obtain Networks of Evolutionary Processors (NEP). A NEP is a set of processors connected by a graph, each processor only deals with symbolic information using rules. In short, objects in processors can evolve and pass through processors until a stable configuration is reach. This paper just shows some ideas about these two models.

Keywords: Natural Computation, Membrane Systems, Neural Networks, Networks of Evolutionary Processors.

ACM Classification Keywords: F.1.1 Models of Computation: Self-modifying machines (neural networks); F.4.1 Mathematical Logic: Computational logic

Introduction

Natural sciences, and especially biology, represented a rich source of modeling paradigms. Well-defined areas of artificial intelligence (genetic algorithms, neural networks), mathematics, and theoretical computer science (L systems, DNA computing) are massively influenced by the behavior of various biological entities and phenomena. In the last decades or so, new emerging fields of so-called "natural computing" identify new (unconventional) computational paradigms in different forms. There are attempts to define and investigate new mathematical or theoretical models inspired by nature, as well as investigations into defining programming paradigms that implement computational approaches suggested by biochemical phenomena. Especially since Adleman's experiment, these investigations received a new perspective. One hopes that global system-level behavior may be translated into interactions of a myriad of components with simple behavior and limited computing and communication capabilities that are able to express and solve, via various optimizations, complex problems otherwise hard to approach.

In the last decade and especially after Adleman's experiment [1] a number of computational paradigms, inspired or gleaned from biochemical phenomena, are becoming of growing interest building a wealth of models, called generically Molecular Computing. New advances in, on the one hand, molecular and theoretical biology, and on the other hand, mathematical and computational sciences promise to make it possible in the near future to have accurate systemic models of complex biological phenomena. Recent advances in cellular Biology led to new models, hierarchically organized, defining a new emergent research area called Cellular Computing.

Numeric Models - Neural Networks

Neural networks are non-linear systems whose structure is based on principles observed in biological neuronal systems. A neural network could be seen as a system that can be able to answer a query or give an output as answer to a specific input. The in/out combination, i.e. the transfer function of the network is not programmed,

* Supported by INTAS 00-626 and TIC 2003-09319-c03-03.

but obtained through a ”training” process on empiric datasets. In practice the network learns the function that links input together with output by processing correct input/output couples. Actually, for each given input, within the learning process, the network gives a certain output that is not exactly the desired output, so the training algorithm modifies some parameters of the network in the desired direction. Hence, every time an example is input, the algorithm adjusts its network parameters to the optimal values for the given solution: in this way the algorithm tries to reach the best solution for all the examples. These parameters we are speaking about are essentially the weights or linking factors between each neuron that forms our network. Neural Networks' application fields are typically those where classic algorithms fail because of their inflexibility (they need precise input datasets). Usually problems with imprecise input datasets are those whose number of possible input datasets is so big that they can't be classified. For example in image recognition are used probabilistic algorithms whose efficiency is lower than neural networks' and whose characteristics are low flexibility and high development complexity. Another field where classic algorithms are in troubles is the analysis of those phenomena whose mathematical rules are unknown. There are indeed rather complex algorithms which can analyse these phenomena but, from comparisons on the results, it comes out that neural networks result far more efficient: these algorithms use Fourier's transform to decompose phenomena in frequential components and for this reason they result highly complex and they can only extract a limited number of harmonics generating a big number of approximations. A neural network trained with complex phenomena's data is able to estimate also frequential components, this means that it realizes in its inside a Fourier's transform even if it was not trained for that! One of the most important neural networks' applications is undoubtfully the estimation of complex phenomena such as meteorological, financial, socio-economical or urban events. Thanks to a neural network it's possible to predict, analyzing historical series of datasets just as with these systems but there is no need to restrict the problem or use Fourier's transform. A defect common to all those methods it's to restrict the problem setting certain hypothesis that can turn out to be wrong. We just have to train the neural network with hystorical series of data given by the phenomenon we are studying. Calibrating a neural network means to determinate the parameters of the connections (synapsis) through the training process. Once calibrated there is need to test the netowrk efficiency with known datasets, which has not been used in the learning process. There is a great number of Neural Networks which are substantially distingushed by: type of use, learning model (supervised/non-supervised), learning algorithm, architecture, etc.

Multilayer perceptrons (MLPs) are layered feed forward networks typically trained with static backpropagation. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are easy to use, and that they can approximate any input-output map. In principle, backpropagation provides a way to train networks with any number of hidden units arranged in any number of layers. In fact, the network does not have to be organized in layers any pattern of connectivity that permits a partial ordering of the nodes from input to output is allowed. In other words, there must be a way to order the units such that all connections go from ”earlier” (closer to the input) to ”later” ones (closer to the output). This is equivalent to stating that their connection pattern must not contain any cycles. Networks that respect this constraint are called feed forward networks; their connection pattern forms a directed acyclic graph or dag.

Jordan and Elman networks extend the multilayer perceptron with context units, which are processing elements that remember past activity. Context units provide the network with the ability to extract temporal information from the data. In Elman networks, the activity of the first hidden layer are copied to the context units, while the Jordan network copies the output of the network. Jordan and Elman networks combine the past values of the context unit with the present input x to obtain the present net output. The Jordan context unit acts as a so called lowpass filter, which creates an output that is the weighted (average) value of some of its most recent past outputs.

Time lagged recurrent networks are MLPs extended with short term memory structures. Most real-world data contains information in its time structure. Yet, most neural networks are purely static classifiers. TLRNs are the state of the art in nonlinear time series prediction, system identification and temporal pattern classification.

Symbolic Models - Cellular Computing

P-systems represent a class of distributed and parallel computing devices of a biological type that was introduced in [14] which are included in the wider field of cellular computing. Several variants of this model have been investigated and the literature on the subject is now rapidly growing. The main results in this area show that P-systems are a very powerful and efficient computational model [15], [16], [13]. There are variants that might be classified according to different criteria. They may be regarded as language generators or acceptors, working with strings or multisets, developing synchronous or asynchronous computation. Two main classes of P-systems can be identified in the area of membrane computing [15]: cell-like P-systems and tissue-like P-systems. The former type is inspired by the internal organization of living cells with different compartments and membranes hierarchically arranged; formally this structure is associated with a tree. Tissue P-systems have been motivated by the structure and behaviour of multicellular organisms where they form a multitude of different tissues performing various functions [2]; the structure of the system is instead represented as a graph where nodes are associated with the cells which are allowed to communicate alongside the edges of the graph.

More recently, a notion of population P-systems has been introduced [3], [4] as a model for tissue P-systems where the structure of the underlying graph can be modified during a computation by varying the set of nodes and the set of edges in the graph. Specifically, nodes are associated with cells, each of them representing a basic functional unit of the system, and edges model bonds among these cells that are dynamically created and destroyed. Although mainly inspired by the cell behavior in living tissues, population P-systems may be also regarded as an abstraction of a population of bio-entities aggregated together in a more complex bio-unit (e.g. social insects like ants, bees, wasps etc, organized in colonies or bacteria of different types). This is the main reason why we use the term population instead of tissue albeit the term cell is retained to denoting an individual in the system. The concept also recalls other similar computational models: grammar systems [8], eco-grammar systems [9], or more recently, networks of parallel/evolutionary processors [10].

Universality results have been obtained [4] for a number of variants of population P-systems. The following different rules are considered: transformation rules for modifying the objects that are present inside the cells, communication rules for moving objects from a cell to another one, cell division rules for introducing new cells in the system, cell differentiation rules for changing the types of the cells, and cell death rules for removing cells from the system. As well as this, bond-making rules are considered that are used to modify the links between the existing cells (i.e., the set of edges in the graph) at the end of each step of evolution performed by means of the aforementioned rules. In other words, a population P-system in [4] is basically defined as an evolution-communication P-system [7] but with the important difference that the structure of the system is not rigid and it is represented as an arbitrary graph. In particular, bond making rules are able to influence cell capability of moving objects from a place to another one by varying the set of edges in the underlying graph.

Another interesting variant of population P-systems is obtained by considering the general mechanism of cell communication based on signal molecules as a mechanism for triggering particular transformations inside of a cell once a particular signal-object has been received from some other cell in the system [3]. This leads to a notion of population P-systems where the sets of rules associated with the cell can vary according to the presence of particular objects inside and outside the cells. Yet again, the introduction of this mechanism is motivated by the features shared by biological systems at various levels where the behavior of an individual is affected both by its internal state and by the external stimuli received. Some results concerning the power of population P-systems with a rule activating mechanism have been obtained [5].

Further developments of the area of population P-systems are expected to cover alternative ways of defining the result of a computation and the use of string objects. Population P-systems in fact attempt to model aspects of biological systems formed by many different individual components cooperating in a coherent way for the benefit of the system as a whole; a more appropriate notion of computation is therefore necessary in order to characterize the emergent behavior of the system. Existing approaches in the area of grammar system such parallel communicating grammar systems [8] or eco-grammar systems [9], rely on the use of a single sentential form that is rewritten in parallel by different interacting/cooperating grammar components. In particular, in the case of eco-grammar systems, this sentential form is associated with the environment and it can be rewritten both by rules corresponding to action taken from the individual components in the system and by dedicated rules associated with the environment. In a similar way, we can consider string-processing population P-systems where the result of a computation is given by a string (or a language) produced in the environment at the end of a computation. However, with respect to grammar systems, population P-systems present some other interesting features like the possibility of moving objects from a place to another one, the possibility of forming bonds among the cells, the possibility of introducing new cells in the system by means of cell division, which need to be formalized for the particular case of string objects. In this respect, we aim to present some reasonable variants of population P-systems with string objects.

Membranes in P-systems can be connected using a graph and all of them can be treated as skin ones forming a so called Network of Evolutionary Processors. A network of evolutionary processors of size n is a construct NEP = $(V, N_1, N_2, \ldots, N_n, G)$, where V is an alphabet and for each $1 \le i \le n$, $N_i = (M_i, A_i, PI_i, PO_i)$ is the i-th evolutionary node processor of the network. The parameters of every processor are:

- M_i is a finite set of evolution rules of one of the following forms only:

  $a \prod b$, $a, b \in V$ (substitution rules)

  $a \prod \varepsilon$, $a \in V$ (deletion rules)

  $\varepsilon \prod a$, $a \in V$ (insertion rules)

  More clearly, the set of evolution rules of any processor contains either substitution or deletion or insertion rules.
- A_i is a finite set of strings over V. The set A_i is the set of initial strings in the i-th node. Actually, in what follows, we consider that each string appearing in any node at any step has an arbitrarily large number of copies in that node, so that we shall identify multisets by their supports.
- PI_i and PO_i are subsets of V^*representing the input and the output filter, respectively. These filters are defined by the membership condition, namely a string $w \in V^*$ can pass the input filter (the output filter) if $w \in PI_i$ ($w \in PO_i$).

$G = (N_1, N_2, \ldots, N_n, E)$ is an undirected graph called the underlying graph of the network. The edges of G, that is the elements of E, are given in the form of sets of two nodes. The complete graph with n vertices is denoted by K_n. By a configuration (state) of an NEP as above we mean an n-tuple $C = (L_1, L_2, \ldots, L_n)$$, with $L_i \subseteq V^*$ for all $1 \le i \le n$. A configuration represents the sets of strings (remember that each string appears in an arbitrarily large number of copies) which are present in any node at a given moment; clearly the initial configuration of the network is $C_0 = (A_1, A_2, \ldots, A_n)$.

A configuration can change either by an evolutionary step or by a communicating step. When changing by an evolutionary step, each component L_i of the configuration is changed in accordance with the evolutionary rules associated with the node i. When changing by a communication step, each node processor N_i sends all copies of the strings it has which are able to pass its output filter to all the node processors connected to N_i and receives all copies of the strings sent by any node processor connected with N_i providing that they can pass its input filter.

Theorem 1. Each recursively enumerable language can be generated by a complete NEP of size 5. [21]

Theorem 2. Each recursively enumerable language can be generated by a star NEP of size 5. [22]

Theorem 3. The bounded PCP can be solved by an NEP in size and time linearly bounded by the product of K and the length of the longest string of the two Post lists. [23]

A simple NEP of size n is a construct SNEP = (V, N_1, N_2,… , N_n, G), where, V and G have the same interpretation as for NEPs, and for each $1 \leq i \leq n$, N_i = (M_i, A_i, PI_i, FI_i, PO_i, FO_i) is the i-th evolutionary node processor of the network. M_i and A_i from above have the same interpretation as for an evolutionary node in a NEP, but:

- PI_i and FI_i are subsets of V representing the input filter. This filter, as well as the output filter, is defined by random context conditions, PI_i forms the permitting context condition and FI_i forms the forbidding context condition. A string $w \in V^*$ can pass the input filter of the node processor i, if w contains each element of PI_i but no element of FI_i. Note that any of the random context conditions may be empty, in this case the corresponding context check is omitted. We write ρ_i (w) = true, if w can pass the input filter of the node processor i and ρ_i (w) = false, otherwise.
- PO_i and FO_i are subsets of V representing the output filter. Analogously, a string can pass the output filter of a node processor if it satisfies the random context conditions associated with that node. Similarly, we write τ_i (w) = true, if w can pass the input filter of the node processor i and τ_i (w) = false, otherwise.

Theorem 4. The families of regular and context-free languages are incomparable with the family of languages generated by simple NEPs. [21]

Theorem 5. The "3-colorability problem" can be solved in O(m + n) time by a complete simple NEP of size 7m+2, where n is the number of vertices and m is the number of edges of the input graph. [21]

Filtered Connections in NEPs

Main idea in NEPs is based on the fact that filters are inside processors in order to control what objects can pass through connections, but these filters make complex proccessors. If such filters are in connections, instead in processors, the simplicity of processors will increase compare to classical NEPs.

A network of evolutionary processors with filtered connections of size n is a construct FNEP = (V, N_1, N_2,… , N_n, G) where V is an alphabet and for each $1 < i < n$, N_i=(M_i, A_i) is the i-th evolutionary node processor of the network. The parameters of every processor are:

- M_i is a finite set of evolution rules (substitution, deletion or insertion rules).
- A_i is a finite set of strings over V. The set A is the set of initial strings in the i-th node. Actually, in what follows, we consider that each string appearing in any node at any step has an arbitrarily large number of copies in that node, so that we shall identify multisets by their supports.

Finally, G=(N_1, N_2, … ,N_n, (E,F)) is an directed graph called the underlying graph of the network. The edges of G, that is the elements of (E,F), are given in the form (e_i, f_i) where f_i is the filter associated to connection e_i. Elements in F are just object sets, an element w pass the filter in f_i if $w \in f_i$. The complete graph with n vertices is denoted by K_n. By a configuration (state) of an NEP as above we mean an n-tuple C=(L_1, L_2, … ,L_n), with $L_i \subseteq V^*$ for all $1 < i < n$. A configuration represents the sets of strings (remember that each string appears in an arbitrarily large number of copies) which are present in any node at a given moment; clearly the initial configuration of the network is C_0 = (A_1, A_2, … ,A_n).

A configuration can change either by an evolutionary step or by a communicating step. When changing by an evolutionary step, each component L_i of the configuration is changed in accordance with the evolutionary rules associated with the node i. When changing by a communication step, each node processor N_i sends all copies of the strings it has to all the node processors connected to N_i and receives all copies of the strings sent by any node processor connected with N_i (providing that all sent/received information pass filters in connections).

Theorem 6. Solved problems using NEPs can be solve using NEPs with filtered connections.

Given two processor of a NEP, N_i and N_j connected by edge e_{ij} and with filters (FI_i,FO_i) and (FI_j,FO_j), they can be transformed into two processors of a NEP with filtered connections in the following way:

- Remove filters from processors N_i and N_j to obtain processors of NEPFC M_i and M_j.
- Add the filter $f_{ij}=FO_i \wedge FI_j$ to a connection from processor M_i -> M_j.
- Add the filter $f_{ji}=FO_j \wedge FI_i$ to a connection from processor M_i <- M_j.

It is clear that this kind of transformation produces a NEPFC with the same behaviour than a NEP.

Theorem 7. A NEPFC with m processors and less than c=2m connections can not be transformed into an equivalent NEP.

Each c connection is an equation with two unknows, so there are 2c unknows (input and output filters in NEPs) and there exists 2m filters to compute. So if the c<2m the system has infinite solutions but the behaviour will not be the same in all cases. If c=2m the system has only one solution, and if c>2m the system has only one solution.

Therefore, if NEPs and NEPFCs are equivalent under some constraints then all theorems in NEPs are valid for NEPFC. This new model can solve NP-problems in linear time.

Conclusions

This paper has introduced the novel computational paradigm Networks of Evolutionay Processors. Connectionists models such as Neural Networks can be taken into account to develop NEP architecture in order to improve behaviour. As a future research, learning concepts in neural networks can be adapted in a NEP architecture provided the numeric-symbolic difference in both models. NEPs and NEPFCs can be consired universal models since the are able to solve NP-problems. The great disadvantage is that a given NEP/NEPFC can only solve a given problem, if it is necessary to solve another problem (maybe a little variation) then another different NEP/NEPFC has to be implemented. The idea of learning tries to undertake such disadvantage proposing a model able to solve different kinds of problems (that is a general class of problems). Learning proposed can be based on the self organizing maps. There are a lof of open problems that need to be solved in order to show the computational power of this model, but the possibility to compute NP-problems is promising apart from the massive parallelization and non-determinism of the model.

Bibliography

[1] Adleman, L.M. 1994. Molecular computation of solutions to combinatorial problems. Science, 226, 1021-1024

[2] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. 2002. The Molecular Biology of the Cell. Fourth Edition. Garland Publ. Inc., London

[3] Bernardini, F., Gheorghe, M. 2004. Cell Communication in Tissue Psystems and Cell Division in Population P-Systems. In [17], 74-91

[4] Bernardini, F., Gheorghe, M. 2004. Population P-Systems. Journal of Universal Computer Science, 10, 509-539

[5] Bernardini, F., Gheorghe, M. 2005. Cell Communication in Tissue P-Systems: Universality Results. Soft Computing (to appear)

[6] Busby, S., de Lorenzo, V. 2001. Cell regulation - putting together pieces of the big puzzle. Curr. Op. Microbiol, 4, 117-118.

[7] Cavaliere, M. 2003. Evolution Communication P-Systems. In [16], 134-145 and in [17], 206-223.

[8] Csuhaj-Varju, E., Dassow, J., Kelemen, J., Paun, Gh. 1997. Grammar Systems. A Grammatical Approach to Distribution and Cooperation. Gordon and Breach, London

[9] Csuhaj-Varju, E., Kelemen, J., Kelemenova, A., Paun, Gh. 1997. EcoGrammar Systems: A Grammatical Framework for Studying Life-Like Interactions. Artificial Life, 3, 1-28.

[10] Csuhaj-Varju, E., Salomaa, A., 1997. Networks of Parallel Language Processors. In New Trends in Formal Languages. Control, Cooperation, and Combinatorics. In Paun, Gh., Salomaa, A. (eds), Lecture Notes in Computer Science, 1218, Springer-Verlag, Berlin, Heidelberg, New York, 299-318

[11] Krasnogor, N., Gheorghe, M., Terrazas, G., Diggle, S., Williams, P., Camara, M. 2005. Bulletin of the EATCS (to appear)

[12] Jennings, N.R. 2000. On agent-based software engineering. Artificial Intelligence, 117, 277-296.

[13] Martin-Vide, C., Mauri, G., Paun, Gh., Rozenberg, G., Salomaa, A. (eds) 2004. Membrane computing. International workshop, WMC 2003, Tarragona, Spain, July 2003. Revised papers. Lecture Notes in Computer Science, 2933, Springer, Berlin Heidelberg New York

[14] Paun, Gh. 2000. Computing with Membranes. Journal of Computer and System Sciences, 61, 108-143

[15] Paun, Gh. 2002. Membrane computing. An introduction. Springer, Berlin Heidelberg New York

[16] Paun, Gh., Rozenberg, G., Salomaa, A., Zandron, C. (eds) 2003. Membrane computing. International workshop, WMC-CdeA 02, Curtea de Arges, Romania, August 19-23, 2002. Revised papers. Lecture Notes in Computer Science, 2597, Springer, Berlin Heidelberg New York

[17] Paun, Gh., Riscos-Nunez, A., Romero-Jimenez, A., Sancho-Caparrini, F. (eds) 2004. Second brainstorming week on membrane computing, Seville, 2-7 February 2004. Technical Report 01/2004, Department of Computer Science and Artificial Intelligence, University of Seville, Spain

[18] Swift, S., Downie, J.A., Whitehead, N.A., Barnard, A.M.L., Salmond, G.P.C., Williams, P. 2001. Quorum sensing as a population-densitydependent determinant of bacterial physiology. Adv Micro Physiol, 45, 199270.

[19] Williams, P., Camara, M., Hardman, A., Swift, S., Milton, D., Hope, V.J., Winzer, K., Middleton, B., Pritchard, D.I., Bycroft, B.W. 2000. Quorum sensing and the population dependent control of virulence. Phil. Trans Roy Soc London B, 355(1397), 667-680.

[20] Winzer, K., Hardie, K.H., Williams, P. 2002. Bacterial cell-to-cell communication: sorry can't talk now – gone to lunch!. Curr Op. Microbiol, 5, 216-222.

[21] Juan Castellanos, Carlos Martin-Vide, Victor Mitrana, Jose M. Sempere; *Networks of Evolutionary Processors*.

[22] Juan Castellanos, Carlos Martin-Vide, Victor Mitrana, Jose M. Sempere; *Solving NP-Complete Problems with Networks of Evolutionary Processors*. IWANN.

[23] Carlos Martin-Vide, Victor Mitrana, Mario J. Perez-Jimenez, Fernando Sancho-Caparrini; *Hybrid Networks of Evolutionary Processors*. GECCO 2003. Lecturne Notes on Computer Science 2723, pp. 401-412. 2003.

Authors' Information

Luis Fernando de Mingo López – Dept. Organización y Estructura de la Información, Escuela Universitaria de Informática, Universidad Politécnica de Madrid, Crta. De Valencia km. 7, 28031 Madrid, Spain; e-mail: lfmingo@eui.upm.es

Eugenio Santos Menéndez – Dept. Organización y Estructura de la Información, Escuela Universitaria de Informática, Universidad Politécnica de Madrid, Crta. De Valencia km. 7, 28031 Madrid, Spain; e-mail: esantos@eui.upm.es

Francisco Gisbert – Dept. Organización y Estructura de la Información, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain; e-mail: fgisbert@fi.upm.es

DIGITISATION OF CULTURAL HERITAGE: BETWEEN EU PRIORITIES AND BULGARIAN REALITIES[1]

Milena Dobreva, Nikola Ikonomov

Abstract: The paper presents the Bulgarian setting in digital preservation of and access to cultural and scientific heritage. It mentions key Bulgarian institutions, which take or should take part in digitisation endeavours. It also presents examples of building and adapting specialised tools in the field, and more specifically SPWC, ACT and XEditMan.

Keywords: digital preservation of and access to cultural and scientific heritage, legislative issues, SPWC, ACT, XEditMan

ACM Classification Keywords: I.3.3. Digitizing and scanning; I.3.6. Standards; D.2.1. Methodologies

Introduction

Before the current period of economic transition, until 1990, Bulgaria was the Eastern-European country with highest expertise in information technologies within the CMEA (Council for Mutual Economic Assistance). During this time, digitisation of cultural and scientific heritage was not a separate field of work – but domains such as library automation systems, databases of cultural heritage objects; corpora of ancient and mediaeval texts had been already developed for several decades. Such work was also done in Bulgaria but not on a large-scale systematic basis.

In the years after 1990, Bulgaria has been undergoing a period of structural changes and economic transition. The acquired technological excellence in information and communication technologies (ICT) had been transferred from huge institutions to small and medium-sized enterprises functioning in a highly competitive environment. Digitisation activities, which require large investments and are not bringing quick profit, have not become attractive for the companies from the ICT sector.

Additionally, culture, education and science sectors have been suffering from inadequate funding during the transition period (the share of gross national product spent for science for example in the last years is about 0,29% which is 10 times less than in the EC). This general setting was not favourable for the establishing of national and institutional digitisation programmes.

At the same time, Bulgarian collections house over 12,500 manuscripts of Slavonic, Greek, Latin, Ottoman Turkish and other origin. Another key example is the epigraphic inscriptions from the Antiquity period, which form the third largest collection in the world following those of Italy and Greece. Precious monuments of immovable heritage, nine objects in the UNESCO World Heritage List, numerous archæological findings, Old Bulgarian runic inscriptions — all these materials are of interest not only for the local community, but also for the wider European community of which Bulgaria is a cultural part and indeed for all of mankind globally by virtue of the shared meaning of the culture of the Other. Yet, electronic information on these resources is still hardly accessible in its fullness not only to foreign experts, but also to regional and local specialists.

In this paper we will present the key types of institutions playing different roles in digitization processes with a brief comment on their actual involvement in digitisation.

---

[1] The part of this article presenting institutional involvement is an extended and updated version of the respective part of a report on Bulgaria prepared by Milena Dobreva for a MINERVAPlus Global Report (**Coordinating digitisation in Europe**. Progress report of the National Representatives Group: coordination mechanisms for digitisation policies and programmes 2004, forthcoming).

Bulgarian Institutions and Their Level of Involvement in Digitisation Activities

Six types of organisations are potentially interested in digitisation of cultural heritage: government bodies, repositories, research and/or educational institutions, companies, and foundations.

Government bodies are entrusted with the supervision of such activities.

Three institutions should play the key role for establishing digitisation policy in Bulgaria but neither one is currently working in this direction:

1. *The Ministry of Culture and Tourism.*[1] The last structural change was done very recently, in the end of January 2005 when Tourism was added to the activities of the ministry in recognition of the fact that cultural tourism will be one of the basic specialisation sectors for the Bulgarian economy in the next years.
2. *ICT Development Agency at the Ministry of Transport and Communications*[2], Republic of Bulgaria. The agency was created in 1995 and currently is the body responsible for the development of the information and communication technologies in Bulgaria. In the last few years it provided funding for projects aimed at presentation of cultural heritage in electronic form. One example is the first XML repository of catalogue descriptions of Old Bulgarian manuscripts preserved in Bulgaria – a project carried out by the Institute of Mathematics and Informatics and funded by the Agency in 2004. However, a coherent strategy has not been created and respectively followed.
3. *Ministry of Education and Science.*[3] Digitisation is not a technical activity; it develops rapidly and involves new research results.

Repositories (libraries, archives and museums), which seem the most natural initiators of digitisation projects because of the close relationship between digitisation and preservation, are currently in the position of observers due to lack of funding on the one hand, and copyright issues for digital collections, on the other hand. There are about 7000 public, university, scientific, specialised libraries and information centres in the country. As most important institutions in this group we should mention:

1. *General Department of Archives at the Council of Ministers of Republic of Bulgaria*.[4] The General Department of Archives initiated pilot work in digitisation of archival documents with the publication of the documentary CD compendium “The Independence of Bulgaria and the Bulgarian Army” containing materials from the Central Military Archive in Veliko Turnovo in 2003. The vision on digitisation activities of the State Archives was presented recently [Markov 2004].
2. *The National Library “Saint Cyril and Saint Methodius”*[5] plays a leading role in the process of expert decision-making related to measures of digital cataloguing and publishing of mediæval manuscript heritage and early printed books. Its prescriptions in these fields are adopted in other libraries in the country, which have such collections. The National Library is also the basic driving force for digital cataloguing of modern books. Although the library experts have quite extensive experience in following the current practices, real digitisation work has not been planned [Moussakova, Dipchikova 2004].
3. *The National Museum of History*[6] does not seem to be currently involved in any digitisation-related work.

---

[1] http://www.culture.government.bg/ date of last visit 25.5.2005.

[2] http://www.ict.bg/, date of last visit 25.5.2005.

[3] http://www.minedu.government.bg, date of last visit 26.5.2005.

[4] http://www.archives.government.bg/index_en.html, date of last visit 25.5.2005.

[5] http://www.nationallibrary.bg/, page does not open on 25.5.2005.

[6] http://www.historymuseum.org/, date of last visit 25.5.2005.

Research and/or educational institutions are the most active initiators of small-scale digitisation projects in Bulgaria. They usually do not have the funds and resources for running mass digitisation projects, but are the most active promoters of this field of work.

1. *The Institute of Mathematics and Informatics*[1] of the Bulgarian Academy of Sciences (IMI) plays the leading role in this direction. Digitisation of Scientific Heritage department[2] was established in IMI in 2004. The institute took part in projects related to digitisation of mathematical heritage; cataloguing and electronic publishing of mediæval Slavonic manuscripts. In addition, IMI organised in the last years three summer schools and four specialised workshops related to digitisation of cultural and scientific heritage which were targeted at Central European countries' participants and have regional impact.
   The Institute produced the most extensive XML catalogue (over 800 catalogue records) of Old Bulgarian manuscripts stored in Bulgaria [Pavlov 2004] in cooperation with specialists from the Faculty of Mathematics and Informatics of the Sofia University "Kliment Ohrdiski" and the National Library "St Cyril and St Methodius" IMI is the coordinator of the international project Knowledge Transfer for the Digitisation of Cultural and Scientific Heritage in Bulgaria (KT-DigiCult-BG), supported by the Marie Curie programme, Framework Programme 6 of the European Commission which is implemented in 2004-2008.
   The Institute is a partner in the COMTOOCI project supported by eCulture program of the EC, which is coordinated by the Institute for computational linguistics in Pisa, Italy. In this project its role is to support the localisation and local implementation of a specialised software for philological and librarian work in the cultural institutions which was developed by the Italian institute.
   IMI also works on presentation of folklore archives in digital form in cooperation with the Institute for folklore of the Bulgarian Academy of Sciences.
2. *The Institute for Bulgarian Language*[3] (IBL) works on digital preservation and use of audio archives containing live recordings presenting various Bulgarian dialects. These records originally were collected in the 50s and 60s in the 20c, and their conversion in electronic form was absolutely necessary since the original tapes started to deteriorate.
3. Amongst educational institutions we should mention The State Library Institute,[4] which recently opened specialized programme Information funds *of the cultural and scientific heritage*. Sofia University offers a general programme on Library and information activities[5].

Companies are interested in presenting sections of cultural heritage to the world, which they believe will be easily realised on the market. Today it is rather difficult to establish customer interest. The Bulgarian market for such products is unsatisfactory. This is why their main market is abroad. As an example of a company, which specializes in digitisation services, we could mention BalkanData[6] - a US-owned company based in Bulgaria. This combination seeks to offer the winning combination of the local technological and intellectual excellence and the low labour costs in the country.

Non-governmental institutions (NGOs). One active organisation in the library field is The Union of Librarians and Information Services Officers (ULISO).[7] It produced in 1997 the National Program for the preservation of Library Collections.

---

[1] www.math.bas.bg, date of last visit 25.5.2005.

[2] http://www.math.bas.bg/digi/indexbg.html, date of last visit 25.5.2005.

[3] http://www.ibl.bas.bg, date of last visit 25.5.2005.

[4] http://www.svubit.org/, date of last visit 25.5.2005.

[5] http://forum.uni-sofia.bg/filo/display.php?page=bibliotekoznanie, date of last visit 25.5.2005.

[6] http://www.balkandata.net/, date of last visit 25.5.2005.

[7] http://www.lib.bg/act.htm, date of last visit 25.5.2005.

Funding bodies (foundations) rarely support projects undertaken in the field of digitisation. In addition, the scale of their support cannot meet the real costs of serious digitisation projects. In the last years the tendency is that such bodies are supporting basically dissemination activities (workshops, conferences, trainings).

Legislative Issues

The main cultural and scientific heritage collections in Bulgaria belong to the State and their maintenance is totally dependent on the State budget. One would expect that the development of a national policy for digitisation would be an easy task when most collections of the cultural heritage are State-owned. Unfortunately, most of the legislation in the cultural sphere does not cover any digitisation aspects. A brief presentation of key legal acts covering issues, which could be approached also in digitisation programmes follow.

The Law for Protection and Development of the Culture[1] (in force since 1 January 2001) defines the basic principles and functions of the national cultural policy and the cultural institutions. However, digitisation is not mentioned amongst the issues that are covered in it.

The Deposit Law[2] (last version in force as of 1 January 2001) addresses works on digital media (electronic documents). According to it, works published on digital media should be presented in three copies to the National Library within two weeks after the publication. The National Library stores these materials as physical copies, and is not seen as a body, which would include the electronic publications into a digital library.

The Regulation for Rendering and Saving Movable Cultural Monuments[3] addresses the matters of finding, collecting, and preserving of movable cultural heritage monuments and making scientific descriptions related to them. Its application is mandatory for all museums, art galleries, museum collections as well as individuals. According to Article 62, the basic form of record and scientific description is the inventory book. The detail and accuracy of records is the responsibility of the directors of the collections. The scientific descriptions of immovable objects are presented as “Scientific passports” of the objects (Article 79). This regulation is in force since 1 January 1974. Understandably, electronic records and links between documentation of various collections were not planned in that time, but changes, which would take into account the current state of technology, have not been made.

The Regulation N 26 of 10.04.1996 of the Development, Use and Management of an Automated Information System “An Archæological map of Bulgaria”[4] seems to be the only legislative act in Bulgaria which treats a matter of digital presentation and storage of data related to the cultural heritage. It addresses the development of a specialized information system. The feeding of the database is the responsibility of the Institute of Archæology of the Bulgarian Academy of Sciences and the National Institute for Cultural Monuments based on primary data supplied from specialists who worked *in situ*. Information can be obtained from this automated system only on the basis of a written request for a service fee. The collection of data and their use were adequate for the state of the technologies in 1996; now this is outdated but changes to adapt the collected data and to provide access via the Internet have not been done.

The Tariff of rates collected by State Cultural Institutions for Services and Provision of Documents and Copies[5], date of last update 5 January 2001 does not include any fees related to digital images despite the actuality of the update.

---

[1] http://www.culture.government.bg/docdetail.html?id=16, in Bulgarian, date of last visit 25.5.2005.

[2] http://www.culture.government.bg/docdetail.html?id=66, in Bulgarian, date of last visit 25.5.2005.

[3] http://www.culture.government.bg/docdetail.html?id=49, in Bulgarian, date of last visit 25.5.2005.

[4] http://www.culture.government.bg/docdetail.html?id=48, in Bulgarian, date of last visit 25.5.2005.

[5] http://www.culture.government.bg/docdetail.html?id=38, in Bulgarian, date of last visit 25.5.2005.

EU Cooperation and Current EU Priorities

Bulgarian institutions are active in searching for international cooperation possibilities. Within the trend of Digital culture (Access to and preservation of cultural heritage) in FP6, we can mention the following projects where Bulgarian institutions participate as members:

- CALIMERA (participant ULISO)
- EPOCH (participant New Bulgarian University)
- MINERVAPLUS (participant – IMI-BAS as an associated member)
- PRESTOSPACE (participant Sirma AI Ltd)
- KT-DigiCult-BG is a project coordinated by IMI-BAS.

IMI-BAS was an initiator of the creation of the South-Eastern European Network for Digitisation of Scientific and Cultural Heritage[1], constituted with the signing of the Borovets declaration of 17 September 2003.

The current priorities under IST 2.5.10 (Access to and preservation of cultural and scientific resources) are targeted to:

- Enriched conceptual representations
- Advanced access methods
- Long-term preservationThe presentation of current Bulgarian setting in the previous sections shows that some Bulgarian institutions are trying to be in line with current developments.

Development of Local Tools vs. Adaptation of Existing Tools

In the digitisation work one crucial matter is what tools will be applied for the practical work. In the last years IMI gained experience in two approaches: *developing local tools* for support of specific task and *localisation of existing platforms* to the Bulgarian environment.

As a **home-made tool** we could mention XEditMan [Pavlov 2004]. This is a tool, which combines an editor and visualisation component for preparing and studying manuscript description of mediaeval manuscripts. Its interface is in Bulgarian and follows the local practices in cataloguing work. The descriptions of manuscripts are produced in XML format following the TEI P4 guidelines.

The experience with development and use of this tool is very positive, since it supports the performance of a specific task and facilitates the preparation of large amount of data in digital form.

As an example of **localised tool**, we could mention **SPWC**, Software Platform for archivist, librarian and philological Work in Cultural Institution. The platform offers a set of tools for document management in cultural institution including digitizing, cataloguing and transcription of primary sources. ILC are active in building specialized workstations for philological work for decades [Bozzi, Corradini 2004] and constantly improve the capabilities and the spread of use of their specialized tools. IMI worked on localisation of the software (translation of the user interface and documentation in Bulgarian), identification of experimental materials (in the case of Bulgaria this are local DTDs – manuscript and archival records), and training and dissemination activities. All abovementioned endeavours were in the frame of the project COMTOOCI (COMputational TOOls for the librarian and philological work in Cultural Institution). The project, supported by the CULTURE 2000 program and coordinated by the Institute for Computational Linguistics (ILC) – Pisa, Italy started in September 2004. In the forthcoming months a pilot installation of SPWC in the General Department of Archives will be done.

[1] http://www.ncd.matf.bg.ac.yu/?page=news&lang=en&file=declaration.htm, date of last visit 25.5.2005.

At the same time, we are studying the possible application of **ACT** [Ribarov 2004]. This software combines the presentation of manuscript images and annotated texts. Because of the high level of variety in mediaeval Slavonic manuscripts, the author chose an approach where previous human annotation of word forms is used in subsequent annotation activities.

We present a comparison of the features of **ACT** and **SPWC** in Table 1.

FEATURES	**ACT**	**SPWC**
Image analysis module		✓
Image representation module	✓	✓
Cataloguing module		✓
Representation of texts	✓	✓
Representation of variants	✓	✓
Multilingual support	✓	✓
Interface in different languages		✓
Annotation of various levels, up to morphology	✓	
Morphological annotation supported on ‘learning by example’ basis	✓	

Conclusion

Under the described lack of national policy, the various institutions in the cultural and scientific heritage sector have the freedom to design their own policies. Unfortunately, this is combined with lack of methodological, financial, technological and human resources support. On this setting, the Digitisation of Scientific Heritage Department at IMI has as a core part of its mission to contribute to the improvement of human resources qualification and support memory institutions through joint activities, which would lead to a difference in the future.

Bibliography

[Bozzi, Corradini 2004] Bozzi A., Corradini M.S. (2004). Aspects and methods of computer-aided textual criticism. In Digital Technology and Philological Disciplines. *Linguistica Computazionale* XX-XXI, (2004). Pisa-Roma, IEPI. 49-66.

[Markov 2004] Markov N., National Archives. In: International Journal Information Theories and Applications (special issue: Proceedings of the International Seminar Digitisation of Cultural and Scientific Heritage, Bansko, 27 August — 3 September 2004). Vol. 11 (2004), No. 3, pp. 282-283.

[Moussakova, Dipchikova 2004] Moussakova E, A. Dipchikova, The Role of the National Library in Preserving National Written Heritage. In: International Journal Information Theories and Applications (special issue: Proceedings of the International Seminar Digitisation of Cultural and Scientific Heritage, Bansko, 27 August — 3 September 2004). Vol. 11 (2004), No. 3, pp. 284-287.

[Pavlov 2004] Pavlov P., XML Presentation of Catalogue Data on Medieval Slavonic Manuscripts: Experience and Perspectives, In: Proceedings of the 33rd Conference of the Union of Bulgarian Mathematicians, Borovets, 1 – 4 April 2004, pp.236–240.

[Ribarov 2004] K. Ribarov *et al.*, "We present the ACT Tool". In: Scripta & e-Scripta, Volume 2, Institute of Literature, Bulgarian Academy of Sciences, Sofia. 2004.

Authors’ Information

Milena Dobreva — Chair of the Department *Digitisation of Scientific Heritage*, Institute of Mathematics and Informatics, BAS, Acad. G. Bonchev St., bl. 8, Sofia-1113, Bulgaria, e-mail: dobreva@math.bas.bg

Nikola Ikonomov — Chair of Laboratory on Phonetics and Speech Communication, Institute for Bulgarian Language, BAS, Shipchenski prohod 52, Sofia-1113, Bulgaria, e-mail: nikonomov@ibl.bas.bg.

CLASSIFICATION OF BIOMEDICAL SIGNALS USING THE DYNAMICS OF THE FALSE NEAREST NEIGHBOURS (DFNN) ALGORITHM[1]

Charles Newton Price, Renato J. de Sobral Cintra, David T. Westwick, Martin Mintchev

Abstract: Accurate and efficient analysis of biomedical signals can be facilitated by proper identification based on their dominant dynamic characteristics (deterministic, chaotic or random). Specific analysis techniques exist to study the dynamics of each of these three categories of signals. However, comprehensive and yet adequately simple screening tools to appropriately classify an unknown incoming biomedical signal are still lacking. This study is aimed at presenting an efficient and simple method to classify model signals into the three categories of deterministic, random or chaotic, using the dynamics of the False Nearest Neighbours (DFNN) algorithm, and then to utilize the developed classification method to assess how some specific biomedical signals position with respect to these categories. Model deterministic, chaotic and random signals were subjected to state space decomposition, followed by specific wavelet and statistical analysis aiming at deriving a comprehensive plot representing the three signal categories in clearly defined clusters. Previously recorded electrogastrographic (EGG) signals subjected to controlled, surgically-invoked uncoupling were submitted to the proposed algorithm, and were classified as chaotic. Although computationally intensive, the developed methodology was found to be extremely useful and convenient to use.

Keywords: Biomedical signals, classification, chaos, multivariate signal analysis, electrogastrography, gastric electrical uncoupling

ACM Classification Keywords: I.5.4 Pattern Recognition: Applications – Signal processing; J.3 Life and Medical Sciences

1. Introduction

Efficient accumulation of accurate knowledge from a wide variety of biomedical phenomena can be obtained from studying and analyzing their dynamics. This dynamics can be assessed by various sensors which monitor, measure, and transform biomedical phenomena into electrical signals that can be analyzed using contemporary electronics and signal processing techniques [1]. Generally, biomedical signals can be to an extent deterministic, random or chaotic [1, 2]. Deterministic signals have the characteristic of predictability, meaning that any future course of the signal could be predicted using some linear analysis tools [1]. Random signals are non-deterministic, in the sense that individual data points of the signal may occur in any order [1], limiting determining the predictability of the future course of the signal to purely stochastic analytical tools. Chaotic signals can be viewed as a connecting mesh between deterministic and random signals, exhibiting behaviour that is slightly predictable, non-periodic, and highly sensitive to initial conditions [2].

Observing that there are three general types of biomedical signals that can be encountered, accurate and efficient study and analysis of these signals can be facilitated by their proper identification as deterministic, random or chaotic, given that specific analysis techniques exist for each type of signals [1]. However, comprehensive and yet adequately simple screening tools to appropriately classify an unknown incoming biomedical signal with respect to these three categories are still lacking.

Recent paper by Gautama et al. [3] presents a method to classify an unknown incoming biomedical signal. The method provides an interpretation of a signal's deterministic and/or stochastic nature in terms of its predictability. Furthermore, it assesses the signal's linear or non-linear nature using surrogate data methods [3]. The result

[1] This study was supported in part by the Natural Sciences and Engineering Research Council of Canada, and by the Gastrointestinal Motility Laboratory (University of Alberta Hospitals) in Edmonton, Alberta, Canada

of this study provides a tool to measure the amount of determinism and randomness in a biomedical signal, useful for detecting a change in health conditions from monitored biomedical signals. However, this method can be seen as an analysis technique that can be applied to a signal once it is classified as deterministic, chaotic or random, rather than as a signal classifier.

The aim of the present work was to develop an efficient and simple method to classify biomedical signals into three categories (deterministic, chaotic or random), using a novel chaos analysis technique which we called the Dynamics of the False Nearest Neighbors (DFNN) algorithm. The proposed method extends the previously developed False Nearest Neighbors (FNN) algorithm [2, 4], to include dynamic FNN characteristics.

2. Methods

Understanding the suggested technique requires an introduction to multivariate signal analysis using state space representation, including time delay and embedding dimension calculations [2, 4, 5].

2.1. State Space Signal Representation

Biomedical signals are usually observed in one-dimensional form, and are represented discretely in the form of a time-domain vector, *s(n)*. It can be inferred that the one-dimensional time-domain vector, *s(n),* is a projection of the signal generator source, represented by an unknown but underlying multidimensional dynamic state vector *x(n)* [2]. The multidimensional dynamic state vector is composed of an unknown number of variables, represented through its dimension *d* [2, 6]. In these notations *n* denotes the current moment in the sampled time-domain.

The transition from a sampled one-dimensional time-domain signal *s(n)* to the corresponding sampled *d*-dimensional state space requires the application of Takens Theorem [6]. Takens Theorem represents a technique to reconstruct an approximation of the unknown dynamic state vector *x(n)* in *d*-dimensional state space by lagging and embedding the observed time series *s(n)*. This reconstructed approximation is the state vector *y(n)* = [*s*(n), *s*(n+T), *s*(n+2T),…, *s*(n+T(*d*–1))], composed of time-delayed samples of *s(n)*, where *T* is the time delay and *d* is the embedding dimension of the system. The accurate calculation of *d* and *T* guarantees through the Embedding Theorem [2], that the sequential order of the reconstructed state vector $y(n) \rightarrow y(n+1)$ is topologically equivalent to the generator state vector $x(n) \rightarrow x(n+1)$, allowing *y(n)* to represent without ambiguity the actual source of the observed multidimensional dynamic vector *x(n)* [2].

Each state space coordinate [s(n), s(n+T), s(n+2T),…,s(n+T(d+1))] constituting a component of *y(n)* defines a point in the state space. As time progresses, the dynamic trajectory of each point in time forms what is called an orbit. An orbit is mathematically defined as the numerical trajectory resulting from the solution of the system [2]. Each orbit constituting *y(n)* is presumed to come from an autonomous set of equations, and therefore, according to the Uniqueness Theorem [2], the trajectory of any orbit is unique and should not overlap with itself. The time delay *T* is an integer multiple of the sampling interval of the signal *s(n)* guaranteeing the extraction of maximal amount of information from the system [2]. The embedding dimension *d* is the minimal state space dimension required to unfold the main orbit of *x(n)* [2]. The main orbit of *x(n)*, known as the attractor, represents the set of points in state space visited by the other orbits of the system long after transients have died out [2].

2.2. Time Delay Calculation

The choice of an accurate time delay *T* guarantees that the time-delayed state space coordinates forming *y(n)* are independent from each other [2]. Choosing too small of a value for *T* clusters the data in state space, while choosing too large of a value for *T* causes the disappearance of the relationships between the points in the attractor [7, 8]. The independence between two coordinates of *y(n)* can be assessed using the mutual information (MI) function [2]. The MI between two *y(n)* coordinates, e.g., *s(n)* and *s(n+T)*, is measured in bits by:

$$MI = log_2 \left\{ \frac{P[\mathrm{s(n)}, s(n+T)]}{P[s(n)]P[s(n+T)]} \right\}, \tag{1}$$

where $P[s(n), s(n+T)]$ is the joint probability density function (JPDF) of *s(n)* and *s(n+T)*. The average mutual information (AMI) of the JPDFs of all coordinates is calculated by:

$$AMI(T) = \sum_{s(n),s(n+T)} P[s(n),s(n+T)] log_2 \left\{ \frac{P[s(n),s(n+T)]}{P[s(n)]P[s(n+T)]} \right\}. \quad (2)$$

The first minimum of the AMI function provides the optimal time delay T, and assures the independence between the coordinates of the multidimensional vector *y(n)* [2, 7, 8].

2.3. Embedding Dimension Calculation

The signal reconstruction in state space requires a dimension that will guarantee no overlap of the trajectory of the orbit constituting *y(n)*. This optimal dimension is obtained after calculating the percentage of False Nearest Neighbours (FNN) between points in state space. FNNs are calculated using reconstructed state space vectors *y(n)* at different embedding dimensions but a constant time-delay [9]. It is accepted that when the FNN percentage drops to zero, the minimum required dimension to unfold the system into its original state around its attractor is reached, which also guarantees that the orbit is unique [2, 9]. The calculation of the FNNs requires the measurement of a distance R_d, defined as the radius between neighbouring vectors in consecutive dimensions. This procedure is referred to as the FNN algorithm [2, 9]. The square of the Euclidian distance representing R_d as seen in dimension *d* is:

$$R_d(n)^2 = \sum_{m=1}^{d} \left[s(n+T(m-1)) - s^{NN}(n+T(m-1)) \right]^2, \quad (3)$$

where *n* is the current index of the discrete signal (in this case *s(n)*) and s^{NN} is the nearest neighbour (NN) of *s(n)*.

The square of the Euclidian distance in dimension *d+1* becomes:

$$R_{d+1}(n)^2 = \sum_{m=1}^{d+1} \left[s(n+T(m-1)) - s^{NN}(n+T(m-1)) \right]^2 = R_d(n)^2 + \left(s(n+dT) - s^{NN}(n+dT) \right)^2 \quad (4)$$

The change in distance between the points at dimensions *d* and *d+1* is:

$$\sqrt{\frac{R_{d+1}^2(n) - R_d(n)^2}{R_d(n)^2}} = \frac{\left| s(n+dT) - s^{NN}(n+dT) \right|}{R_d(n)}. \quad (5)$$

Determining the existence of a false nearest neighbour depends on how the distance between state space vectors behaves as the calculations progress in consecutive dimensions. If the distance increases significantly with the increment of the embedding dimension, then the vectors are false neighbours, and their closeness results from the reconstruction dynamics of the system, not from its underlying dynamics [2, 9]. If the distance is restricted within a certain threshold level close to the state space points, then the state space points are real neighbours resulting from the dynamics of the system. The embedding dimension that adequately represents the system is the dimension that eliminates most of the false neighbours, leaving a system whose trajectories are positioned in state space due to their underlying dynamics, not to their reconstruction dynamics. Figure 1 shows an example of the results of the FNN algorithm applied to model deterministic, chaotic and random signals.

2.4. Dynamics of FNN (DFNN) Algorithm

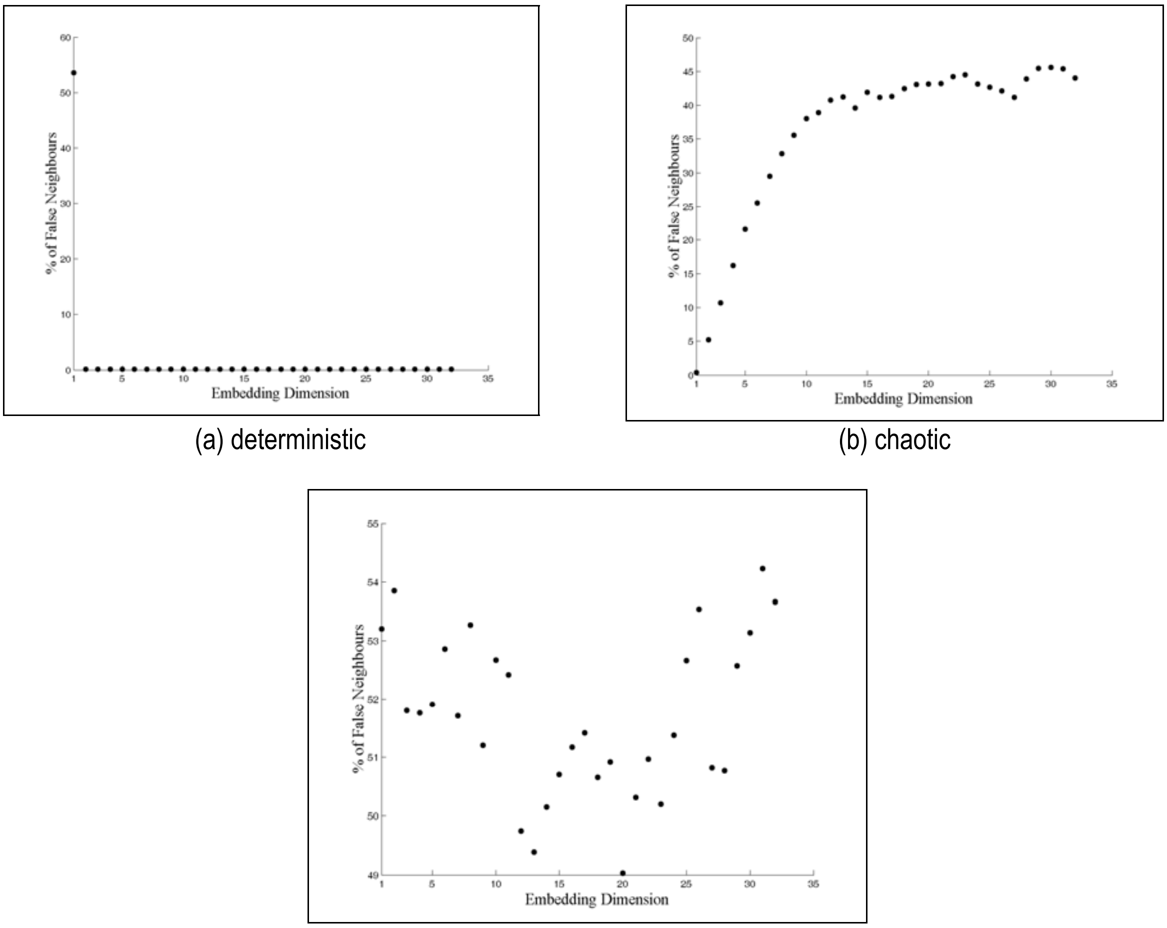
The FNN algorithm is utilized to determine the minimal embedding dimension required to completely reconstruct in state space the source *x(n)* of a one-dimensional time series *s(n)* [2, 9]. Theoretically, the minimal embedded dimension is obtained when the percentage of FNN at a given dimension reaches zero. In practice, not all signals tested through the FNN algorithm reach zero percent FNNs. Two main factors are responsible for this fact. The first is that the larger the embedding dimension of the system, the larger the number of signal samples required for the FNN algorithm [2]. The minimum number of samples required for a given embedding dimension is given by the following equation:

$$m_d = \sqrt{2} \times (e)^d \quad , \tag{6}$$

where *d* is the number of embedding dimensions. The second factor is that an established pattern, or attractor, underlying the dynamics of the system, simply does not exist, as it is the case with white noise [5], for which the FNN algorithm does not reach zero percent FNNs. Therefore, an FNN algorithm failing to converge to zero percent FNN indicates a signal which dimension is too high for the number of samples available.

The important benefit of the proposed DFNN algorithm is that it analyzes the reconstruction dynamics of the signals submitted to the FNN algorithm. Thus, in contrast to the well-established FNN algorithm [9], the DFNN approach does not aim at finding the optimal embedding dimension of the processed signals, but focuses on the processing of the curve representing the FNN dynamics as a function of the embedded dimensions (see Figure 1).

Figure 1 – Sample FNN dynamics of models of (a) deterministic, (b) chaotic, and (c) random signals.



(a) deterministic

(b) chaotic

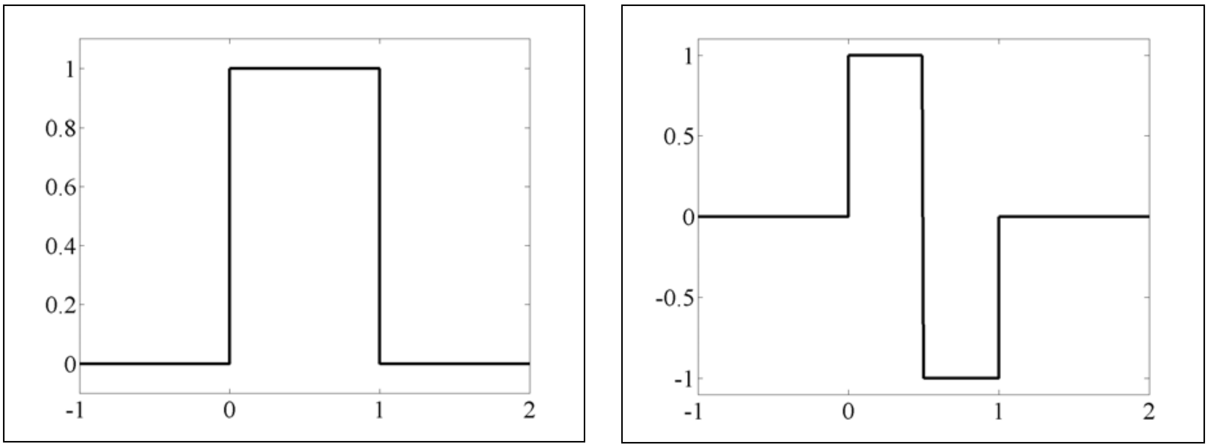
(c) random signals

The independence of the coordinates of the reconstructed state space vector *y(n)* is guaranteed by an accurate choice of time delay *T* [2, 7, 8]. According to the Uniqueness Theorem [2], each reconstruction is unique, therefore the use of an inaccurate embedding dimension when the signal is reconstructed still guarantees a unique representation of the signal, even though it would not represent properly its underlying dynamics until the correct embedding dimension is calculated with the help of the FNN algorithm. What this implies is that the percent of FNNs incrementally calculated at different embedding dimensions in the algorithm is a unique property of the signal under consideration, and therefore, the dynamics of these percent FNNs could also be regarded as a way to represent the signal.

The proposed DFNN algorithm uses a wavelet based pattern recognition technique to classify sampled signals as deterministic, chaotic, or random. The technique is based on analyzing the reconstruction dynamics of the FNNs at consecutively increasing embedding dimensions, ranging from 1 to 32 with the help of wavelet decomposition [10]. The limit of 32 was established due to the fact that 32 is an adequate number of points for a meaningful statistical analysis, because for dyadic wavelet analysis [11] a number that is a power of two is needed, and because it has been shown that the Rossler chaotic system can have an embedding dimension of 25 [12]. It is important to clarify that our aim was not to find the optimal embedding dimension of each signal, but to analyze their reconstruction dynamics as they were submitted to the FNN algorithm.

2.5. Wavelet Decomposition

Wavelet decomposition analysis can be utilized to quantify the shape of the data points extracted from the FNN algorithm [13]. Wavelet analysis coefficients can indicate the resemblance between the shape of a wavelet and a signal. If the resemblance is high, the signal energy is concentrated in few wavelet coefficients. Otherwise, the energy content of the signal is spread throughout these coefficients [14]. Therefore, the aim to quantify the shape of the data points extracted from the FNN algorithm requires (i) finding the wavelet that matches best the wave-shape of the FNN dynamics [14], and (ii) calculating the corresponding wavelet analysis coefficients.



(a) (b)

Figure 2 – (a) Haar scaling function and (b) wavelet.

The proposed DFNN algorithm attempts to classify sampled signals as deterministic, chaotic, or random. An example wave-shape for each signal can be seen in Figure 1. Notice that the shape of the FNN dynamics for each type of signal is different, and that in order to make a comparison one signal type needs to be chosen to be a reference. The deterministic sampled signal pattern was chosen as a reference for determining the wavelet due to its simple shape. The Haar system wavelet was selected, since it matched best the pulse-like wave-shape of deterministic FNN dynamics (compare Figure 1a to Figure 2). Therefore, we hypothesized that through the analysis of the wavelet coefficients, a distinction could be made between deterministic, chaotic and random signal patterns using the FNN dynamics associated with a particular signal.

2.6. Testing the DFNN Algorithm

2.6.1. Categories of Signal Models

In order to test the validity of the proposed DFNN algorithm, signal models with known characteristics were utilized, classified in the existing literature as deterministic, chaotic and random signals [1, 2].

All model signals were sampled frequently enough to comply with the Nyquist Theorem, guaranteeing sufficient digital samples to represent each signal [15]. The length of each signal was 6000 points. The deterministic group of signals included sine, rectangular, triangular, square, sawtooth, quadratic and Dirichlet functions. The models of chaotic signals included the Mackey-Glass Map (MGM), Henon, Ikeda and logistic maps, as well as the Lorentz, and quadratic systems. The random signal realizations were based on the following statistical distributions: Rayleigh, exponential, beta, χ^2 (chi-squared), gamma, impulsive, normal, and uniform.

2.6.2. Testing Protocol

The three categories of model signals (deterministic, chaotic and random) were submitted to the FNN algorithm, where the percent FNN was calculated and recorded for each embedding dimension up to 32. The resulting 32-point representation of each signal (see samples in Figure 1), were submitted to wavelet analysis decomposition in order to evaluate its corresponding wavelet coefficients. The wavelet coefficients were calculated for each model signal in each corresponding category (deterministic, chaotic and random) and were statistically analyzed by calculating the sample mean and standard deviation. The sample mean and standard deviation were considered distinguishable features to be submitted to the Fuzzy C-means clustering algorithm [16]. The Fuzzy C-means algorithm groups these distinguishable features into clusters. Each cluster provides a centroid, representative of each model signal [16]. This centroid is an important feature corresponding to each cluster that was pivotal for categorizing each signal.

2.6.3. Assessing the Robustness of the DFNN Algorithm

The limited amount of model signals used to test the DFNN algorithm could affect the statistical significance of the results. Therefore, two additional tests were designed to further strengthen the robustness of the DFNN algorithm. The first test involved shuffling the data points of deterministic and chaotic signals to transform them into random signals. These shuffled surrogate signals were then submitted to the DFNN algorithm, and categorized according to their position relative to their nearest cluster using the similarity measure (SM) [17].

The chosen SM is based on the Euclidian metric and is represented by a number between zero and unity, zero representing no similarity, and unity representing maximal similarity. It is calculated by the following equation:

$$SM = \frac{1}{1+l}, \tag{7}$$

where l is the Euclidian distance between the centroid of a given cluster (deterministic, chaotic, or random) and the point under consideration.

The second test involved filtering random model signals using a fourth order low-pass digital Butterworth filter. Our hypothesis was that as the random signals undergo low-pass filtering at gradually reduced cut-off frequencies, the level of randomness would be reduced, and the positioning of these new signals in relation to each of the deterministic, chaotic and random centroids, would change after being submitted to the DFNN algorithm. We expected that with the low-pass filtering of these signals with gradually decreasing cut-off frequencies, the level of randomness would drops, and the positioning of the signals would shift away from the random centroid towards the chaotic and deterministic centroids.

2.6.4. Experiment with Electrogastrographic Signals to Detect Gastric Uncoupling

Gastric uncoupling is the loss of electrical synchronization in the stomach [18]. Since gastric motility is electrically controlled, such uncoupling may result in clinical complications such as gastroparesis [18]. In a canine experiment performed by Mintchev et al [18], gastric electrical uncoupling was artificially induced by surgically inhibiting the propagation of electric potentials throughout the length of the stomach using circumferential surgical cuts in the stomach physically separating sections of the organ. Electrogastrography (EGG) is a non-invasive method to record gastric electrical activity [19], and was utilized in this experiment in an attempt to validate its ability to recognize gastric electrical uncoupling. Three kinds of 8-channel EGG signals were recorded from 16 dogs: basal (B), after the first circumferential cut (FC), and after the second cut (SC), each representing three different levels of electrical desynchronization: (i) no uncoupling; (ii) mild induced uncoupling; and (iii) severe induced uncoupling. It has been shown that the amount of electrical uncoupling exhibited in the recorded signals increased with the number of circumferential cuts.

Utilizing the proposed DFNN algorithm, the B, FC, and SC signals from the EGG recordings were tested to assess how each of these signals positions itself with respect to the clusters of deterministic, chaotic, and random signals pre-identified in the experiments with the model signals. We hypothesized that since EGG signals were found to be chaotic [20], they would position themselves in the chaotic cluster of the plot, and that uncoupling will be detected by noticing that basal EGG signal patterns position closer to the deterministic cluster, while SC signal patterns position themselves closer to the random cluster.

Similarly to the model signals, each of the three types of EGG signals (B, FC, SC) was subjected to the DFNN algorithm. The resulting 128 wave-shape patterns for each state (8 EGG channels per state from each of the 16 dogs) were decomposed using wavelet analysis, and the mean and standard deviation were calculated for the coefficients of each EGG signal type with the aim to show how the B, FC, and SC signals position themselves with respect to the deterministic, chaotic and random regions defined using the model signals. The position of the B, FC, and SC signal patterns in each cluster were quantified using the same SM technique used to test the robustness of the DFNN algorithm [17].

3. Results

3.1. DFNN Algorithm

The percent FNN up to a dimension of 32 were calculated for each model signal using a software package called Visual Recurrence Analysis (VRA) [21]. The calculated dimensions for each model signal resulted in unique wave-shapes (see examples in Figure 1). A total of 49 wave shapes were obtained: 9 from the deterministic model signals, 26 from the chaotic model signals, and 14 from the random model signals. Each of these wave-shapes was submitted to wavelet analysis decomposition using the Haar wavelet, the decomposition resulting in 32 coefficients per model signal. The mean and standard deviation of approximation coefficients per model signal were calculated, with a sample of the results shown in Table 1. A plot of the mean against the standard deviation for each model signal was built (Figure 3). Notice the tendency for the deterministic signals to cluster on the left of the plot, the chaotic signals to cluster near the centre of the plot, and the random signals to cluster to the right of the plot.

To formalize the uniqueness of each group as representative of each model signal, the Fuzzy C-means algorithm was applied to the points of Figure 3. This resulted in a centroid being defined for each model signal group (graphically shown in Figure 3, numerically shown in Table 2), to clearly partition the deterministic, chaotic, and random model signals into specific regions of the plot.

Table 1 – Sample means and standard deviations of the wavelet coefficients obtained from some model signals.

Signal Type	Signal Model	Sample Mean	Sample Standard Deviation
Deterministic	Sine	3.2306	12.4698
	Triangular	2.6224	10.4899
	Square	0.0018	0.0071
	Sawtooth	0.0301	0.0071
Chaotic	Henon	36.7863	11.2378
	Ikeda	55.6975	2.2675
	Logistic	48.7745	19.1534
	Lorentz	40.1871	4.1175
	MGM	25.7374	11.1079
	Quadratic	37.8974	17.3651
Random	Impulse	70.3845	0.5949
	Normal	69.7791	0.6863
	Uniform	70.7354	0.6900

Table 2 - Centroid coordinates for each signal model group.

Signal Type	Centroid coordinate
Deterministic centroid	Mean: 3.3877 SD: 7.2483
Chaotic centroid	Mean: 40.9797 SD: 10.7933
Random centroid	Mean: 69.9643 SD: 1.8843

Table 3 – Means and standard deviations of the similarity measures from all EGG signals.

Signal Type	SM to Deterministic Cluster	SM to Chaotic Cluster	SM to Random Cluster
Basal (B)	Mean: 0.0304 SD: 0.0038	Mean: 0.1329 SD: 0.0449	Mean: 0.0274 SD: 0.0045
First Cut (FC)	Mean: 0.0293 SD: 0.0039	Mean: 0.1549 SD: 0.0586	Mean: 0.0278 SD: 0.0031
Second Cut (SC)	Mean: 0.0298 SD: 0.0039	Mean: 0.1546 SD: 0.0595	Mean: 0.0271 SD: 0.0025

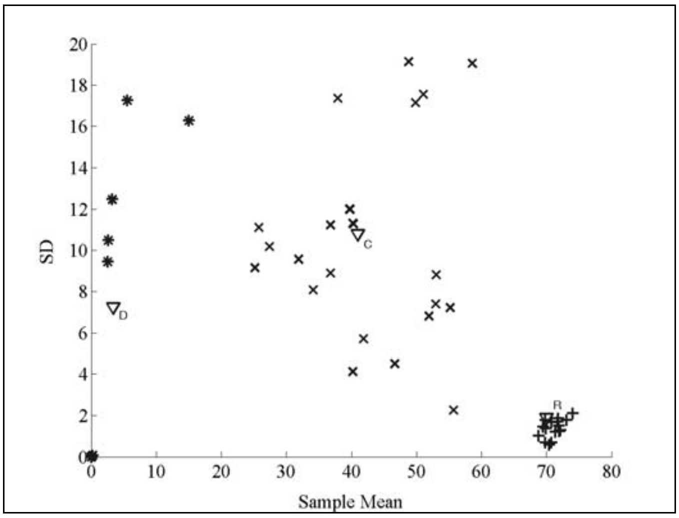


Figure 3 – Sample means and standard deviations of the wavelet coefficients obtained from deterministic (∗), chaotic (×), and random (+) model signals. The centroids (∇) for each cluster were calculated using the Fuzzy C-Means algorithm.

3.2. Robustness of the Algorithm

In the first test performed, the data points of two deterministic and chaotic signals were shuffled and the resulting signals submitted to the DFNN algorithm. All of the shuffled signals positioned themselves in the random region as expected (Figure 4).

The second test for robustness involved filtering random signals using a fourth order low-pass digital Butterworth filter. Three types of random signals were utilized (represented by exponential, uniform and normal probability density functions) and filtered at different normalized cut-off frequencies (0.8, 0.5, 0.3, 0.1, 0.01). With the filtering at decreasing cut-off frequencies, the signals shifted their position further away from the random centroid. This tendency is visualized in Figure 5.

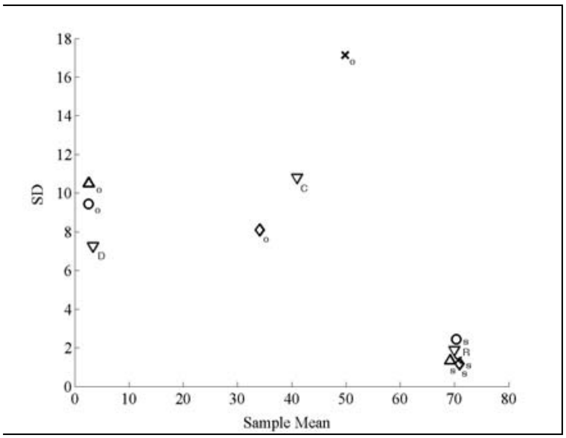


Figure 4 – Positions of the shuffled (**s**) and the original (**o**) deterministic (❍ - sine wave, Δ - triangular wave) and chaotic (× - logistic map, ◊ - Lorenz map) signals.

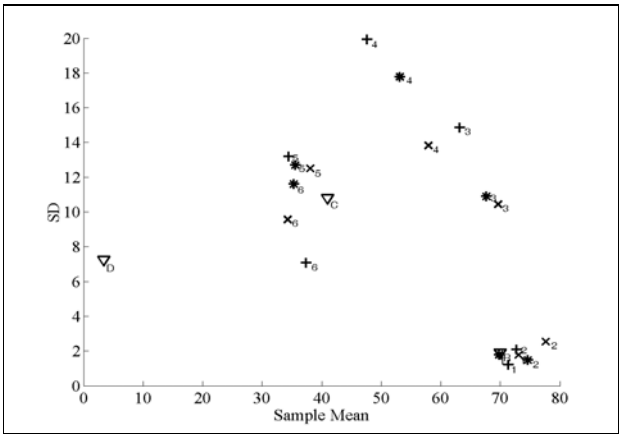
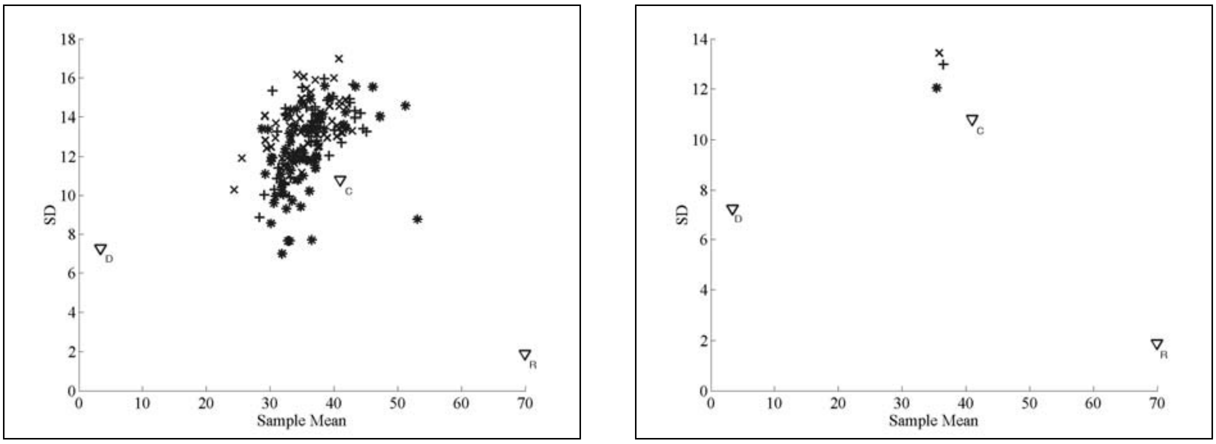


Figure 5 – Three random signals with different probability density functions, exponential (∗), uniform (+) and normal (×), filtered with a fourth order low-pass digital Butterworth filter at the following normalized cut-off frequencies: (**1**) no filtering, (**2**) 0.8, (**3**) 0.5, (**4**) 0.3, (**5**) 0.2, (**6**) 0.01.

3.3. Detection of Gastric Electrical Uncoupling

Gastric electrical uncoupling as assessed by the DFNN algorithm can be demonstrated by a single representative point for each EGG signal type, calculated by obtaining the mean of the means and the standard deviations for each of the B, FC and SC signal wavelet coefficients (Figure 6a), resulting in three representative points shown in Figure 6b. Quantitatively, the calculations were performed using Equation 6, where the similarity measure SM was calculated for each of the B, FC, and SC signals with respect to the centroid of each of the deterministic, chaotic and random regions of the plot obtained from the model signals. The overall means and standard deviations for all SM calculations are shown in Table 3. Notice that the similarity of the B, FC and SC signals to the chaotic region was quite strong due to the high SM value, while their similarity to the deterministic and random regions was very weak due to a low SM value. Nevertheless, slight shift was noted towards the random centroid after the first circumferential cut (point FC on Figure 6b), and after the second cut the standard deviation of the obtained wavelet coefficients increased notably (point SC on Figure 6b).



(a) (b)

Figure 6 **(a)** Sample mean and standard deviation of the wavelet coefficients from B (∗), FC (×), and SC (+) EGG signals.
(b) Positioning of the mean coordinate points representing B (∗), FC (×), and SC (+) signals with respect to the centroids (∇) obtained from model signal clusters.

4. Disscussion

In the recent years chaos analysis of biomedical signals evolved into a powerful digital signal processing avenue [2, 4, 5, 22], often overshadowing the well established deterministic and stochastic signal processing tools [1]. However, a comprehensive methodology to determine the adequate digital signal processing tool set (deterministic, chaotic or stochastic) for a specific biomedical signal is still lacking.

In this study we developed an innovative procedure to examine whether given biomedical signals of interest belong to predefined clusters of deterministic, chaotic or random patterns obtained from model signals typical for each of these three categories. The intent was to algorithmically facilitate an informed quantitative decision on which signal processing tools were better suited for the processing of the biomedical signals under consideration. The proposed DFNN algorithm, combined with wavelet decomposition and subsequent statistical analysis were found to be excellent candidates for fulfilling this mission. The research was motivated by the observation that the shapes of the curves produced by the FNN algorithm appeared to be visually related to the signal type. Therefore, a pattern recognition technique based on a dyadic wavelet expansion of the FNN characteristic was developed. The method was tested on a selected set of artificially constructed signals, and then used to assess how some specific biomedical signals [23] position themselves in these categories. It is important to note that the method was tested on a selected set of model signals, and thus further testing with a variety of model signals might be appropriate to fully assess the capabilities and the limitations of the proposed technique.

The methodology resulted in a convenient and very clear clustering of deterministic, chaotic, and random signal patterns extracted from model signals (see Figure 3). Subsequent analysis of electrogastrographic signals in different states (basal, after mild invoked uncoupling, and after severe invoked uncoupling) confirmed previous suggestions that the EGG signals are inheritantly chaotic [20, 24]. Moreover, it was observed that the dominant chaotic nature of these signals, demonstrated by the fact that the DFNN algorithm resulted in their classification well in the middle of the predefined chaotic cluster (see Figure 4), most likely precluded a clear and significant shift from the basal pattern (B) when the signals recorded after the invoked uncouplings (FC and SC) were considered.

5. Conclusion

An innovative technique for classifying biomedical signals in three categories, deterministic, chaotic, and random was developed. The methodology was quantitatively tested using model signals belonging to each of these three categories, and actual electrogastrographic signals subjected to experimentally controlled uncoupling. The technique could be very useful in making an informed decision which digital signal processing toolset would be most appropriate for a specific type of biomedical signals.

Acknowledgement

This study was supported in part by the Natural Sciences and Engineering Research Council of Canada, and by the Gastrointestinal Motility Laboratory (University of Alberta Hospitals) in Edmonton, Alberta, Canada

References

[1] R. E. Challis and R. I. Kitney, "Bio-Medical Signal-Processing (in 4 Parts). 1. Time-Domain Methods," *Medical & Biological Engineering & Computing*, vol. 28, pp. 509-524, 1990.

[2] H. D. I. Abarbanel, *Analysis of observed chaotic data*. New York: Springer, 1996.

[3] T. Gautama, D. P. Mandic, and M. M. Van Hulle, "A novel method for determining the nature of time series," *Biomedical Engineering, IEEE Transactions on*, vol. 51, pp. 728-736, 2004.

[4] B. Chen and N. Wang, "Determining EMG embedding and fractal dimensions and its application," presented at Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE, 2000.

[5] B. Huang and W. Kinsner, "Impact of low-rate sampling on the reconstruction of ECG in phase-space," presented at Electrical and Computer Engineering, 2000 Canadian Conference on, 2000.

[6] F. Takens, "Detecting strange attractors in turbulance," in *Lecture notes in mathematics*, vol. 898, D. A. Rand and L. S. Young, Eds. Berlin: Springer, 1981, pp. 366-381.

[7] W. Liebert and H. G. Schuster, "Proper choice of the time delay for the analysis of chaotic time series," *Physics Letters A*, vol. 142, pp. 107-111, 1989.

[8] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, pp. 1134-1140, 1986.

[9] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical reconstruction," *Physical Review A*, vol. 45, pp. 3403-3411, 1992.

[10] A. M. Reza, "From Fourier Transform to Wavelet Transform, Basic Concepts," Spire Lab, UWM, White Paper October 27 1999.

[11] M. Unser, "Wavelet Theory Demystified," *IEEE Transactions on Signal Processing*, vol. 51, pp. 470 - 483, 2003.

[12] G. Chen, G. Chen, and R. J. P. de Figueiredo, "Feedback control of unknown chaotic dynamical systems based on time-series data," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on [see also Circuits and Systems I: Regular Papers, IEEE Transactions on]*, vol. 46, pp. 640-644, 1999.

[13] A. Ohsumi, H. Ijima, and T. Kuroishi, "Online detection of pulse sequence in random noise using a wavelet," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 47, pp. 2526-2531, 1999.

[14] S. G. Mallat, *A Wavelet Tour of Signal Processing, 2nd ed.*: Academic Press, 1999.

[15] P. P. Vaidyanathan, "Generalizations of the sampling theorem: Seven decades after Nyquist," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on [see also Circuits and Systems I: Regular Papers, IEEE Transactions on]*, vol. 48, pp. 1094-1109, 2001.

[16] S. Miyamoto, "An overview and new methods in fuzzy clustering," presented at Knowledge-Based Intelligent Electronic Systems, 1998. Proceedings KES '98. 1998 Second International Conference on, 1998.

[17] D. S. Yeung and X. Z. Wang, "Using a neuro-fuzzy technique to improve the clustering based on similarity," presented at Systems, Man, and Cybernetics, 2000 IEEE International Conference on, 2000.

[18] M. P. Mintchev, S. J. Otto, and K. L. Bowes, "Electrogastrography can recognize gastric electrical uncoupling in dogs," *Gastroenterology*, vol. 112, pp. 2006-2011, 1997.

[19] M. A. M. T. Verhagen, L. J. Van Schelven, M. Samson, and A. J. P. M. Smout, "Pitfalls in the analysis of electrogastrographic recordings," *Gastroenterology*, vol. 117, pp. 453-460, 1999.

[20] J. Y. Carre, A. Høst-Madsen, K. L. Bowes, and M. P. Mintchev, "Analysis of the dynamics of the level of chaos associated with gastric electrical uncoupling in dogs," *Medical and Biological Engineering and Computing*, vol. 39, pp. 322-329, 2001.

[21] E. Kononov, "Visual Recurrence Analysis (VRA)," 4.5 ed. http://home.netcom.com/~eugenek, 2003.

[22] I. Yaylali, H. Kocak, and P. Jayakar, "Detection of seizures from small samples using nonlinear dynamic system theory," *Biomedical Engineering, IEEE Transactions on*, vol. 43, pp. 743-751, 1996.

[23] C. Newton Price and M. P. Mintchev, "Quantitative evaluation of the dynamics of external factors influencing canine gastric electrical activity before and after uncoupling," *Journal of Medical Engineering and Technology*, vol. 26, pp. 239-246, 2002.

[24] G. Lindberg, "Is the electrogastrogram a chaotic parameter?," In: Proceedings of the Fourth International Workshop on Electrogastrography, San Francisco, CA, May, 1996.

Authors' Information

Charles Newton Price – M.Sc. graduate student,
Department of Electrical and Computer Engineering; University of Calgary, Calgary, Alberta, Canada T2N 1N4

Renato J. de Sobral Cintra – Ph.D. graduate student,
Department of Electrical and Computer Engineering; University of Calgary, Calgary, Alberta, Canada T2N 1N4
Department of Electronics & Systems, Federal University of Pernambuco; Recife, Pernambuco, Brazil

David T. Westwick – Associate Prof., Dr.,
Department of Electrical and Computer Engineering; University of Calgary, Calgary, Alberta, Canada T2N 1N4

Martin P. Mintchev – Prof., Dr.,
Department of Electrical and Computer Engineering; University of Calgary; 2500 University Drive NW; Calgary, Alberta, Canada T2N 1N4
Department of Surgery, University of Alberta; Edmonton, Alberta, Canada T6G 2B7
phone: (403) 220-5309; fax: (403) 282-6855; e-mail: mintchev@enel.ucalgary.ca

BOTTLENECK PROBLEM SOLUTION USING BIOLOGICAL MODELS OF ATTENTION IN HIGH RESOLUTION TRACKING SENSORS

Alexander Fish, Orly Yadid-Pecht

Abstract: Every high resolution imaging system suffers from the bottleneck problem. This problem relates to the huge amount of data transmission from the sensor array to a digital signal processing (DSP) and to bottleneck in performance, caused by the requirement to process a large amount of information in parallel. The same problem exists in biological vision systems, where the information, sensed by many millions of receptors should be transmitted and processed in real time. Models, describing the bottleneck problem solutions in biological systems fall in the field of visual attention. This paper presents the bottleneck problem existing in imagers used for real time salient target tracking and proposes a simple solution by employing models of attention, found in biological systems. The bottleneck problem in imaging systems is presented, the existing models of visual attention are discussed and the architecture of the proposed imager is shown.

Keywords: Bottleneck problem, image processing, tracking imager, models of attention

ACM Classification Keywords: B.7.0 Integrated circuits: General, I.4.8 Image processing and computer vision: scene analysis: tracking

1. Introduction

Driven by the demands of commercial, consumer, space and security applications, image sensors became a very hot topic and a major category of high-volume semiconductor production [1]. This is due to the imminent introduction of imaging devices in high volume consumer applications such as cell phones, automobiles, PC-based video applications, "smart" toys and, of course, digital still and video cameras. While most of consumer applications are satisfied with relatively low-resolution imagers, image sensors used for target tracking in space, navigation and security applications require high spatial resolution. In addition, these tracking sensors are usually supposed to provide real time tracking after multiple targets in the field of view (FOV), such as stars, missiles and others. The demands for high resolution and real time performance result in a bottleneck problem relating to the large amount of information transmission from the imager to the digital signal processing (DSP) or processor and in bottleneck in performance, caused by the requirement to process a large amount of information in parallel. The simple solution to the bottleneck in performance is to use more advanced processors to implement the required tracking algorithms or to use dedicated hardware built specially for the required algorithm implementation [2]. However, the solution to the performance bottleneck still does not relax the requirement for the large data transmission between sensor and processor.

The same problem exists in biological vision systems. Compared to the state-of-the-art artificial imaging systems, having about twenty millions sensors, the human eye has more than one hundred million receptors (rods and cones). Thus, the question is how the biological vision systems succeed to transmit and to process such a large amount of information in real time? The answer is that to cope with potential overload, the brain is equipped with a variety of attentional mechanisms [3]. These mechanisms have two important functions: (a) attention can be used to select relevant information and/or to ignore the irrelevant or interfering information; (b) attention can modulate or enhance the selected information according to the state and goals of the perceiver. Numerous research efforts in physiology were triggered during the last five decades to understand the attention mechanism [4-12]. Generally, works related to physiological analysis of the human attention system can be divided into two main groups: those that present a spatial (spotlight) model for visual attention [4-6] and those following object-based attention [7-12]. The main difference between these models is that the object-based theory is based on the assumption that attention is referenced to a target or perceptual groups in the visual field, while the spotlight theory indicates that attention selects a place at which to enhance the efficiency of information processing.

The design of efficient real time tracking systems mostly depends on deep understanding of the model of visual attention [12-14]. This paper briefly describes spotlight and object-based models of attention and proposes a solution for the bottleneck problem in image systems for salient targets tracking based on the study and utilization of the spatial (spotlight) model of attention. Two possible sensor architectures are presented and discussed.

Section 2 briefly describes the bottleneck problem in high resolution imaging systems. A review of existing models of attention is presented in Section 3. Section 4 presents descriptions of two sensor architectures, comparing it with the existing spatial (spotlight) models of attention. Section 5 concludes the paper.

2. The Bottleneck Problem in High Resolution Image Systems

As mentioned in section 1, two bottlenecks exist in high-resolution image systems: the data transmission bottleneck and performance bottleneck. The solution for the bottleneck in performance relates to increasing the processing power. This can be performed in the following ways: (a) to use more advanced processors to implement the algorithms, (b) to use dedicated hardware built especially for the required algorithms implementation and (c) employing more than one DSP/processor. Although these solutions are expensive and can dramatically increase the cost of the whole system, they provide simple and trustworthy solutions. Fig. 1 shows an example of such a kind of an imaging system. As can be seen, the system consists of the image sensors array (can either be implemented as a Charge Coupled Device (CCD) or as a standard CMOS imager, as will be described below), a number of processors (or DSPs) and a memory.

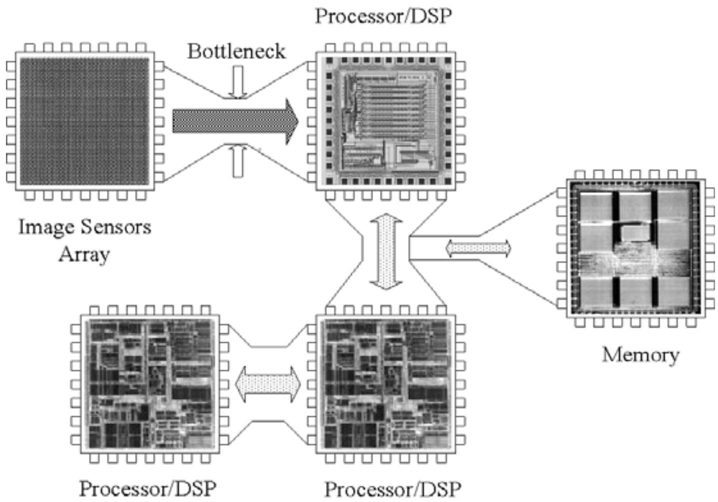


Fig.1. An example of a typical image system, incorporating an image sensor array, processors/DSPs for image processing and memory

The data transmission bottleneck from the image sensors array to the processor/DSP (see Fig. 1) is more difficult for solution. The most efficient way to solve the problem is on-chip implementation of some algorithms that can relax the information bottleneck by reducing the amount of data for transmission. However, this solution is almost impossible in CCD sensors, which cannot easily be integrated with standard CMOS analog and digital circuits due to additional fabrication complexity and increased cost. On the other hand, CMOS technology provides the possibility for integrating imaging and image processing algorithms functions onto a single chip, creating so called "smart" image sensors. Unlike CCD image sensors, CMOS imagers use digital memory style readout, using row decoders and column amplifiers. This readout allows random access to pixels so that selective readout of windows of interest is allowed. In this paper all further discussions will be related to CMOS image sensors.

Fig. 2 shows an example of an imaging system employing a "smart" CMOS image sensor and a single processor/DSP. As can be seen, in this system the image processing can be performed at three different levels: (a) at the pixel level – CMOS technology allows insertion of additional circuitry into the pixel, (b) on-chip image processing and, of course (c) image processing by processor/DSP. The larger amount of processing performed in the first two levels, the less amount of information necessary to be transmitted from the CMOS imager to the DSP. In addition, smaller computation resources are required.

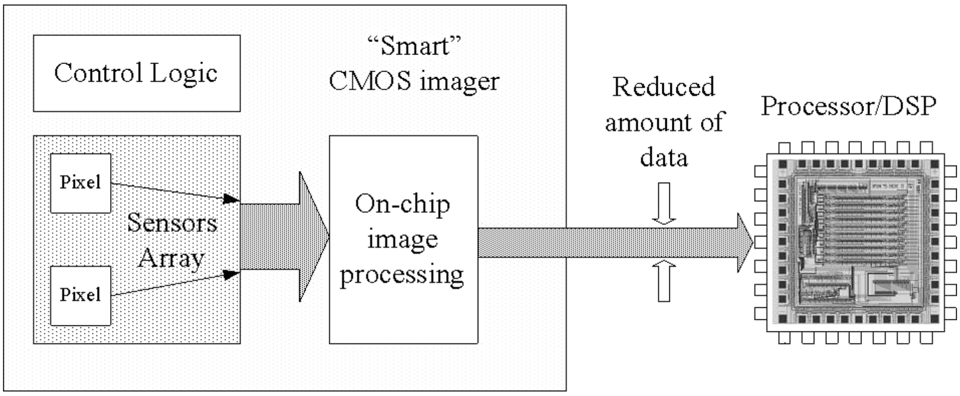


Fig.2. An example of an imaging system, employing a "smart" CMOS image sensor with on-chip processing and processors/DSPs for image processing

Generally, image processing at early stages (pixel level and on-chip level) can solve both data transmission bottleneck and performance bottleneck problems. Tracking imager architectures, proposed in this paper, will try to imitate the models of attention emphasizing on the requirement to perform the most of processing at early stages.

3. Visual Models of Attention

Although many research efforts were triggered during the last decades and numerous models of attention have been proposed over the years, there is still much confusion as to the nature and role of attention. Generally, two models of attention exist: spatial (spotlight) or early attention and object-based, or late attention. While the object-based theory suggests that the visual world is parsed into objects or perceptual groups, the spatial (spotlight) model purports that attention is directed to unparsed regions of space. Experimental research provides some degree of support to both models of attention. While both models are useful in understanding the processing of visual information, the spotlight model suffers from more drawbacks than the object-based model. However, the spotlight model is simpler and can be more useful for tracking imager implementations, as will be shown below.

3.1 The Spatial (Spotlight) Model

The model of spotlight visual attention mainly grew out of the application of information theory developed by Shannon. In electronic systems, similar to physiological, the amount of the incoming information is limited by the system resources.

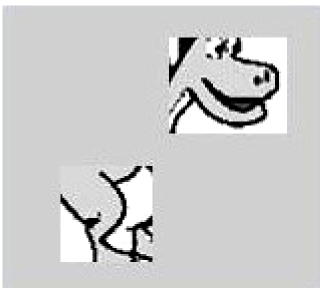


Fig.3 (a). An example of spatial filtering

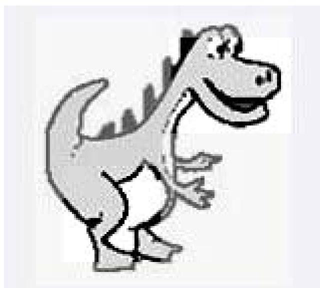


Fig.3 (b). An example of spotlight model of attention

There are two main models of spotlight attention. The simplest model can be looked upon as a spatial filter, where what falls outside the attentional spotlight is assumed not to be processed. In the second model, the spotlight serves to concentrate attentional resources to a particular region in space, thus enhancing processing at that location and almost eliminating processing of the unattended regions. The main difference between these models is that in the first one the spotlight only passively blocks the irrelevant information, while in the second model it actively directs the "processing efforts" to the chosen region.

Fig. 3(a) and Fig 3(b) visually clarify the difference between the spatial filtering and spotlight attention.

A conventional view of the spotlight model assumes that only a single region of interest is processed at a certain time point and supposes smooth movement to other regions of interest. Later versions of the spotlight model assume that the attentional spotlight can be divided between several regions in space. In addition, the latter support the theory that the spotlight moves discretely from one region to the other.

3.2 Object-based Model

As reviewed above, the spotlight metaphor is useful for understanding how attention is deployed across space. However, this metaphor has serious limitations. A detailed analysis of spotlight model drawbacks can be found in [3]. Object-based attention model suit to more practical experiments in physiology and is based on the assumption that attention is referred to discrete objects in the visual field. However being more practical, in contrast to the spotlight model, where one would predict that two nearby or overlapping objects are attended as a single object, in the object-based model this divided attention between objects results in less efficient processing than attending to a single object. It should be noted that spotlight and object-based attention theories are not contradictory but rather complementary. Nevertheless, in many cases the object-based theory explains many phenomena better than the spotlight model does.

The object-based model is more complicated for implementation, since it requires objects' recognition, while the spotlight model only requires identifying the regions of interest, where the attentional resources will be concentrated for further processing.

4. The Proposed Architecture for the CMOS Tracking Imager

In this Section two possible architectures of tracking CMOS imagers are discussed. While these architectures seem similar, the first one employing the spatial filtering model and the second one employing the spotlight attention model. The operation of both imagers will be described with the reference to an example of input scene in FOV, as shown in Fig. 4. In this scene three different types of regions can be observed: (a) two regions consisting of large salient targets (stars), (b) a number of regions consisting of small salient stars and (c) regions that don't consist of any targets of interest. The observer is usually interested in tracking the targets mentioned in group (a), but sometimes there is interest in targets both from groups (a) and (b). Moreover, sometimes the observer is interested in tracking the targets from group (b) and the targets from group (a) serve as salient distractors. Note, the term real time tracking relates to the ability to calculate the center of mass (COM) coordinates of the tracked target in real time.

Fig.4 An example of input scene in FOV

4.1 The Spatial Filtering Based Architecture

Fig. 5 shows the architecture of the CMOS tracking image system based on the spatial filtering model. The proposed sensor has two modes of operation: the fully autonomous mode and the semi-autonomous. In the fully autonomous mode all functions required for target tracking are performed by the sensor (at the pixel level and by on-chip image processing). The only data transmitted to the processor/DSP in this case is the tracked targets coordinates. This mode is very efficient by means of bottleneck problems solution; however it allows less flexibility and influence on the tracking process by the user. In the semi-autonomous mode, part of the functionality is performed in the circuit level and the chip level and part of processing is done by the processor/DSP. In this case feedback from the processor/DSP to the sensor exists and more flexibility is achieved; however more data flows from the sensors to the processor/DSP and back and more processor/DSP resources are used to complete the real time tracking (see Fig. 5).

The real time targets tracking is accomplished in two stages of operation: target acquisition and target tracking. In the acquisition mode N most salient targets of interest (the number of targets N can be predefined by the systems or can be user-defined) in the FOV are found. Then, N windows of interest with programmable size around the targets are defined, using the control logic block. These windows define the active regions, where the subsequent processing will occur, similar to the flexible spotlight size in the biological systems. In the tracking stage, the system sequentially attends only to the previously chosen regions, while completely inhibiting the dataflow from the other regions. This way the system based on the spatial model of attention allows distractors elimination, oppositely to a case of the spotlight model. According to the spotlight model appearance of the additional "salient" targets during the tracking of given targets of interest causes temporary or even permanent loss of the desired target.

Thanks to the control logic and to the CMOS imager flexibility, the proposed concept permits choosing the attended regions in the desired order, independent on the targets saliency. In addition it allows shifting the attention from one active region to the other, independent of the distance between the targets.

As can be seen more information is transmitted from the sensor array to the on-chip system-processing block during the acquisition mode than during the tracking mode. The reason for this is that during the acquisition mode the whole image is captured and transmitted to the on-chip image-processing block for further processing. In case of the fully autonomous mode of operation, the on-chip processing finds the center of mass coordinates of all targets of interest and windows of interest are defined by the control logic. In a case of semi-autonomous mode, some processing is performed in the on-chip processing block and the data is transmitted to the processor/DSP for further processing and windows of interest definition. Note, the acquisition mode is required only once at the beginning of tracking. During the tracking mode, only information from the chosen windows of interest is transmitted, dramatically reducing the bottleneck problem between the sensor and on-chip image-processing block and between the imager chip to processor/DSP.

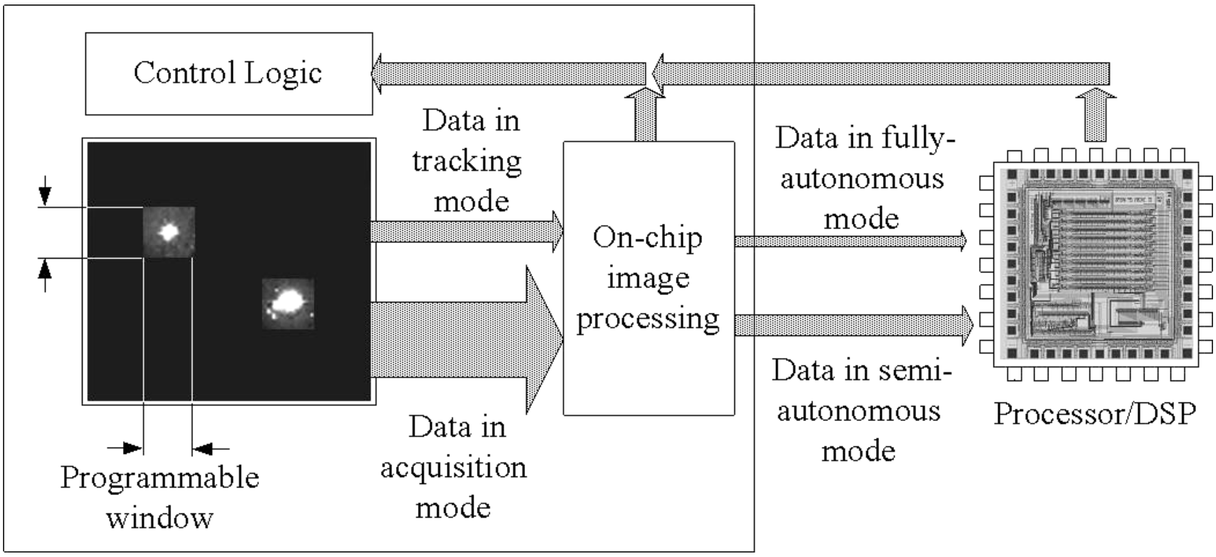


Fig.5 Architecture of the CMOS tracking image system based on the spatial filtering model

4.2 The Spotlight Based Architecture

Fig. 6 shows the architecture of the CMOS tracking image system based on the spotlight model. The principle of this system is very similar to the concept, presented in sub-section 4.1.

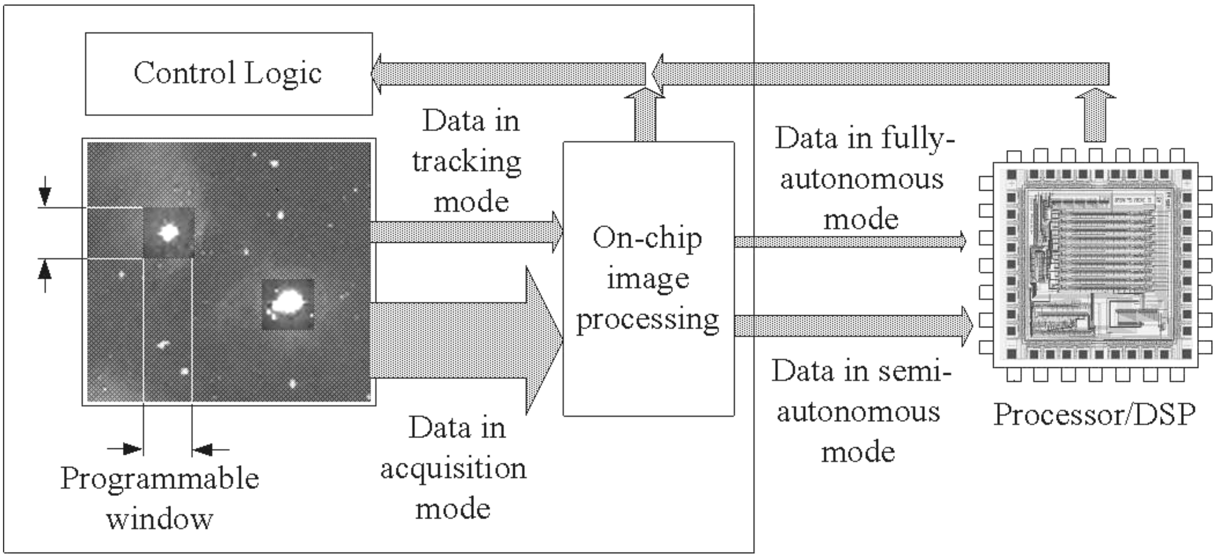


Fig.6 Architecture of the CMOS tracking image system based on the spotlight model

However, there is one important difference. While the spatial filter based imager filters all information that does not fall into the windows of interest, the spotlight based system, only reduces the amount of information transmitted from these regions for further processing. This is performed by a unique feature of CMOS image sensors that allow implementation of adaptive multiple resolution sensors [15]. As can be seen in Fig. 5, the two most important regions of interest are captured with full resolution (in the same way like in the system presented on sub-section 4.1), while all other regions are captured with reduced resolution. On one hand this allows to change the defined windows of interest according to the events in the FOV. On the other hand, this architecture is still significantly reduces the amount of information transmission from the sensor.

5. Conclusions

Two architectures of tracking image systems were proposed. Both the data transmission bottleneck and the performance bottleneck are reduced in the proposed imagers due to employing the spatial filtering and spotlight models of attention, found in biological systems. The proposed imagers can be easily implemented in a standard CMOS technology. Both imagers can operate in the full autonomous and semi-autonomous modes of operation. A brief description of the spatial and object-based models of attention was presented and an explanation of the proposed image systems operation was provided. Further research includes implementation of the proposed sensors in an advanced CMOS technology.

Acknowledgements

We would like to thank Alexander Spivakovsky and Evgeny Artyomov for their helpful suggestions during the preparation of this work. We also would like to thank the Israeli Ministry of Science and Technology for funding this project.

Bibliography

[1] O. Yadid-Pecht, R. Etienne-Cummings. CMOS imagers: from phototransduction to image processing. Kluwer Academic Publishers, 2004.

[2] R. William, H. Engineering. Using FPGAs for DSP Image Processing. FPGA journal, available http://www.fpgajournal.com/articles/imaging_hunt.htm

[3] Chun, M. M., & Wolfe, J. M. Visual Attention. In E. B. Goldstein (Ed.), Blackwell's Handbook of Perception, Vol. Ch 9, pp. 272-310). Oxford, UK: Blackwell, 2001.

[4] R. W. Remington, L. Pierce. Moving attention: Evidence for Time-invariant shifts of visual selection attention. Perception and Psychophysics, vol. 35, pp. 393-399, 1984.

[5] J. Moran, R. Desimone. Selective attention gates visual processing in the Extrastriate Cortex. Science, vol. 229, pp. 784-787, 1985.

[6] G. Sperling, E. Weichselgartner. Episodic theory of the dynamics of spatial attention. Psychological review, vol. 102, no. 3, pp. 503-532, 1995.

[7] A. Treisman. Features and targets: The fourteenth Barlett memorial lecture. Quarterly Journal of Experimental Psychology, 40A, pp. 201-237, 1988.

[8] A. Treisman. Feature binding, attention and target perception. Philosophical Transactions of the Royal Society of London, vol. 353, pp. 1295-1306, 1998.

[9] A. Treisman, G. Gelade. A feature-integration theory of attention. Cognitive Psychology, vol. 12, pp. 97-136, 1980.

[10] A. Treisman, D. Kahneman, J. Burkell. Perceptual targets and the cost of filtering. Perception & Psychophysics, vol. 33, pp. 527-532, 1983.

[11] S. Yantis. Targets, attention, and perceptual experience in “Visual Attention". R. D. Wright (Eds.), pp. 187-214. Oxford, NY: Oxford University Press, 1998.

[12] M. I. Posner. Orienting of attention. Quarterly Journal of Experimental Psychology, vol. 32,pp. 3-25, 1980.

[13] C. Koch, B. Mathur. Neuromorphic vision chips. IEEE Spectrum, vol. 33, pp. 38–46, May 1996

[14] C. Koch, S. Ullman. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. Human Neurobiology, vol. 4, pp. 219–227, 1985.

[15] E. Artyomov, O. Yadid-Pecht. Adaptive Multiple Resolution CMOS Active Pixel Sensor. ISCAS 2004, Vancouver, Canada, May 2004, Vol. IV, pp. 836-839.

Authors' Information

Alexander Fish – The VLSI Systems Center, Ben-Gurion University, Beer Sheva, Israel; e-mail: afish@ee.bgu.ac.il

Orly Yadid-Pecht – Dept of Electrical and Computer Engineering, University of Calgary, Alberta, Canada; e-mail: oyp@ee.bgu.ac.il

EXAMINATION OF PASSWORDS' INFLUATION ON THE COMPRESSING PROCESS OF NON-ENCRYPTED OBJECTS

Dimitrina Polimirova-Nickolova, Eugene Nickolov

Abstract: The principal methods of compression and different types of non-encrypted objects are described. An analysis is made of the results obtained from examinations of the speed of compression for objects when using passwords with different length. The size of the new file obtained after compression is also analyzed. Some evaluations are made with regard to the methods and the objects used in the examinations. In conclusion some deductions are drawn as well as recommendations for future work.

Keywords: Password, Methods of Compression, Non-Encrypted Applications, Archive Programs, Level of Compression, File Extensions, Information Security.

ACM Classification Keywords: D.4.6 Security and Protection: information flow controls

The Situation

The information technologies' progress leads to increasing need of creation and use of compressed objects. With regards to this, examination and analysis are made of different methods of compression and their varieties, which purpose is the creation of high-speed, effective compression of information flows working in real time.

The compression represents a method for reducing the size of some object. In the case of digital data, the compression is connected to the reduction of the bytes'number by removing the unneeded and/or non-critical information which results in decrease of object's size. The methods of compression are used when the user wishes to economize space when storing the object, to gain time, when it is emailed or to reduce the risk for loss or modification of the information.

When using the different methods of compression a password could be put which allows the achievement of highest information security of the object. The password represents confidential authentication information, formed by symbol strings.

In this article are examined and analyzed non-encrypted applications, on which different methods of compression and passwords with different size are applied. We shall use the term "non-encrypted application" for all applications which are executed under the control of the operation system, process some input data and produce relevant output results while the information flows are in normal (non-encrypted) form all the time. In this examination the non-encrypted applications are represented by 18 types of extensions belonging to 6 basic types archiving programs. They are chosen among the 300 archiving programs known at the moment. They are:

1) E-mail archiving programs – this type of archiving programs use the relative homogeneity of information flow (e-mail traffic) to select the most appropriate methods for compression.

2) Converting archiving programs – these archiving programs are able to convert objects compressed by some method in objects compressed by another method.

3) Multiple archiving programs – these programs execute some successive archiving processing on the different parts of some object using methods of compression with different properties.

4) Image archiving programs – these programs help solving a very important problem of the present day connected to the real-time processing of video and image web-objects – the obligatory immediate compression of the object after its creation. The transmission and the processing of the object are executed entirely in compressed state, till the end moment of its reproduction by the appropriate media.

5) Data archiving programs – these programs are specialized in the processing and the use of compressed objects, got from information flows owning the characteristics of "data" (in this case we are concerned by the circumstance that the data in the different phases of their existence pass in compressed mode, are kept for some time in this "minimized" form, after which thy are decompressed.).

6) Executable archiving programs – the goal in these programs is to achieve some specificity of the compression, connected to the possibilities for running the compressed objects.

The Problem

The object of examination are the methods of compression (different types of compressing programs) and their varieties (different levels of compression, size of the used password etc.); the stress will be laid on the password length and its effect on the compression process [2, 3]. The methods will be analysed after their application on non-encrypted objects from the 6 types of archiving programs specified above. By contrast with the compressin, the archiving is a process of storing a data structure for a later use. In most cases this data structure (which is usually a single object) is stored in a file, but besides this it could be written in memory or transmitted to another application.

Some kinds of examinations are made, concerning the effect of the password size on certain characteristics of the compression of non-encrypted applications.

The first examination analyses the SPEED of the compression process. The following tasks are put in connection to this examination:

1) Determination of the *objects* which will be analyzed and the *methods of compression*, applied on them. The objects are the selected 18 types, mentioned above and described in Table (1a – 1f), by which a formal presentation of non-encrypted applications is obtained. The initial size of each type is 1 024 000 bytes.

Table 1a
E-mail archiving programs

Extension	Program / Information
DBX	Outlook Express Email Folder
IDX	Outlook Express Mailbox Index
PCE	Eudora Mailbox Name Map

Table 1b
Converting archiving programs

Extension	Program / Information
ACE	WinAce Compressed File
RAR	WinRAR Compressed Archive
ZIP	Compressed Archive File

Table 1c
Multiple

Extension	Program / Information
ARJ	ARJ Compressed Archive
JAR	JAR Archive
TAR	Tape Archive File

Table 1d
Image archiving programs

Extension	Program / Information
AIS	ACDSee Image Sequence File
B&W	Image Lab
BIL	AreView Image File (ESRI)

Table 1e
Data archiving programs

Extension	Program / Information
DOC	Word Document (Microsoft)
PDF	Acrobat Portable Document Format
TXT	Text File

Table 1f
Executable archiving programs

Extension	Program / Information
EXE	Executable File (Microsoft)
PE	Portable Executable File
PL	Linux Shell Executable Binary

2) Evaluation of the *time*, needed for compression without password by the means of 5 compressing programs, selected among the dozens, known up to now– ACE, GZip, JAR, RAR, ZIP. With the purpose of simplifiing the exposition we shall desribe only the examinations for level of compression Normal (Figure 1).

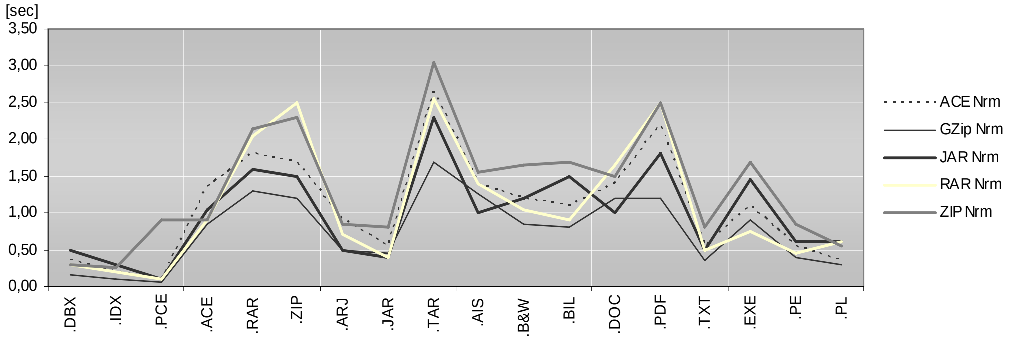


Figure 1

3) Compression of the initial objects with *8, 16 and 32-digit password* and measurement (by software means) of the time, needed for their processing. Fixed passwords are used with length of 8, 16 and 32 symbols, which for simplicity are formed only by the numbers from 0 to 7, ranged consecutively in ascending order, and repeated periodically (Figure 2a – 2c).

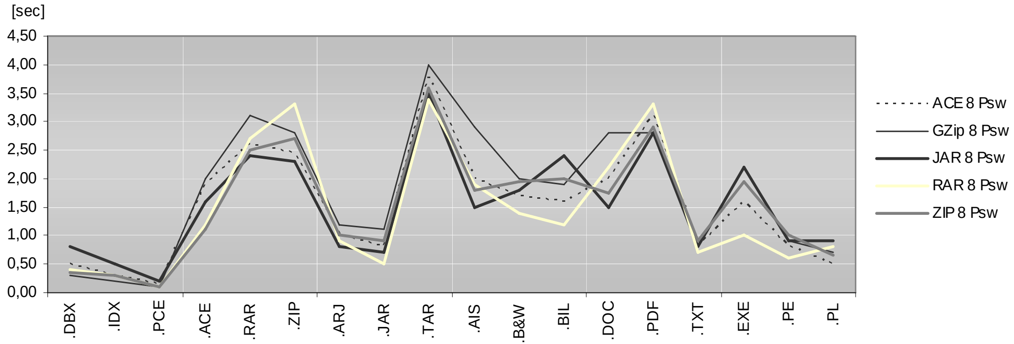


Figure 2a

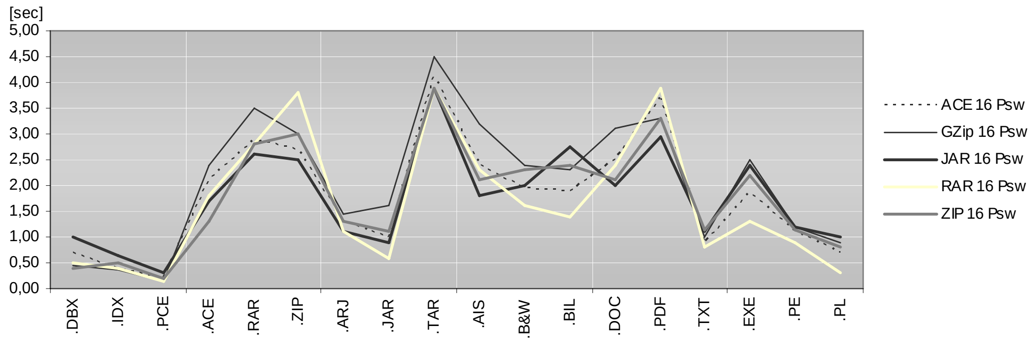


Figure 2b

Figure 2c

4) Juxtaposing of the time, needed for the compression of objects when *8, 16 and 32-digit password* is used (Figure 3).

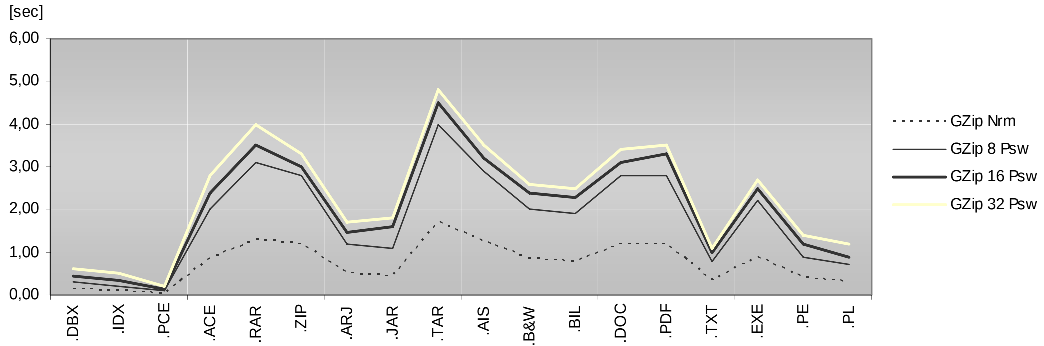


Figure 3

The second examination analyses the SIZE of the resulting file after compression by using passwords with different length. The following tasks are put in connection to this examination:

1) To compress different file formats with different compressing programs with level of compression Normal *without using passwords* (Figure 4a, b, c, d, e, f).

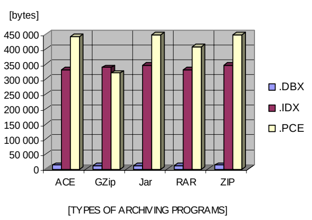


Figure 4a

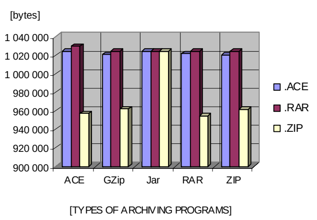


Figure 4b

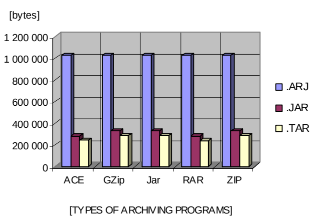


Figure 4c

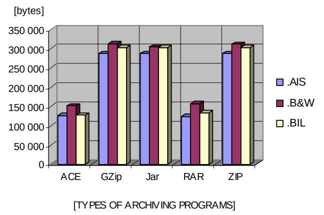


Figure 4d

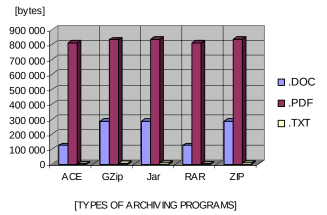


Figure 4e

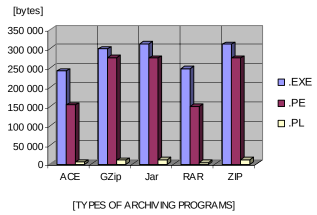


Figure 4f

2) To compress different file formats with different compressing programs using *8, 16 and 32-digit passwords* (Figure 5).

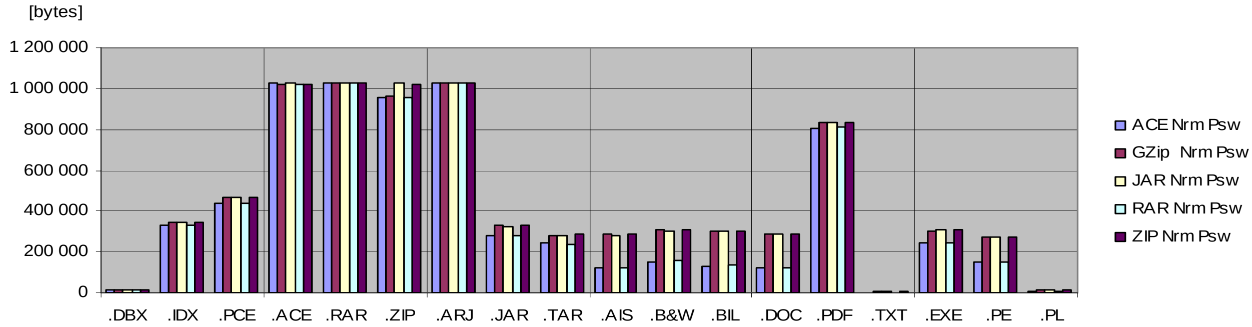


Figure 5

3) To juxtapose the size of the files, compressed *without password* with the size of the files compressed with *8, 16 and 32-digit passwords*.

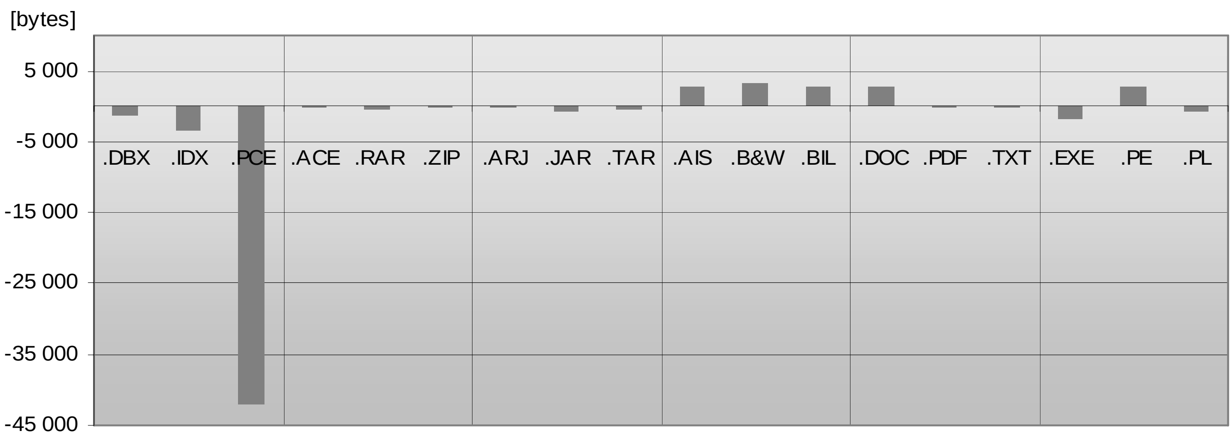


Figure 6

With the purpose of simplifiing the exposition, on Figure 6 are demonstated the results only for one method of compression (GZip) with regards to the different size of files *with* and *without* password.

The following assessments could be made from the experiments, which were carried out:

1) With regard to the SPEED:

a) The evaluation with regard to the selected objects for compression is positive and the assumptions that are made have not influence on the obtained results. The evaluation with regard to the selected methods of compression is also positive and the experiments that are carried out could be generalized for the other methods. The selected size of the objects is sufficient for conducting the necessary evaluations, conclusions and recommendations.

b) The evaluation with regard to the time, needed for compression *without password* shows a maximum for .TAR and minimum for .PCE. The examinations made for a level of compression Normal (Nrm) illustrate enough the main evaluations, conclusions and recommendations.

c) The evaluation with regard to the compression with *8, 16 and 32-digit passwords* shows that the number and the types of the symbols are correctly chosen. The maximum with regard to the speed is for .TAR, the minimum – for .PCE; this evaluation is valid for the three examined password lengths.

d) The evaluation with regard to the time, needed for compression with *8, 16 and 32-digit passwords,* shows a confirmation of the maximum for .TAR and of the minimum for .PCE for the chosen most appropriate method of compression GZip, along with some local extremums which could be seen in the graph on Figure 3. The relative increase of the time for compression with password in comparison with the compression without password is largest in the case of 8-digit password. The relative increase of the time for compression with 16 and 32-digit password is minimal and practically the same for all other extensions.

2) With regard to the SIZE:

a) The evaluation with regard to the compression of different file formats with the selected types of compressing programs for level of compression Normal *without using a password* shows the presence of minimums, maximums and other local extremums, illustrated on Figure 4a – 4f. The absolute maximum is for .TXT compressed with ACE, and the absolute minimum - for .ARJ compressed with JAR.

b) The evaluation with regard to the compression of different file formats with the selected types of compressing programs for level of compression Normal *with a password* shows the presence of minimums, maximums and other local extremums, illustrated on Figure 5. The use of 8, 16 and 32-digit passwords practically does not influence the size of the resulting objects.

c) The evaluation with regard to the juxtaposition of the sizes of files compressed *without password* and *with password* for a selected method of compression (GZip) shows the largest difference for .PCE and the smallest for .ACE.

Conclusions and Future Work

The examinations that are made and the extensive experiments that are carried out, show the significant perspectives in the scientific research on these problems. The influence of the methods of compression on the different applications together with the use of password with different length is considerable and could be used for decision-making connected to the information security and the risk evaluation in processing of information flows [1].

Bibliography

[1] Bahram Javidi, Optical and Digital Techniques for Information Security (Advanced Sciences and Technologies for Security Applications), Springer, June, 2005.

[2] David Salomon, Data Compression: The Complete Reference, Springer Verlag New York, Inc., 2004.

[3] Rob Shimonski, Introduction to Password Cracking, IBM Developer Works Hacking Techniques article, July 2002.

Authors' Information

Dimitrina Polimirova-Nickolova - PhD Student, Research Associate, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, E-mail: polimira@nlcv.bas.bg.

Eugene Nickolov - Prof.DSc, PhD, Eng, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, E-mail: eugene@nlcv.bas.bg.

CONCEPTUAL IDEA OF IDENTIFICATION OF PATTERNS AND PROBLEM SOLVING IN THE HUMAN'S MEMORY AND THE POSSIBILITIES TO USE IT IN ARTIFICIAL INTELLIGENT

Zinoviy Rabynovych

Abstract: Given cybernetic idea is formed on the basis of neurophysiologic, neuropsychological, neurocybernetic data and verisimilar hypotheses, which fill gaps of formers, of the author as well. First of all attention is focused on general principles of a Memory organization in the brain and processes which take part in it that realize such psychical functions as perception and identification of input information about patterns and a problem solving, which is specified by the input and output conditions, as well. Realization of the second function, essentially cogitative, is discussed in the aspects of figurative and lingual thinking on the levels of intuition and understanding. The reasons of advisability and principles of bionic approach to creation of appropriate tools of artificial intelligent are proposed.

Keywords: pattern, perception, identification, solving, problem generator, understanding, intuition.

ACM Classification Keywords: I.2.0. Artificial intelligence: general – cognitive simulation; I.2.4. Artificial intelligence: knowledge representation formalisms and methods – semantic networks; I.2.6. Artificial intelligence: learning; I.5.1. Pattern recognition: models – neural nets; J.3. Life and medical sciences – biology and genetics.

Introduction

The conceptual level of Modeling natural mechanisms of psyche means penetrating into it from the top to the bottom – from certain psychical functions to information principles of its physical realization, i.e. from the psychical result to the information mechanism of its obtaining [1-4].

Thus, the Conceptual Model determines (using top-theme) "the understanding of the relationships between structure and function in biology".

Information processes taking place in the physical substance of nervous system consist of two main classes: purely combinational processes (as unrelated to memorizing) which occur in sensory organs, organs moving control etc., and combinational- accumulating processes which occur in the Memory.

One should consider that thinking processes as a starting postulate belong to these processes (including processing of both the input and output information coming into and out of the Memory, respectively); i.e. the Memory is also a thinking medium encircled into the total neural net of the whole nervous system.

Where is the limit of the Memory and how is it organized?

We shall consider below the Conceptual Model of the Memory and processes which occur in it, and direct our attention both to the recognition function relating to the Memory concept as itself and the function of solving problems already relating to purposeful thinking as the most important one for metabolism. These functions are dominating for artificial intelligence problems and the Conceptual Model (CM) will, therefore, be of interest within the subject "From Nature to Artificial".

Memory and Recognition of Patterns

The Memory functions include such concepts as "recognizing", "remembering", "imagining" etc. Unlike perceiving the information from environment, it is suitable to represent all these actions as demonstrating the so-called "mental sight", i.e. a look which is initialized within the Memory and acted as exciting the certain sensory memorized structures in the Memory network, their combination etc. So, what is this structure?

To answer this question and understand the global defining principle of the Memory organization (that is required for the Conceptual Model), it is necessary to start with the basic hypothetical prerequisite which ensues, as it were, from "common sense" (but doesn't coordinate with experimental data which it doesn't follow since it is impossible to observe sufficiently in general and in details).

And such a prerequisite as the main basic hypothesis is the following one:

"The pattern reconstruction in the Memory (imagination, mental sight) is determined by exciting all its elementary components which took part in the pattern perception".

It would be considered as a law of nature, as an immutable but inexplicable fact. Really, how does it happen that fluctuations of the potential of the neural net components are transformed to, visible inside (also audible, tangible, palpable) patterns? But it happens, isn't it? So, this fact should be the basis for the further constructions.

These constructions should already result in capability to form the signals inside the Memory. These signals are able to excite the components of the pattern perceived and fixed in the Memory, and there is already no information about this pattern at the Memory input. Concerning this initial information, the internal exciting signals are the information flow, which is reverse with regard to the information formed in the Memory.

Fig. 1 illustrates the aforementioned fact: Memory is represented with its main fields (see below) and the arrows denote delimitation between the direct and inverse information. So, we conclude that the "Memory" begins, in fact, where the reverse bonds finish. They finish just at the level of "c" conceptors, which repeat "r" receptors. Reverse bonds shouldn't extend over "r" receptors since the illusions occur during the mental sight, i.e. one sees, hears and touches that it isn't sensed at present time. So, the Memory limit is just positioned between "r" receptors and "c" conceptors; therefore, duplicating the first ones by the last ones is necessary for it.

The latter ones, "c" conceptors, are actually the smallest sensory components of the perceived patterns and the most elementary sub-patterns regarding to the Pattern. These sub-patterns hierarchically and naturally group together into larger sub-patterns, then the latter ones group in even larger sub-patterns, etc., up to the sensory concentration of the whole Pattern in the single structure unit which denotes only its symbol.

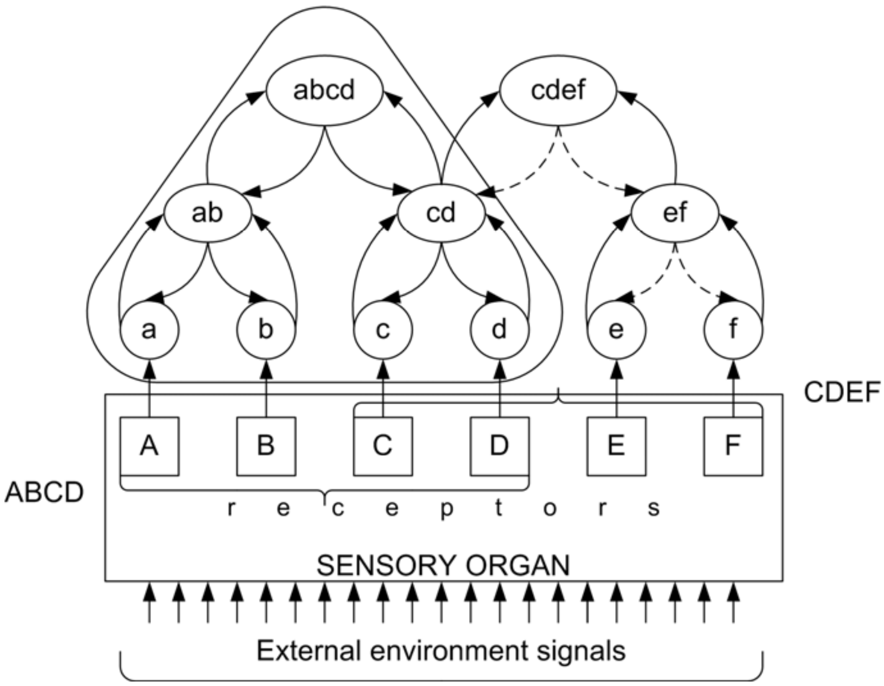


Fig. 1. Elementary structures of pattern perception, memorizing and recognition

Thus, when constructed information signals enter the Memory from outside, the nonlinear pyramidal Model of the pattern expressed by this information will be constructed in the Memory. The down-going reverse bonds from the top of the pattern pyramid up to its base (consisting of conceptors which repeat input receptors) are necessary for memorizing the pattern, i.e. for making possible the pattern reconstruction by "mental sight" according to the aforementioned main hypothesis. So, the Model of the concrete pattern, both as an object and as a structure in the Memory, is the pyramid hierarchical construction with up-going convergent inductive (from partial to general) bonds and down-going divergent deductive (from general to partial) bonds.

The Memory as a whole consists of the set of such Models as the Memory loops associatively bound by the general components. The Memory is a structural realization of the semantic network as a system with knowledge fixed in it.

The aforementioned fact can be illustrated with the network, which is enough simple for visualization (Fig. 1). The network consists of fully constructed (i.e. fixed in the Memory) ABCD pattern and CDEF pattern entered into the Memory but not fixed in it.

It is caused by the fact that the pyramid (i.e. the Model) of the ABCD pattern already has reverse bonds and pyramid of the CDEF pattern isn't still constructed in full.

These patterns have one general CD sub-pattern which is the associative element of both pyramids. Such using of the general parts of the pattern provides: first, economy of constructing the semantic structures in the Memory; second, spontaneous (i.e. self-originated) parallelism during processing the pattern information in it that is essentially favorable for effectiveness of this processing (see below). Finishing the construction of the CDEF pattern means creating the reverse bonds in it (shown via dotted line).

Finishing the construction, i.e. changing the Model of demonstrated pattern into the memorized one, can be performed by several successive demonstrations of the same pattern that results in beating the genetically innate reverse bonds or even in their appearance.

The creation of the pattern structures in the Memory (Models) is scientifically valid. The aforementioned conception on constructing these structures, nevertheless, requires the following plausible hypothesis: constructing the direct up-going bonds for perceiving the concrete patterns by the Memory precedes an appearance of the reverse down-going bonds which ensure their memorization (see the main hypothesis) by learning.

Thus, a brain Memory medium of a new-born child (or other individuals with enough developed nervous system) is already saturated by bonds needed to obtain the information from organs of sense. The Memory itself is originated in full by constructing the already reverse bonds in metabolic process begun from the birthday (and while preparing the relevant capabilities even before it). Therefore, the brain as the thinking mechanism creates itself but under the environment impact. Let rather look aside and notice that so-called innate abilities are apparently caused by peculiarities of the genetically innate direct bonds. The demonstration of these abilities (as the construction of the relevant full pattern structures in the Memory) already occurs as a result of learning; i.e. phenomenal talents appear as a result of coincidence of two factors: the genetically constructed relevant structures in the Memory Medium and its subsequent learning as an impact of external factors.

As a consequence, the brain's Memory in its conceptual representation is a hierarchical semantic network restricted from above with ending of ascending connections (straight lines coming from sensory organs) and from bottom - with ending of backward connections to structure units, which directly perceive information entering the Memory (formed with specific sensory organs from specific signals). And whole thinking from simple recognition of pattern information to its analysis and synthesis and subsequent actions on sets of patterns, connected with creation and transformation of different situations, are performed in the closed restricted space, connected with dual-side informational links with external (toward it) environment. The first link is intended for information perceiving from the environment, the second – for information (directing, informative and other, created directly in Memory) production into the environment.

Now we consider recognition as the simplest psychical function, which determines the presence of the Model of the recognized pattern in the Memory as its definite structure.

If such structure exists (e.g. abcd Model described in Fig. 1) then the secondary spike of exciting this structure already passed by the reverse bonds will take place after presenting the prototype of the structure and its switching-off,. It means implementing the "mental sight", i.e. recognizing the presented pattern in accordance with the main hypothesis. If the fully reverse bonds are absent, i.e. the pattern isn't filled in (such as cdef in Fig. 1) the secondary spike will not take place.

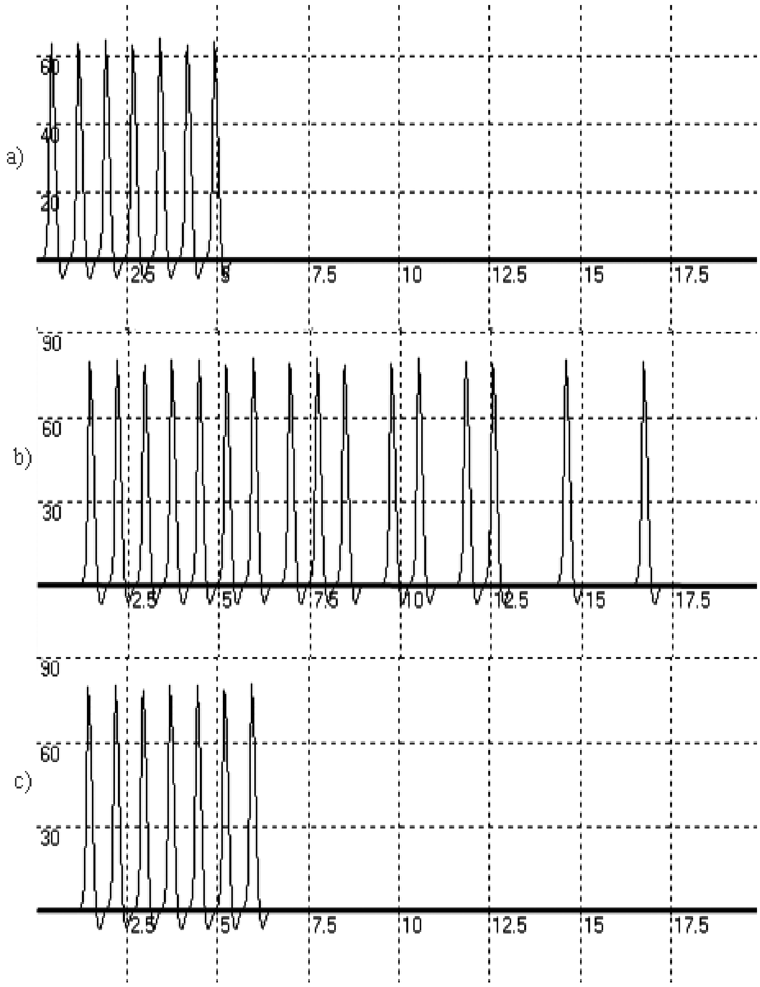


Fig. 2. Model's work result of pattern recognition process:

a) Inputs to Memory onto elementary conceptors (outputs of receptors)

b) Lower lever elementary conceptors' output with presence of backward links

c) Lower lever elementary conceptors' output with presence of broken backward links

Thus, it is hypothetically claimed that recognition of the demonstrated pattern is determined with the secondary spike of exciting the relevant structure as its remembrance.

This hypothesis in supported by Modeling of the components of the olfactory sensory system (performed by neurophysiologist G.S. Voronkov (Moscow State University, Russia). Results of computational experiment represented on Fig. 2 were obtained by PhD students (Tkachuk. S, Iamborak R., Iamborak V.), with assistance of biological network Modeling tools developed by them (the secondary spike doesn't occur if the existing reverse bonds in the olfactory sensory system are interrupted and it obviously occurs if the reverse bonds exist, i.e. there is the full Model of the demonstrated pattern in the Memory).

Since exciting processes can be both probabilistic and possibilistic, the results are not defined in full and misoperations (e.g. due to associative bonds (cdef) between the Models of two patterns) are certainly possible in a recognizing process. But this indeed takes place in fact.

"Recognizing" in the presented above interpretation is the function implemented quite automatically in any living organism, which has the relevant nervous system. The implementation of this function in the broad sense of this term already including "comprehension" should relate to the thinking process in which described actions are implemented only as its first stage. During the construction of mathematical Models of Memory patterns the apparatus of growing pyramidal networks [5] can be effectively used, as a type of semantic networks.

Memory and Solving of Problems

Introduced conception of pattern structures in the Memory as their Models is general for all thinking functions since it is a basis for organizing all the Memory.

It is very important for Modeling the human thinking processes to consider the Memory as a whole system consisting of two subsystems: sensory one, lingual one [2] and higher associative subsystem, which store the patterns and their lingual notations and notions. The structures of these subsystems are interrelated by direct and reverse bonds, which define a correspondence between them and their mutual effect by transfer of the excitation. Note, that language variety is realized by the additional lingual subsystems, whose structures wouldn't bind with the sensory system structures but interact with it only by the structure of subsystem of single language. The type of bonds between sensory and one or another lingual subsystems defines capability and the level of recognized thinking in one or another language.

The process of human thinking, as specified above, is determined by interaction between the sensory and lingual subsystems of the Medium at the cognitive and intuitive (as uncognitive) levels.

The cognitive thinking is organically bound with the lingual expression of thoughts, i.e. individual talks, as it were, with himself or herself. Thus, the cognitive thinking is called verbal although the language wouldn't be a speech. For example, the relevant signals even enter the organs of speech in the first dominating case.

The principally sequential character of cognitive thinking results from it, because it is impossible to pronounce more than one thought at the same time.

In general, the cognitive thoughts are represented by the so-called "complete" dynamic structures, which integrate the relevant excited structures of the sensory and lingual subsystems at different levels of their hierarchy.

Excitement of incomplete structures is related to uncognitive thinking, which isn't restricted by strong interaction of structures of the sensory and lingual subsystems.

Therefore, such dynamic structures can arise at once at different levels of these subsystems not resulted in "pronouncing" the excited "sense".

So, the amount of information processed (even spontaneously) can be much greater here than during cognitive thinking.

Furthermore, such combinations can arise at the intuitive thinking level, which do not have lingual equivalents and so they don't rise to the consciousness level (e.g. savage thinking).

Therefore, the component of uncognitive intuitive thinking is of great importance besides the cognitive component of the purposeful process of solving the problems (man does think!).

The following hypothesis will be quite natural in accordance with aforementioned facts [1].

The Problem to be solved is specified in the Memory by the Models of original and goal situations, and its solution is an activated chain of cause-effect bonds that results in transforming the first one to the second one. The process of constructing the chain consists of two interrelated processes operating at the same time: sequential cognitive process (as reasoning) and spontaneous activation of the structures in the Memory by their associative bonds with the Models of the original and goal situations. We shall ignore below the "Model" term and terminologically equate the structure in the Memory to the situation itself.

Realizing the problem to be solved (the Problem situation) creates, some "tension" in the Memory, and so the special "problem generator" (PG) [1] term is very useful for obvious illustrating and considering the process specified in the hypothesis, which poles are the original and goal situations and its "tension" maintains the existence of the problem situation.

Creating the activated chain, which locks these poles and means solving the Problem, liquidates this "tension", i.e. terminates the PG existence. The links of the specified chain are intermediate situations between the original and goal situations (Fig. 3) and they can be defined not only by single-sided transformation but also by cross-transformation of these situations. However, the locking chain wouldn't be created in the continuous process (e.g. if there isn't enough knowledge in the Memory) that promotes creating the new intermediate PG which defines the break in the constructed chain of its poles locking, i.e. a new pair of original and goal situations.

The Problem solution can result in creating the new structure (i.e. the new knowledge) in the Memory due to beating the new bonds between its components. If there are enough intension and lifetime of the dynamic structure, this process is similar to conversion of the information dynamically stored (i.e. as a short-term memorizing in the statistically fixed Memory).

As the "distance" between the original and goal situations decreases due to creating the links of the required chain, which locks them, increase of the activity of the second process and the process as a whole can result in avalanche sudden locking of the PG poles, i.e. solving the Problem as a result of enlightenment. Furthermore, it can occur quite unexpectedly and accidentally just as a result of the second process, only if the first process which is the cognitive reasoning doesn't exist. The second process occurs nevertheless because PG is already excited.

In general, enlightenment is a property of creative processes, which can be schematically considered as a sequence with step-by-step domination of cognitive and intuitive thinking [6].

In the first case, the obtained result is comprehended and the new intermediate chain (sub-chain) is proposed. In the second case (i.e. at the next stage), this sub-chain is already reached and changed, if possible, up to obtaining the final result. Thus, the total process of solving the problem is probabilistic (or possibilistic) with a wide range of its quantitative characteristics. As such its speed and the time depends on the complexity of the problem being solved and the level of excitement of *PG.* The first factor is stipulated with length of the chain of interconnected structures, which communicate the initial and target situations (i.e. 'the distance' between them). The second factor is stipulated mainly with how strongly a human is dedicated to the problem in concern, i.e. how the emotion of interest is involved into its solution.

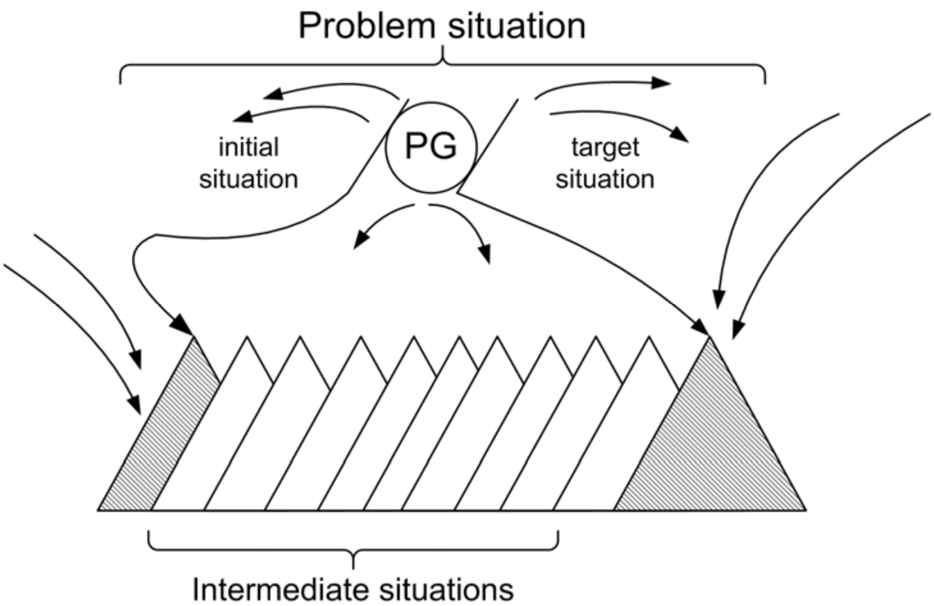


Fig. 3. Chain of the problem solving as a pattern situations transmission

Memory and Emotions

As long as the procedure of information processing in memory (in general in neural system) is characterized not only with sense but also with emotional part, let us according to the subject of the report define the major features of interaction between these contributors.

Division of a whole neuron network into separate fields and preparation of informational signal-receptors and memory environment, which accepts these signals, hold also in case, when emotional components are presented. Such components perform emotional influence by creation and transmission of specialized mediators.

These influences, which are established and transmitted in the whole hierarchy of memory, define its particular emotional state (e.g. depression or alternatively inspiration). In turn this state represents a background for mental processes taking place inside the memory. In other words, it guides the activation and interaction of memory structures.

Consequently, the emotional part actively participates in mental process and influences its functional performance. Nevertheless such performance, embodied with corresponding neuro-physiological indicators, naturally effects the emotional condition. For instance, when mental function is involved for solution of a complex problem, as it has been already stated, the emotion of interest (or even inspiration) plays the major role. Moreover, when theoretical problem is solved (the poles of PG are connected) with sudden conjecture, the emotion of extreme joy can appear.

As such, information processing inside the memory can be defined as emotional and sense simultaneously. And this processing in its conceptual representation can be easily represented in terms of described principles of memory organization.

Naturally, major emotional organs are located on the memory level, which accepts 1:1 receptor signals. It means, that during perceiving a pattern by memory from outside they will be organically included into its formed structure.

During the pattern recognition according to the main hypothesis all components of memory will be excited, i.e. the mental pattern obtains the “emotional colouring”.

Consequently, memory as a system with structured organization during the information processing creates two independent results in thinking – sense and emotional (correspondingly – idea and humour, which stipulates the state of memory).

Emotional organs are located in neural system and outside memory. They participate in interaction with memory similarly to guided fulfilling mechanisms (like, moving apparatus).

I.e. emotional mediators are presented not only in internal loops of memory structures, but also in reflective loops of its connections with outside (in respect to memory) environment (see above).

Also mutual emotional influence appears between them in such interaction. In other words memory does not only manage, but also is managed by external environment in respect of its adjustment to a particular condition.

Concerning the emotional part of information processing it is necessary to stress, that it especially reflects in subconscious intuitive human thinking and at the same time has a significant value in his/her conscious thinking (for example, in speed-up of searching for particular information in his/her internal database).

As a whole, the introduced Conceptual Model explains a lot of psychological phenomena, which are related to processes of thinking, as well as clarifies the material substance of its mechanisms (including human’s capability, erudition< smartness, inspiration etc). It supports our sureness in its plausibility.

It is worth to pay special attention to the fact, that though perceptive meaning is introduced in the Model, the Model itself can serve as a sufficient basis for its further advanced concretizations.

About Practicability of Conceptual Model Bionic Use

Conceptual Model (*CM)* has already essential bionic value [1]. But it makes no sense to apply nature as a whole to machines (e.g. legs and wheels). For example, when implementing the analogues of *CM* characteristics in computer architecture, it is necessary to remember its destination first of all.

For developing the universal high-performance and high-intellectual computers (i.e. computers which have enough advanced internal intelligence), it is quite reasonable to reflect the following characteristics in their architectures (as some analogues of the main *CM* principles) [8]:

- Distributed processing the information and its operational storage (i.e. processing part of the machine is to be some memory-processor medium).
- The two-component machine computational process: first, the successive one which perceives the user’s tasks, initiates, organizes and controls the process of their performing; and second, parallel one which is the component of the total computational process and is responsible for performing the tasks in each own branch.

- Machine tasks and knowledge representation as semantic and associative networks realized by graphs and their hierarchical processing; knowledge is represented by complex date structures at its upper level and in details at the lower level.
- Possibility to adapt to tasks and to organize the total computational process step-by-step with its dynamic scheduling and controlling the results obtained at each step.

The universal computer with the architecture built on the basis of specified principles must promote the effective realizing of different information technologies of alternative classes: a symbolism and connectionism including neural-computer ones. But, in the second case the technologies are realized at the program Models, which can also have sufficiently high characteristics (e.g. a great number of neuron-like elements) and can be structurally realized in part by paralleling processing in memory-processor medium. Furthermore, such computer must promote the effective technology realizing and integrating different technology processes.

A new class of multimicroprocessor cluster computers with specified properties (the so-called intellectual solving machines (ISM)) and, in particular, it Models for broad using was developed in the Institute of Cybernetics of NAS of Ukraine) (this project was supported by the grant from the USA chief of the project – Prof. Koval V, beside him the main authors Prof. Bulavenko O. and Prof. Rabynovych Z.) [9, 10].

ISMs in accordance with above-mentioned fact combine the distributed information processing with the internal higher-level language (which has the developed means of knowledge presenting and processing) and dynamic centralized-decentralized (successive and parallel respectively) control. Exactly this set of characteristics stipulates the belonging of ISM to the new class.

The high level of machine intellect structurally realized in their cluster architectures contribute the essential increasing the efficiency of user-machine interaction and possibility to perform, at that, complex computations which need the deep fractalbility of processing data. It provided by using in programming the high and very high level languages and automatic hierarchical organization of program transformation and performance.

Bibliography

[1] Rabynovych Z. L.: A particular bionic approach to structural Modeling of purposeful thinking. // Cybernetics, No 2.(1979) 115–118

[2] Voronkov G. S., Rabynovych Z. L.: Sensor and lingual system - two forms of knowledge representation. // News of artificial intelligence. (1993) 116–124

[3] Rabynovych Z. L.: About Thinking Mechanisms and Intelligent Computers. // Cybernetics and system analysis, No 3. (1993) 69–78

[4] Haken H., Haken-Krel M.: Mystery of perception. // Translation from German, Moscow: Institute of computer researches (2002)

[5] Gladun V.P.: Decision planning. //Kyiv: Nauk. Dumka (1987) 167p.

[6] Glushkov V.M., Rabynovych Z. L.: Problems of deductive constructions automation. // Information management and intelligence. Moscow (1976) part 4, chapter 2, 300 – 326.

[7] Geoffrey E. Hinton mapping part-whole hierarchies into connectionist networks. //Artif. Intellig. 46 (1990) No 1/2, 47–75

[8] Rabynovych Z.L.: About artificial intelligence concept and its development. // Cybernetics and system analysis, No 2 (1995) 163–173

[9] Koval V.N., Bulavenko O.N., Rabynovych Z.L.: Intellectual solving machines as a basis of highly productive computational systems. // Controlling systems and machines. 36 (1998) 43--52.

[10] Koval V., Bulavenko O., Rabynovych Z.: Parallel Architectures and Their Development on the Basis of Intelligent Solving Machines. // Proc. of the Intern. Conf. of Parallel Computing in Electrical Engineering. - Warsaw, Poland, September 22-25 (2002) 21–26.

Author's Information

Zinoviy Rabynovych – professor, doctor of technical sciences, Academician Glushkov Institute of Cybernetics, Academician Glushkov avenue 40, Kyiv-187, 03680 Ukraine, e-mail: glad@aduis.kiev.ua

MODEL CONSTRUCTION AND RESEARCH USING SYSTEM COMPOSITIONAL APPROACH ON NATURAL HIERARCHICAL NEURAL NETWORKS. DEVELOPMENT OF COMPUTER TOOLBOX

Yuriy Belov, Sergiy Tkachuk, Roman Iamborak

Abstract: System compositional approach to model construction and research of informational processes, which take place in biological hierarchical neural networks, is being discussed. A computer toolbox has been successfully developed for solution of tasks from this scientific sphere. A series of computational experiments investigating the work of this toolbox on olfactory bulb model has been carried out. The well-known psycho-physical phenomena have been reproduced in experiments.

Keywords: system compositional approach, mathematical and computer modelling, elementary sensorium, hierarchical neural networks, computer toolbox, olfactory bulb.

ACM Classification Keywords: J.3 Life and medical sciences - biology and genetics; I.2.6. Artificial intelligence: learning - connectionism and neural nets; I.5.1 Pattern recognition: models - neural nets; I.6.3 Simulation and modelling: applications; H.1.m Models and principles: miscellaneous.

Introduction

Physics, mathematics and modern computer science are universally recognized instruments for research of complex processes and phenomena of the real world. In addition to traditional research domains of these sciences more and more disciplines are being involved into their sphere of interest. In scientific literature dedicated to interdisciplinary exploration the expression “mathematical and/or computer model” occupies the most prominent place.

This paper is dedicated to the topic of mathematical and computer modelling and research of cognitive processes in the brain.

Recently a lot of scientific researches were conducted, where those phenomena and processes are investigated, which have never been involved into the sphere of physico-mathematical and computer applications. The tendency to formalization appears especially in those knowledge domains, where a direct experiment giving the possibility to collect reasonably complete and objective information about the reality under research is impossible in practise. It is commonly known that neuron sciences for instance occupy one of the leading places in modern biology according to the number of physicians, mathematicians and computer science specialists involved competing with molecular biology, genetics, and biotechnologies. According to the complexity of appearing interdisciplinary problems neuro-sciences even leave others behind.

Fast accumulation of enormous amount of experimental data, especially in the last decades of twentieth century and the beginning of the new century has prepared a foundation for trying to develop (on a basis of modern imaginations and possibilities) a new conception concerning the natural mechanisms of recognition, memory and purposeful thinking. Also alternative approaches exist, which are dictated by queries of both fundamental and modern practical medicine and by search of new non-traditional ways of creation of “intellectual” technics.

The idea, that theoretical constructions can appear only on a basis of wide experimental material reflecting the subject under investigation completely, is still popular in the scientific society. However, the history of natural science does not prove this conception on the one hand, and numerous examples urge us to think, that the motivating stimulus of developing and creating a new conception is usually a limited set of fundamental facts on the other hand. Though, the experiment gives food for theoretical constructions no any doubt and serves as a foundation for a future theory. It is worth to emphasize, that similarly to the theory, which is supported with experimental facts, the experiment gives useful information only if it is carrying out according to a specific theoretical conception.

In this paper we discuss questions concerning with mathematical modelling and research of cognitive processes inside the human brain. For this matter computer toolbox was introduced and discussed. While being created first of all it was oriented to networks with complex architecture, namely non-linear hierarchical neural networks of interacting neurons and neuron ensembles (which are made of simpler neural networks in turn) with taking into account energy dissipation as a matter of fact. The last circumstance be known let us take into account and model very important aspects connected with self-organization. It is worth to emphasize we will research and model multilevel hierarchical neural networks with forward (ascending, aggregating), backward (descending, decomposition), and cyclic (parallel, positive and negative) connections. At the same time we will operate with so-called basic structures, which probably the complex brain-like structures (e.g. memory), generally speaking, of large dimension are constructed from. We hope that developed in cooperation with our colleagues [1,2] approach will help to achieve deeper understanding of human's nature and brain activity. Necessity for research and modelling of such neural networks with complex architecture appears when solving the tasks of multilevel information processing inside the brain and for computer modelling, complex behaviour, decision making, etc.

At the present moment it ought to be said, that existing experimental data and conceptions concerning the neuronal activity characteristics and interaction principles have not yet led to complete understanding of such information processing procedures as memorization, recognition, thinking, etc. inside the brain. The mechanisms concerning the functioning of attention, distinction of unconscious and conscious psychical processes, influence of emotions, etc. are still not explained.

Mathematical Model of Elementary Sensorium: Basic Notions

By the *neural model* of sensorium we further imply a model, consisting of neurons and synapses, which incorporate a complex hierarchical network structure (generally speaking, of high dimensionality) of interacting neurons and neuronal ensembles. *Synapse* is treated as a connection of two *neurons*. Neurons and synapses are treated atomic.

Neurons are connected between each other using synapses. In turn, synapses and neurons are connected with pre- and postsynaptic *membranes*. If a pathway of a signal transmission is from the neuron to synapse, then the membrane is called *presynaptic*, if a signal is transmitted in reverse direction, then the membrane is called *postsynaptic.*

Let's regard the representation of sensory (non-verbal) information inside the brain. Cnsider, taking into account [3], that:

1. there is a model of outside world in the brain (neuronal engram);
2. information about sensory environment is transmitted into the brain being encoded by sensory systems;
3. the model of sensory environment is represented as sensory systems with their supermodal level (neuronal model);
4. basic units of nervous system, neurons, correspond to the objects of outside environment;
5. objects of sensory environment effect on neuronal model;
6. changes of the model – "informational processes in the sensory part of brain"

Main Basic Elements and Compositions of the Model of Elementary Sensorium

Let's pay more attention to the discussion of the main basic notions and elements as well as of the compositions of the model of elementary sensorium.

The model consists of *synaptic levels* (SL). There are *symbol* and *quasi-symbol neurons* (SN and QSN) on every synaptic level, which form *symbol* and *quasi-symbol fields* respectively. The main difference between symbol and quasi-symbol neurons is in what functions they perform and how their activity is interpreted, though both have a similar structure. Symbol neurons correspond to particular objects as a whole. Those quasi-symbol neurons, which are connected by positive backward links with some symbol neuron, represent properties of the single object, which the symbol neuron corresponds to. The higher level of hierarchy they are located on, the more complex object (and more complex properties of former) they represent [2]. Let's define symbol and quasi-symbol neurons in aggregate be *principal*.

Let's define symbol neurons SL-0 as *receptors*. Both symbol and quasi-symbol neurons can exist on every level. The only exception from this rule is SL-0 – the level of receptor neurons. This one does not contain quasi-symbol neurons. Receptor neurons correspond to indecomposable elementary objects, which the system of generators of higher level SL symbol neurons are defined by. Receptors are symbol neurons themselves.

Let us distinguish quasi-symbol neurons of SL-1 separately. We define them be quasi-receptor neurons as long as they duplicate receptor neurons [2]. The set of quasi-receptor neurons are denoted as *quasi-receptor field*.

Symbol and quasi-symbol neurons in the model are organized into the *basic structures* (BS), which form a hierarchical neural network. The notion of basic structure is introduced based on neuron structures defined in [2-3]. Every basic structure is defined by certain symbol neuron located, for determinacy, on SL-*i*. This neuron is termed *determinative* for BS. BS consists of the determinative neuron *N* itself, the set of quasi-symbol neurons K_i on SL-*i*, which have positive feedforward and feedback with SN *N*, the set of symbol neurons S_{i-1} from SL- *i*-1, whose axons converge to the determinative neuron of BS. All synapses and inserted neurons, which a connection between *N* and K_i, *N* and S_{i-1}, K_i and S_{i-1} is realized by, belong to the basic structure also. Aforementioned basic structure is defined as BS corresponding to the neuron *N*.

it is significant , that the neural network, which is not provided with mechanisms for new BS creation, cannot be trained to recognize new objects. It is capable to recognize only those objects, which corresponded symbol neurons exist for.

Further the principal parts of basic structure components are defined.

Let's consider the *i*-th synaptic level. A *symbol group* corresponds to each SN. Let's define the symbol group of the determinative neuron of the *i*-th synaptic level as a part of corresponding BS, which consists of quasi-symbol neurons, synapses, inserted neurons and synapses, which belong to the *i*-th synaptic level and mediate connections between the symbol neuron and quasi-symbol neurons corresponded to it. It is significant to note, that the connections of the symbol neuron with other ones in the same SL are not included here.

Let us define a notion of *converging group* for the symbol neuron *N* from SL-*i*. This group is formed with symbol neurons of SL- *i*-1, which alter, most often indirectly via synapses and inserted neurons, the state of *N* (i.e. alter its membrane potential), and also with all intermediate neurons and synapses, i.e. neurons whose output signals are input signals for *N*. Note, that even though this influence can be mediated with other neurons, it cannot be mediated with other principal neurons.

A notion of type of the neuron and single-type neurons is very important. Informally, the neurons are single-type neurons if they excite to the same quality input signals. Let's introduce formal notions. Let's define a notion of *type of neuron* for symbol neurons. On SL-0 the types of neurons are given as initial characteristics of the neural network and are the elements of some set of elementary types. This set is denoted as *RT*. For SL-*i* ($i \geq 1$) the notion is given inductively. Let's consider the symbol neuron *N* on SL-*i*. Let the symbol neurons from the converging group of neuron *N* have types $t_1, t_2, \ldots, t_n$. Then the type of neuron *N* is $\{t_1 \cup t_2 \cup \ldots \cup t_n\}$ by definition. The type of quasi-symbol neuron is defined by the types of symbol neurons, whose output signals are input signals for the considered quasi-symbol neuron. It is significant, that in the model the types of these neurons coincide. if types of two neurons coincide, then they are *single-type neurons*,. Obviously the *single-type relation of neurons* is equivalence relation.

Based on the paper [4] as well as on papers [1,2] for more precise modelling it is worth to take into account, that before the impulses of single-type symbol neurons reach the target symbol neuron on the next level, the initial signals could undergo some modifications, while passing through the inserted neurons and the row of synapses. At the same time, the signals from single-type neurons can interact independently on the signals of neurons of other types. As a result a notion of *uniform converging group* is introduced. Its definition is just the same as one of converging group with a bit difference: uniform converging group comprises by those and only those neurons of converging group from SL-*i*-1, which are single-type neurons. Consequently, for the symbol neuron its converging group is decomposable into a set of uniform converging groups. Note, that there is a particular set of synapses and neurons, which the uniform converging groups interact through. At the same time these neurons and synapses are not contained by any uniform converging group themselves.

A *projective group* of quasi-symbol neuron *N* of SL-*i* is a set of neurons and synapses, which consists of quasi-symbol neuron *N*, single-type symbol neurons from SL-*i*-1, whose outputs are inputs of *N*, and also synapses and inserted neurons which these connections are mediated by.

A *descending group* of quasi-symbol neuron N_k of SL-*i* is formed with neuron N_k itself, all quasi-symbol neurons from SL-*i*-1 accepting the input (possible indirectly) from N_k without intermediate principal neurons, and also all intermediate neurons and synapses (if any). Note, that descending groups appear for neurons on SL-*i* for $i \geq 2$, as long as quasi-symbol neurons appear starting from SL-1.

Horizontal pair of symbol neuron N_s of SL-*i* is a neuron N_s' from SL-*i*, N_s itself, to which a signal is passed to from N_s' without other intermediate principal neurons. All synapses and inserted neurons, through which the signal is passed from N_s' to N_s belong to horizontal pair as well. A notion of *horizontal co-pair* is similar to one of horizontal pair with a single difference: *N* is not a recipient but a source of a signal.

A *horizontal group* of symbol neuron *N* is a union of all its horizontal pairs.

A *horizontal co-group* of symbol neuron *N* is a union of all its horizontal co-pairs.

Basic Properties of Notions Defined for Elementary Sensorium

Taking into account neuro-physiological data [4], particular relations have to be held between uniform converging groups, projective groups and symbol group. Let us define them formally. Consider a symbol neuron N_s is in SL-*i*. Let N_k be a quasi-symbol neuron, which belongs to a symbol group of neuron N_s. By definition for aforementioned notions the following condition hold:

SCP1. Let $n_1, n_2, \ldots, n_p$ be a set S of all symbol neurons, which are included in projective group of quasi-symbol neuron N_k on SL-*i*. Then S is equal to a set of all neurons SL-*i*-1, which belong to a particular uniform converging group of the symbol neuron N_s. At the same time N_k is included into the symbol group of N_s. The inverse assertion holds as well. The set of symbol neurons S of SL-*i*-1 of some uniform converging group N_s coincides with a set of symbol neurons, which such quasi-symbol neuron N_k' exists for, that S is a set of symbol neurons of the projective group N_k', while N_k' itself belongs to the symbol group of N_s. A projective group with a set of symbol neurons *S* and a uniform converging group with a set of symbol neurons *S* on SL-*i*-1 are referred as *corresponding*.

SCP2. Vertebrates have the following feature for some sensory systems, for olfactory system [4] in particular: often in the corresponding uniform converging and projective groups intermediate elements between the set *S* and target symbol and quasi-symbol neurons are equal. The neuron processes, diverging on the output, are different only. Some of them are inputs of the symbol neuron, others – quasi-symbol. Further such corresponding groups are referred as *adjacent*. Note, that inside the olfactory bulb (OB) exactly the adjacent projective and uniform converging groups hold.

Let us specify a property, which links uniform converging and descending groups (SCD1). Let two single-type symbol neurons N_s^1 and N_s^2 of SL-*i*-1 belong to the converging group of the symbol neuron N_s on SL-*i*. These neurons belong to the same projective group of certain quasi-symbol neuron N_k (see SCP1). Let quasi-symbol neurons $N_{k,1}^1, \ldots, N_{k,m}^1$ and $N_{k,1}^2, \ldots, N_{k,n}^2$ (and only they) belong to symbol groups N_s^1 and N_s^2. Then these neurons belong to the descending group of the quasi-symbol neuron N_k. Let us define the part of descending group of neuron N_k, which consists of quasi-symbol neurons $N_{k,1}^1, \ldots, N_{k,m}^1$ and intermediate synapses and neurons, which $N_{k,1}^1, \ldots, N_{k,m}^1$ are connected with N_k by, as *descending symbol subgroup* of the descending group of quasi-symbol neuron N_k corresponding to the symbol neuron N_s. Defined property is a generalization of some results from of paper [4].

Correspondence between Defined Notions of Sensorium and Olfactory Bulb Elements

Let us give a description of the symbol group, which represented in the neural network described in [4]. Tufted cell (TC) represents a symbol neuron. Mitral cells (MC) which correspond to TC represent the quasi-symbol neurons, and signal goes through a granule cell.

Converging and projective groups in the OB cannot be described as a simplest case. Synaptic connections so-called olfactory zones (OZ) are located between receptors (SL-0) and tufted and mitral cells (SL-1), which interact via interglomerular cells. Also inside OZ a pre-synaptic inhibition exists. I.e. integrally the converging groups on SL-1 in OB are much more complex than the simplest case. Former description is a description of the converging group in OB on SL-1 for the tufted cell [4] as well.

In OB the uniform converging groups are strictly expressed – there are tufted cell N_s certain OZ and also all receptor neurons, whose axons are connected with this OZ. There are also some additional connections between various OZ which belong to the converging group N_s – this interaction realized via inter-glomerular cells. Thus, this fact shows additional connections between uniform converging groups which were mentioned above [4].

Regard the horizontal groups and co-groups presented in OB [4]. High order tufted cells influence low order tufted cells via vertical short-axon cells. Thus, higher order tufted cell with some vertical short-axon cell together and synapses between them forms a horizontal pair with tufted cell of lower order, which has synapses with corresponding vertical short-axon cell. In turn tufted cells of lower order form co-pairs with higher order tufted cells. Similarly tufted cells of the same order influence each other via the horizontal short-axon cells [4]. Here also pairs and co-pairs exist, which in turn are parts of groups and co-groups.

The common part of adjacent projective and uniform converging groups is represented in OB by olfactory zones – they represent that common part, which is specified in the definition of adjacent groups [4].

Description of Computer Toolbox for Biological Neural Networks Modelling

The toolbox is a computer programme, input data of it is a neural network and its inputs described in XML language [5]. The input neural network is defined as oriented graph.

Oriented Graph of Neural Network. Vertices and Arcs. The first stage of the neural network construction is specification of vertices. Vertices are intended to define specific points of the model. Under “specific points” we understand locations of the neural network, where some signal transformation, nonlinear as a rule, takes place. During the modelling these locations with sufficient accuracy can be substituted a single point for, i.e. vertex. The examples of specific points are pre-synaptic and post-synaptic membranes, axon hills, etc. The arcs of the graph define the direction of signal transmission. They have such attributes as type, length, and coefficient of signal increase/decrease.

Neurons and Synapses

To specify the network in a more intuitive way the basic types of elements of biological neural network are distinguished. Also using them the method, how to pass signals along the arcs, is specified. In a graph, representing neural network, neurons and synapses represented by its sub-graphs, every arc belongs to only one network element, neuron, or synapse (Fig. 1).

Some vertices can belong to both neuron and synapse simultaneously. In this case vertices model either pre- or postsynaptic membranes. Some vertices in the neural network are input vertices. They correspond to the endings of dendrites of receptor neurons in natural neural networks. For each input vertex the input signal is specified by a set of pairs (instant of time, level of signal).

If the level of signal is needed for a certain instant of time which is not specified in input the signal is calculated as a result of linear interpolation. For all that, the inquired instant of time should be hit between the minimal and maximal instants given in input. All vertices are also output vertices, i.e. it is possible to obtain the level of output signal in any particular instant of time from them.

Designed data structure for vertex description in neural network provides an ability to store the history of signal level in the vertex. it gives ability to analyze the signal changes in the vertex during experiment carrying out. After completion of simulation of the network behaviour the toolbox enables to view the history of every vertex.

Time in the Model. Time in the model is discrete; toolbox enables to define the quantization interval. While simulating the level of signal is recalculated in each point in every quantization instant of time. Conventional time units are used in toolbox.

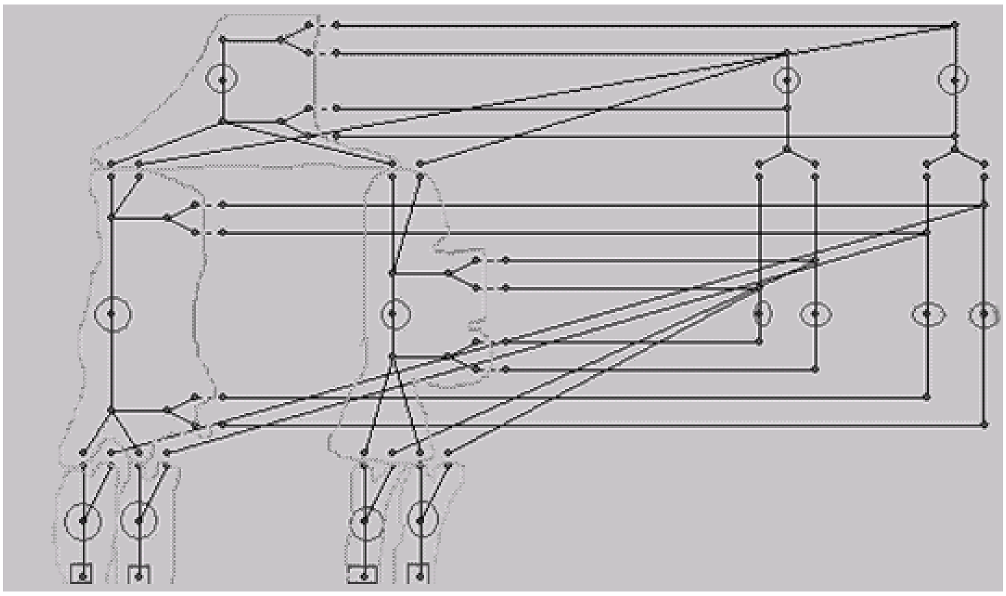


Fig. 1 Example of neural network model in computer toolbox.
For clearness some neurons are rounded with curves. Input vertices are marked with squares. Vertices, where generation of action potential takes place, are marked with circles.

Input Data in the Model. During simulation of aforementioned processes, which take place inside the neural network, there is an ability of real-time visual representation of signal level in any vertex in any quantization instant of time. The more intensive signal level in the vertex in particular instant of time, the bigger circle diameter with a centre in this vertex. Also it is possible to view a diagram of signal level dynamics in any point upon the modelling completion. It is possible to visualize a summary signal level of a particular neuron.

Input Data Representation. The input data is read from XML-document, where the neural network structure is specified. Let us refer to three main elements of this document.

1. *ports* – description of vertices. Each vertex is defined by a title, unique identifier, 2D-coordinates, type (regular vertex or, so-called, generator of action potential (AP), see below). For AP generators the number of vertex (which stores AP sample), a threshold signal level, and coefficients of length and amplitude increment relatively to the sample is defined. Also there is an optional parameter for every vertex – list, which usually defines a signal level of input activity by means of pairs <instant of time, the signal level in a vertex in this instant of time>.

2. *synapses* – description of synapses. The main characteristics of the synapse are its class (chemical or electrical), type and the list of arcs. Every arc is defined by an ordered pair of vertex numbers, length and weight. Note, the notions of length and transmission time of signal in toolbox are synonyms. In the simplest case synapse consists of two vertices, which correspond to pre- and postsynaptic membranes, and one arc, which corresponds to synaptic gap connecting two vertices. There is a possibility to use several arcs to describe synapses more appropriate using facilities presented in toolbox.

3. *neurons* – description of neurons. Each neuron is a particular type of. At this moment there is only one neuron type implemented in toolbox – simple neuron. Similar to synapse, neuron has a list of arcs, the order of which is just the same as the order of arcs in synapses.

Input receptor signals of the neural network are also defined in XML-document, which consists of a list of elements. Each element defines a signal for one of the input vertices as a list of pairs, which specify the level of input signal in a particular instant of time.

Calculation of Signal Values. Each vertex of the network is characterized by certain state in any instant of discrete time $\Delta t \cdot i$, where $i \in [0..n]$, Δt – quantization interval. State is defined by a current signal level of the vertex, the current phase of the action potential and by other parameters depending on the type of the vertex. The state of the network in the current instant of time $\Delta t \cdot (i+1)$ is defined by states of vertices in the network graph.

Let's define a signal processing, which takes place in the vertices. Two types of vertices are used in toolbox: *simple vertices* and *AP generators*. Signal values are defined on inputs of simple vertices in discrete instants of time. Linear spline is constructed based on these points. It can be used in output signal value calculation in any instant of time. AP generators have more complex behaviour. Dependency of the output signal on the input signal is almost the same as in previous item. The important exception, however, is in the following. If signal level reaches the critical level of depolarization and in this instant the vertex is not in state of refractor period, then the action potential with predefined parameters is generated. The type of action potential of the vertex is determined based on the sample taken in [6, pp.27-54], given by a list of pairs of coordinates – dependency of membrane potential on time. In every instant of discrete time in every vertex the values of all adjacent vertices are being corrected, taking into account, which object this vertex belongs to.

Arcs in simple neuron have such characteristics as length and coefficient of signal change, i.e. the signal while passing through the arc is being processed with linear transformation. We stress the simple neuron (not on just any neuron) to emphasize the flexibility and extensibility of toolbox. While time signals in vertices are changed depending on their arc connections with other vertices: $v_j(t) = \sum_{e \in I_j} v_{s(e)}(t - l_e) \cdot c_e$, where generally $v_k(t)$ – the signal value in the vertex *k* in the instant of time *t*, I_j – set of arcs entering the vertex *j*, l_e – the time of signal transmission along the arc *e*, $s(e)$ – the beginning of the arc *e*, c_e - weight coefficient of the arc *e*. Note, that this transformation is inherent to all vertices and is the first transformation, which can be followed by specific transformations related to every particular type of vertex.

In most simple implementations of synapse models the signals are transmitted in a similar way with arcs, but exceptions are regions where mechanisms of plasticity are implemented. Plasticity is implemented by change of the coefficient of signal transmission in synapse according to the following rule:

$$c_{ij}^1 = \begin{cases} c_{ij}^0 \cdot \lambda_{inc}, \ v_j(t + l_{ij}) \geq v_n \\ c_{ij}^0 \cdot \lambda_{dec}, \ v_j(t + l_{ij}) < v_n \end{cases},$$

where c_{ij}^0 – current coefficient of signal conductivity, weight coefficient of a synapse, c_{ij}^1 - new coefficient of signal conductivity while passing along the arc (i, j), $v_j(t + l_{ij})$ – signal level in the vertex *j* in the instant of time $t + l_{ij}$, $\lambda_{inc} \geq 1$ – increment coefficient, $\lambda_{dec} \in (0;1]$ – decrement coefficient of synapse weight, v_n – constant, which specifies a border between the increment and decrement of weight coefficient of a synapse. In order to model more precise complicated synapse type is implemented, where the signal is described by integral transformation:

$v_j(t + l_{ij}) = c_{ij} \int_{t-\Delta t}^{t} v_i(\tau) \cdot e^{-\lambda(t-\tau)} d\tau$, where $v_i(\tau)$ – signal level in the vertex i in the instant at the time τ, $v_j(t + l_{ij})$ – signal level in the vertex *j* at the instant of time $t + l_{ij}$, l_{ij} – time of signal transmission along the arc (i, j), c_{ij} – weight coefficient of an arc, Δt – time interval, which is taken into account during the output signal calculation, $\lambda > 0$ – parameter defining signal decrement. In such synapses signal level on post-synaptic membrane at the instant of time $t + l_{ij}$ depends on the level of signal on pre-synaptic membrane during the period of time $[t - \Delta t, t]$. Thus, while calculating the current state of a particular network vertex not only one previous state is taken into account, but all network states, which occurred during a whole period of time. Consequently, more precise modelling results can be obtained. The summary level of signal of each neuron at the instant of time t is calculated with a sum of signals of all arcs of the neuron, where signal of the arc s_{ij} is calculated like $s_{ij}(t) = \int_0^{l_{ij}} v(\lambda)d\lambda \approx \sum_{k=0}^{[l_{ij}/\Delta l]} v(\Delta l \cdot k)\Delta l$, where Δl – sampling interval in numerical integration, $v(\Delta l \cdot k)$ – the level of signal at the distance $\Delta l \cdot k$ from the beginning of the arc, l_{ij} – length of the arc (i, j).

Verification of Conformity of Toolbox Using Olfactory Bulb Model

In this section the testing of toolbox functionality using olfactory bulb model [4] is described. Testing has been performed on the neural network described precisely in [4]. Experiments to prove OB phenomena [7-8] have been carried out.

The constructed neural network of olfactory bulb bases on experimental data described in [4], in major follows the basic conception. Aforementioned programming environment has been used in olfactory bulb modelling. Parameters of OB model are described in XML language.

Inputs are represented by four vertices, i.e. four types of receptors were examined in model, which react differently on complex odours in adequate odour environment [4]. The results of experiments carried out are followed. Signals of the inputs of groups 1-3 during experiments 2-3 have been passed during conventional time intervals 0–5 and 10–15. In experiments 1-3 the signal has been passed into the last output corresponding to mechanoreceptors during all the time while experimenting. Let's describe experiment carried out and obtained results in more details.

Testing of Mechanoreceptors. Pure air was passed to input. Consequently one mitral cell only excited. Other principal neurons did not generate action potentials.

Recognition of Odour in Case of Several Types of Receptors are Excited. Stimuli a, b, c, d were passed to inputs in concentration enough for excitation [4]. MC1 and TC14 reached excitation. Cells MC1 and TC14 generated AP. The rest of tufted cells have not been activated except of TC124, which gave a poor response during odour recognition.

Recognition of Full Odorant Spectrum. A full spectrum of stimuli passed to inputs. All receptors were excited. Consequently, all mitral cells and almost all tufted cells were also excited. However with time all of them were inhibited by TC1234.

Excitation of Principal Neurons non-Connected with Mechanoreceptors. In the presence of low air speed complex odorants TC12, TC13, TC123, and TC23 were recognized. There wasn't air flow component in mentioned odours. It happened only in case when air speed is low – low level of signal at the input of mechanoreceptors in comparison with other types of receptors.

Synaptic Plasticity. We have implemented plasticity of synapses connected with principal neurons. According to computer simulation it is possible to conclude, that aforementioned modification of synapses taken in account in toolbox based on the modelling of plasticity mechanisms results in following. When repeated inputs to receptors passing, the response of corresponding mitral and tufted cells increased in frequency of generated action potentials and in duration of rhythmical activity. It may be concluded, that synaptic plasticity is an important component of short-term memory.

We succeeded to reproduce all phenomena, which had been planned during experimentation. This proves the conformity of toolbox to commonly known morphological, electro-physiological and psychological data.

Conclusions

The system compositional approach to mathematical and computer modelling of certain type of natural hierarchical neural networks is discussed. Fundamental basic components and compositions of the model of elementary sensorium are described. Also basic properties of introduced definitions and notions are established.

Computer toolbox for modelling of informational processes in biological hierarchical neural networks is developed.

A series of computational experiments concerning the functionality of toolbox was carried out on the model of olfactory bulb, where common-known psycho-physical phenomena were reproduced.

Bibliography

[1] Z.L.Rabinovich. About mechanisms of thinking and intellectual computers // Cybernetics and system analysis, 1993, #3, p.63-78
[2] Z.L.Rabinovich, G.S.Voronkov. Representation and processing of knowledge in interaction of human sensory and language neuron systems // Cybernetics and system analysis, 1998, #2, p.3-12
[3] G.S.Voronkov. Information and brain: the sight of neurophysiologist // Neuron computers: development, application. # 1-2, 2002, p.79-86
[4] G.S.Voronkov, V.A.Izotov. Computer modelling of information processing mechanisms in olfactory system. I. Model of structural and functional organization of olfactory bulb elements and receptor epithelium // Biophysics. 2001, vol. 46, prod. 4, p.696-703
[5] M.Graves. Designing XML Databases. Moscow: Williams, 2002, 640p.
[6] R.Shmidt and others. Human physiology: 3 volumes, vol. 1, Moscow: Mir, 1996, 323p.
[7] G.S.Voronkov, V.A.Izotov. Computer modelling of information processing mechanisms in olfactory system. II. Mechanisms of recognition and short-term memory in olfactory bulb: results of computer simulation // Biophysics. 2001, vol. 46, prod. 4, p.704-708.
[8] G.S.Voronkov, V.A.Izotov. Computer modelling of information processing mechanisms in olfactory system. III. Reproduction of psycho-physical phenomena with olfactory bulb model // Biophysics. 2002, vol. 46, prod. 5, p.914-919.

Authors' Information

Yuriy Belov – professor, doctor of physical-mathematical sciences, Taras Shevchenko National University in Kyiv, Ukraine, 03680, Kyiv - 680, Academician Glushkov Avenue 2, building 6, e-mail: belov@ukrnet.net

Sergiy Tkachuk – Ph.D. student, Taras Shevchenko National University in Kyiv, Ukraine, 03680, Kyiv - 680, Academician Glushkov avenue 2, building 6, e-mail: tksergiy@gmail.com

Roman Iamborak – Ph.D. student, Taras Shevchenko National University in Kyiv, Ukraine, 03680, Kyiv - 680, Academician Glushkov Avenue 2, building 6, e-mail: yambor@ukrpost.net

EXPLORATION BY MEANS OF COMPUTER SIMULATION OF NONLINEAR HIERARCHICAL STRUCTURE OF NEURONAL MEMORY ON THE MODEL OF OLFACTORY BULB

Yuriy Belov, Sergiy Tkachuk, Roman Iamborak

Abstract: Results of numerical experiments are introduced. Experiments were carried out by means of computer simulation on olfactory bulb for the purpose of checking of thinking mechanisms conceptual model, introduced in [2]. Key role of quasisymbol neurons in processes of pattern identification, existence of mental view, functions of cyclic connections between symbol and quasisymbol neurons as short-term memory, important role of synaptic plasticity in learning processes are confirmed numerically. Correctness of fundamental ideas put in base of conceptual model is confirmed on olfactory bulb at quantitative level.

Keywords: thinking phenomena, olfactory bulb, numerical experimentation, model, neural network.

ACM Classification Keywords: J.3 Life and medical sciences - biology and genetics; I.2.6. Artificial intelligence: learning - connectionism and neural nets; D.2.5 Software engineering: testing and debugging; E.2 Data storage representations - linked representations, object representation; H.1 Models and principles; I.5.1 Pattern recognition: models - neural nets; I.6.4 Simulation and modelling: model validation and analysis.

Introduction

More and more papers are dedicated to modelling of brain activity and thinking processes in particular lately. This paper concerns the sphere of mathematical and computer modelling and research of cognitive processes in brain. It is continuation of a paper [1]. Because of the great complexity of research object, construction of conception, which doesn't conflict with wide variety of experimental data and conforms to known psychological and psychophysical phenomena, is hard enough. One of few such conceptions is conceptual model described in [2]. This one is used as a base in this paper.

Authors of [2] have carried out qualitative analyses of described conceptual model and haven't found contradictions with experimental materials. That is why authors of this paper have carried out certain qualitative analysis of conceptual model. A computer toolbox for simulation of informational processes in natural neural networks was developed for this purpose.

Olfactory bulb was chosen to carry out numerical experiments because of existence of deep research results in it; detailed data of structure are known [3]. Some essential constituents of thinking such as appearance of learning and identification, memory, imagination occur in the olfactory bulb [4-5]. Authors' attention is concentrated just on them.

This paper introduces experiment statements and their interpretations as well. When carrying out latest ones authors make their aim to confirm correctness of conceptual model [2] by computer simulation as much as possible on experimental object chosen.

Correspondence between Cells of Olfactory Bulb and Conceptual Model

There are unambiguous correspondence among many cells of olfactory bulb and conceptual model proposed by the reason of conceptual model and olfactory bulb is in relation of abstract – specific respectively. Basic correspondences between cells of olfactory bulb and ones of conceptual models are listed in Table 1.

Table 1. Basic correspondences between cells of olfactory bulb and ones of conceptual model

Olfactory bulb [3]	**Conceptual model [2]**
olfactory bulb	neuronal model, which satisfies conditions of conceptual model
tufted cell	symbol neuron
mitral cell	quasisymbol neuron, quasireceptor neuron (more precisely

Experimentation on Olfactory Bulb Model

Every experiment consists of two parts:

1. Experimentation. There is description of actions made by experimenters. Construction of neural network for toolbox, essential input data to former and measurement of network output data were carried out in this part as well.
2. Interpretation of the experimentation results. What way obtained results fit the conceptual model were emphasized in this part in.

Every experiment description follows in detail. Note, planning of experiments and carrying out them have coincided because of absence of possible unexpected difficulties while experimenting.

Output Signals and Identification. Input signals sufficient for activation were being sent to inputs of receptor neuron, corresponding to one olfactory zone [3], during the time interval from 0 till 5 time units. Output was measured from quasireceptor neuron corresponding to olfactory zone above. As a result action potentials (APs) were being generated by receptor neurons for period of time during which input signals were sent. After that formers finished (fig. 1). But generation of APs was going on in quasireceptor neuron after instant of time 5 as well.

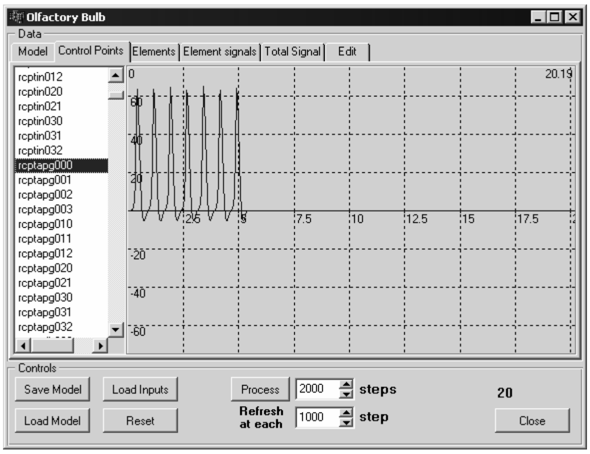


Figure 1. Generation of AP in receptor neuron.

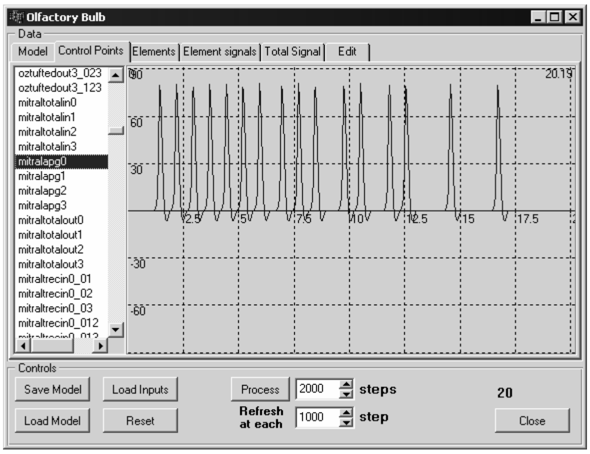


Figure 2. Generation of AP in quasireceptor neuron.

As well as in quasireceptor neuron after stopping sending of input signal to receptor neuron, input generation of AP of symbol neuron was going on too (fig. 3). That may be caused by large weight of connection between symbol and quasirepector neurons.

Output signals of quasireceptor neuron, but not symbol one were analyzed in this experiment as a distinction with [3] it was performed. Former inconsistency between [2] and [3] is caused by fact, that authors of conceptual model described in [2] hold the opinion, which has some differences with one described in [3-5].

Since quasireceptor neuron excited after input signal to receptor neuron had stopped secondary spikes have been got [2]. It is evidence of identification of input stimulus because of quasireceptor neuron corresponding excited receptor neuron has excited. The fact of generation of AP after finishing sending of receptor input signals to model indicates the short-term memorizing of stimulus as well.

Validation of "Mental View" Existence. Input signal was sent to postsynaptic membranes of symbol neurons (but not receptor neurons) during the time interval from 0 till 5. Sending input signals was stopped after. During the time interval from 0 to 5 symbol neuron was generating AP. Some time after instant of time 5 AP was being generated, it stopped later (fig. 4).

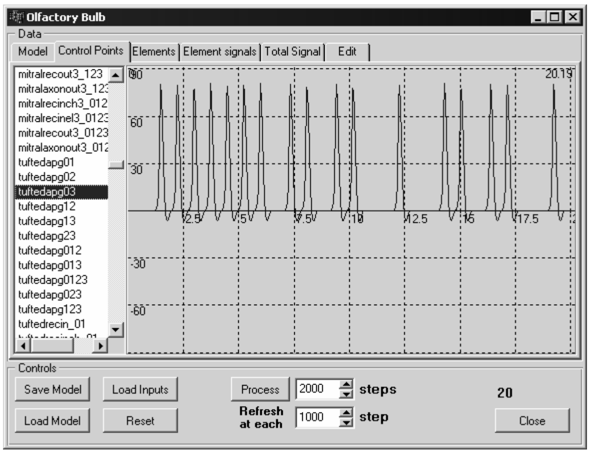


Figure 3. Generation of AP in symbol neuron.

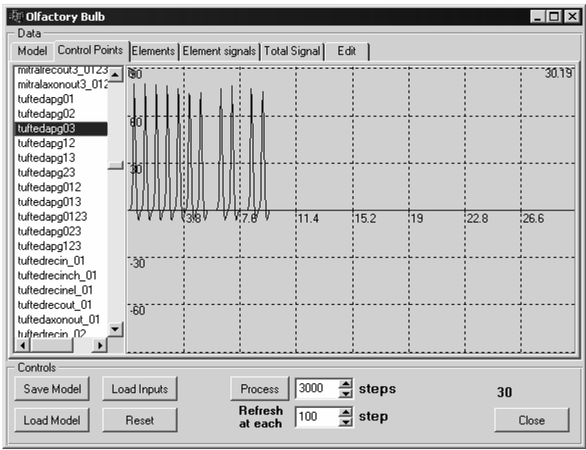


Figure 4. Generation of AP in symbol neuron.

When input signals was begun to send to input of symbol neuron, quasireceptor neurons began to generate APs too. After finishing sending signal to inputs of symbol neuron at instant of time 5 generation of APs in quasireceptor neurons was going on (fig. 5).

Exciting of mitral cells took place in this experiment, quasireceptor neurons were excited in other words. In conceptual model former corresponds to "imagination" of object in the moment this one is not represented in environment. In other worlds mental view [2] existence was confirmed in olfactory bulb.

Short-term Memory as Excitation Feedforward and Feedback Between Symbol and Quasisymbol Neurons. Connections from tufted cells to mitral ones going through granule cells [3] were broken before experimentation in this experiment. During experimenting on olfactory bulb more precise definition of breaking had to be done since connections between symbol neurons and quasisymbol ones in our case are not direct, but though the granule cells [3] and are much complicated than simplest case of conceptual model. As the result there are possible several implementations. Thus three cases of breaking connections from tufted cells to mitral ones were distinguished in neural network model of olfactory bulb:

1. by means of removing granule cell and all input and output connections with former;
2. by means of removing all connections from tufted cells to granule ones and from granule cells to mitral ones;
3. by means of removing all connections from granule cells to mitral ones.

In all three cases input signals were been sent to input of receptor neuron during the time interval from 0 till 5 time units. In the issue APs were generated by receptor neurons after the instant of time 5. APs were stopped after of course. (fig. 6).

Activity of principal neurons (symbol and quasisymbol neurons) had some differences by different means of experiment realization.

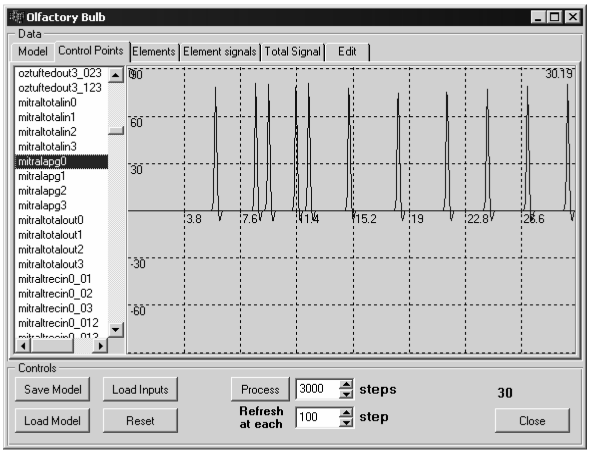


Figure 5. Generation of AP in quasisymbol neuron.

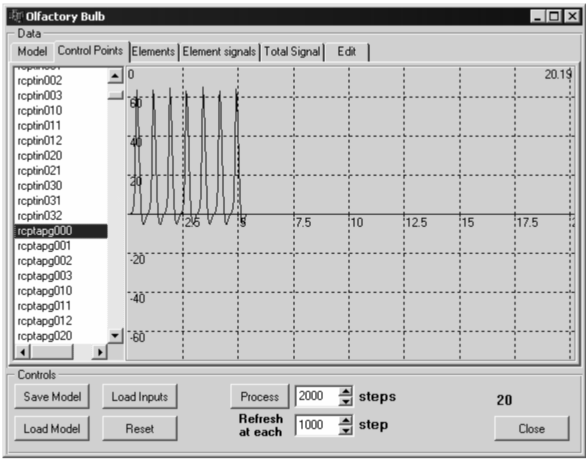


Figure 6. Generation of AP in receptor neuron.

Let consider realization by means of first and second cases. When sending signals to receptor neuron inputs AP were being generated by symbol and quasisymbol neuron. After input signal sending stopped generation of AP stopped there immediately (fig. 7a, 7b, 8).

Breaking connections by means of case 3 when signal sending to receptor neuron inputs stop symbol neurons generated additional AP as a response to inputs from themselves which came to from granule cells.

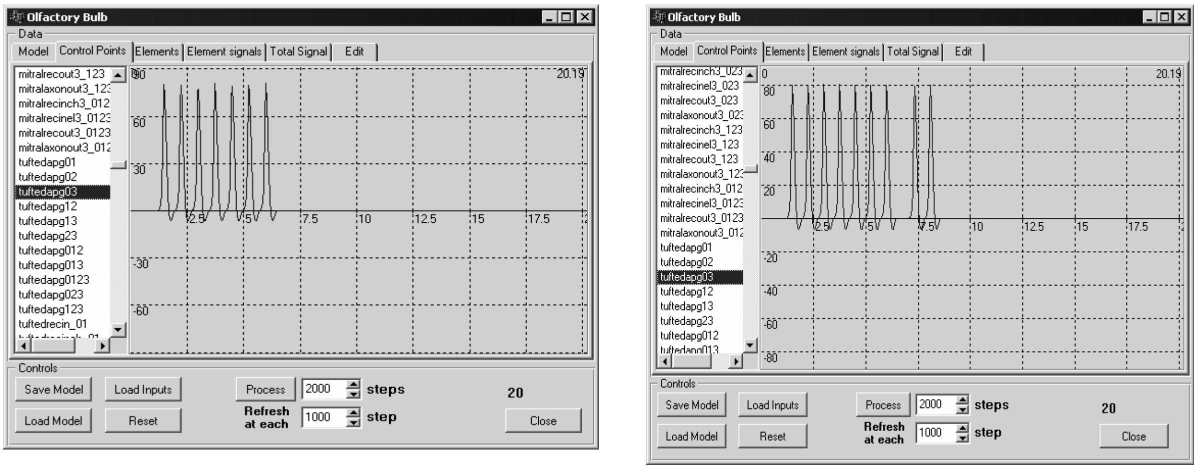
This experiment confirms well known hypotheses adhered by authors as well. It says closed neuronal cycles realize a function of short-term memory

Thus repeated spikes in quasisymbol neurons didn't occurred when cyclic connections mentioned above broken. It can be make up a conclusion that cyclic connection is one of the realization mechanisms of short-term memory.

Learning by the Synaptic Plasticity. Taking into account of modelling of synaptic connection weight growing when sending input signal to receptor neurons one of the short-term and long-term memory mechanisms have Receptor neurons were generating APs during former time intervals (fig. 9).

When sending input stimuli during time interval from 0 till 5 generation of APs by quasisymbol and symbol neurons occurred after input signals sending to receptors had finished (fig. 9-11) along with when input signals were sent.

When sending suitable input stimuli separated by short time (5 time units) while using toolbox it is obvious that output signal spread much longer during the second time interval of input stimuli sending (from 10 till 15) than during the first one (from 0 till 5) (fig. 9-11)



a) b)

Figure 7. Generation of AP in symbol neuron a) by means of removing connection in cases 1 and 2; b) by means of removing connection in case 3.

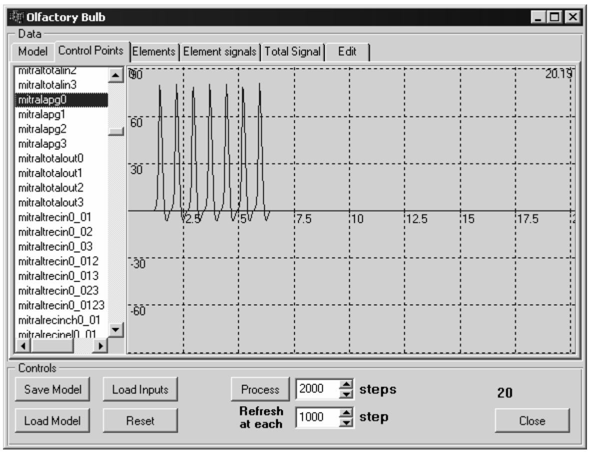


Figure 8. Generation of AP in quasisymbol neuron by means of removing connections in all three cases.

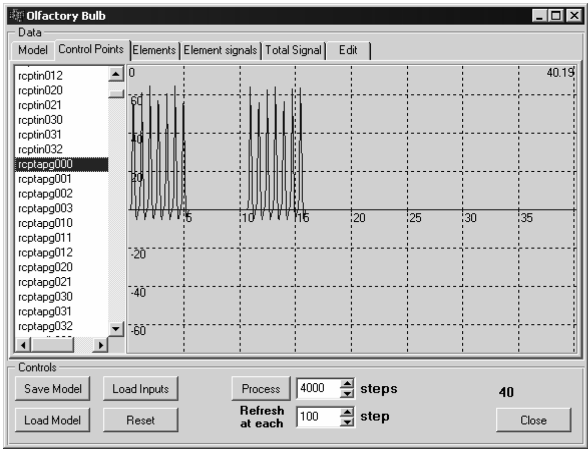


Figure 9. Generation of AP in receptor neurons during the time intervals from 0 till 5 and from 10 till 15.

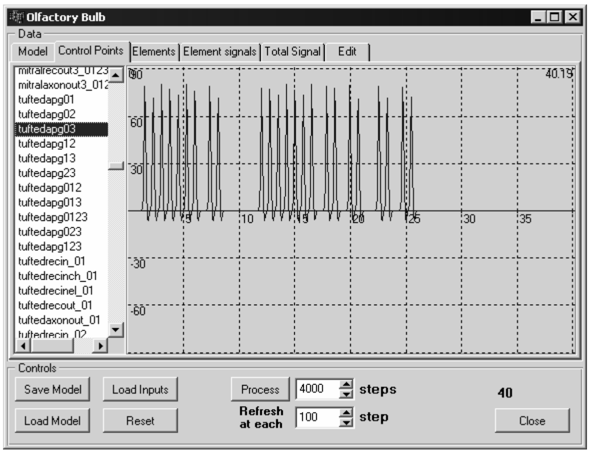


Figure 10. Generation of AP in symbol neurons during the time interval from 0 till 15.

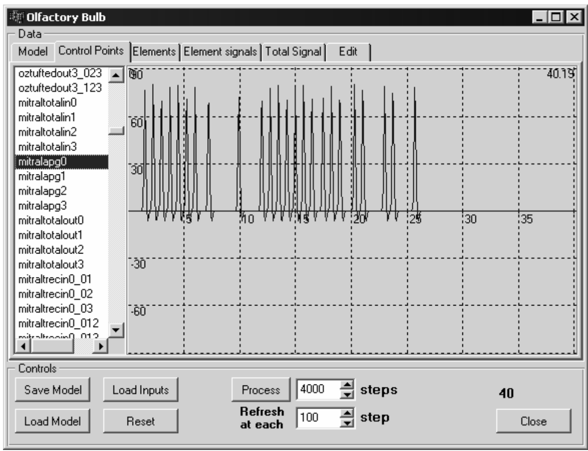


Figure 11. Generation of AP in quasisymbol neurons during the time interval from 0 till 15.

Synaptic plasticity corresponds to increment of weight of feedforwards and feedbacks in conceptual model. Hence one of the mechanisms of learning is realized in terms of conceptual model. Since outputs of mitral cells and tufted ones were spreading longer in presence of growing of synaptic weights in complex neuronal interactions of tufted, mitral and granule cells, it can be concluded that phenomena of synaptic plasticity conform to its function expected in conceptual model entirely.

been realized. It is long-term and short-term synaptic plasticity respectively.

During the time intervals from 0 till 5 and from 10 till 15 units input signals were being sent to receptor neurons. Output signals of symbol and quasisymbol neurons were measured.

Conclusion

Following hypotheses concerning conceptual model have been confirmed in the issue of carrying out of experiments by computer simulation: key function of quasisymbol neurons during the identification of the pattern represented in environment, existence of mental view [2], functions of excitation of cyclic connections (feedforward and feedback) between symbol and quasisymbol neurons as short term memory. Important functions of synaptic plasticity in learning processes are confirmed also.

Described above experiments confirm principal positions of conceptual model on quantitative level. Former positions were discussed as credible hypotheses of its authors before. But it must be emphasized that results of experiments do not ensure the full correctness of conceptual model, they can be treated as partial confirmation of this one.

Principal positions of conceptual model which could be verified on olfactory bulb model were confirmed in this paper. They confirm validity of fundamental backgrounds of conceptual model not only on qualitative level, but on quantitative one too.

Bibliography

[1] Yu. A. Byelov, S. V. Tkachuk, R. V. Iamborak. Mathematical and computer modelling and research of cognitive processes in human brain. Part I. system compositional approach to modelling and research of natural hierarchical neuron networks. development of computer tools. KDS-2005. Proceeding. Volume 1. 2005, Varna, Bulgaria, pp.23-31.

[2] Z.L. Rabinovich About natural mechanisms of thinking and intellectualization of computing machines. Cybernetics and system analysis. No. 5, 2003, pp.82-88.

[3] V.A. Izotov and G.S. Voronkov. Computer Modeling of the Mechanisms of Information Processing in the Olfactory System. I. A Model of Structural and Functional Organization of Neuron Elements of the Olfactory Bulb and the Receptor Epithelium. Biofizika, Vol. 46, No. 4, 2001, pp. 696–703.

[4] V.A. Izotov and G.S. Voronkov. Computer Modeling of the Mechanisms of Information Processing in the Olfactory System. II. The Mechanisms of Identification and Short-term Storage in the Olfactory Bulb: the Results of the Computer Experimentation. Biofizika, Vol. 46, No. 4, 2001, pp. 704–708.

[5] V.A. Izotov and G.S. Voronkov. Computer-assisted Modeling of the Mechanisms of Information Processing in the Olfactory System. III. Reproduction of Psychophysical Phenomena by the Olfactory Bulb Model. Biofizika, Vol. 47, No. 5, 2002, pp. 914–919.

Authors' Information

Yuriy Belov – professor, doctor of physical-mathematical sciences, Taras Shevchenko National University in Kyiv, Ukraine, 03680, Kyiv - 680, Academician Glushkov avenue 2, building 6, e-mail: belov@ukrnet.net

Sergiy Tkachuk – Ph.D. student, Taras Shevchenko National University in Kyiv, Ukraine, 03680, Kyiv - 680, Academician Glushkov avenue 2, building 6, e-mail: tksergiy@gmail.com

Roman Iamborak – Ph.D. student, Taras Shevchenko National University in Kyiv, Ukraine, 03680, Kyiv - 680, Academician Glushkov Avenue 2, building 6, e-mail: yambor@ukrpost.net

SCIT — UKRAINIAN SUPERCOMPUTER PROJECT

**Valeriy Koval, Sergey Ryabchun, Volodymyr Savyak,
Ivan Sergienko, Anatoliy Yakuba**

Abstract: The paper describes a first supercomputer cluster project in Ukraine, its hardware, software and characteristics. The paper shows the performance results received on systems that were built. There are also shortly described software packages made by cluster users that have already made a return of investments into a cluster project.

Keywords: supercomputer, cluster, computer structure.

ACM Classification Keywords: C.5.1 Super computers, C.1.4 Parallel Architectures.

Introduction

To solve the most important tasks of an economy, technology, defense of Ukraine, that have large and giant computing dimensions, we need to be able to calculate extralarge information arrays. Such extremely large computations are impossible without modern high-performance supercomputers.

Unfortunately, such computational resources are almost unavailable in Ukraine today. This can cause a precarious situation development in a different country's life areas. We can lose leading positions in a science, science intensive products' development, complex objects and processes modeling and design technologies.

It is also impossible to import large supercomputers for the above mentioned tasks, because of embargo (for really powerful supercomputers), their extra-large prices, practically impossible upgrade, requirements to control the usage of imported supercomputers from abroad. In this situation Ukraine and other countries (India, China, Russia, Belarus) need to design its national supercomputers [1].

Today in Glushkov Institute of Cybernetics NAS of Ukraine, two high-performance and highly effective computational cluster systems SCIT-1 and SCIT-2 are running in an operation-testing mode. They are built on the basis of modern microprocessors INTEL® XEON™ и INTEL® ITANIUM® 2.

On the basis of these supercomputer systems, a powerful joint computer resource will be built. It will be available for access for users from different organisations from different regions from all the NAS of Ukraine. The systems built are focused to applications from the fields of molecular biology, genetics, science of materials, solid-state physics, nuclear physics, semiconductor physics, astronomy, geology.

Development Ideology

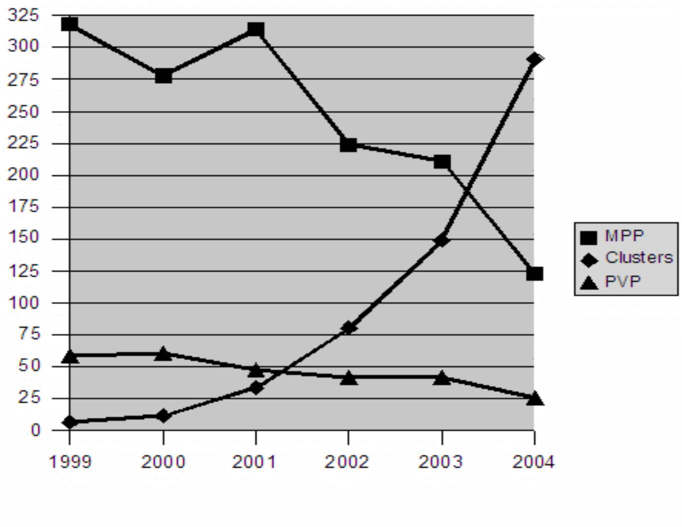
While developing a supercomputer, system scientists and engineers face, a great amount of questions that requires to run a different kind of experiments. The experiments are run to understand a performance, features and characteristics of architecture, hardware platform for computing node solution, node interconnections, networking interfaces, storage system [2].

To make a right **decision on system architecture** we have made an analysis of world supercomputer tendencies. One of major sources we used was top500 list of the largest supercomputer installations. An analysis we made proves us, that a solution of cluster architecture is a right one.

Cluster computer system – is a group of standard hardware and software components, coupled to solve tasks. Standard single processor or SMP (symmetric multiprocessor system) are used as processing elements in a cluster. Standard high-performance interconnect interfaces (Ethernet, Myrinet, SCI, Infiniband, Quadrics) are used to connect processing elements in a cluster system.

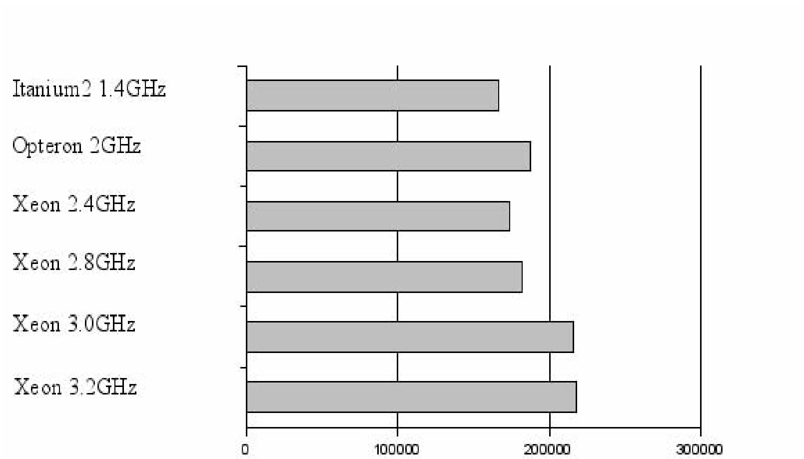
A development of supercomputer systems with cluster architecture is one of the most perspective ways in the world of high-performance computations today. The amount of supercomputer clusters installed in the world in increasing rapidly and the amount of finances spent for this direction is also increased.

Tendencies of a development of supercomputers in the world for **MPP** (Massively Parallel Processing), **PVP** (Parallel Vector Processor) and cluster systems are shown on a Picture 1. As shown on the picture below, clusters are dominated in top500 list. For the several last years, an amount of cluster systems in the list have grown and an amount of **MPP** and **PVP** systems is going down.



Picture 1. World supercomputer tendencies

When making a selection of a hardware platform of computational nodes we analyzed price/performance ratio. As LINPACK is rather narrow test, we choose SPECfp tests understand a performance of nodes on the basis of different kind of real applications. The prices we calculated were taken from Ukrainian IT market operators. The diagram received in analysis is shown on a Picture 2.

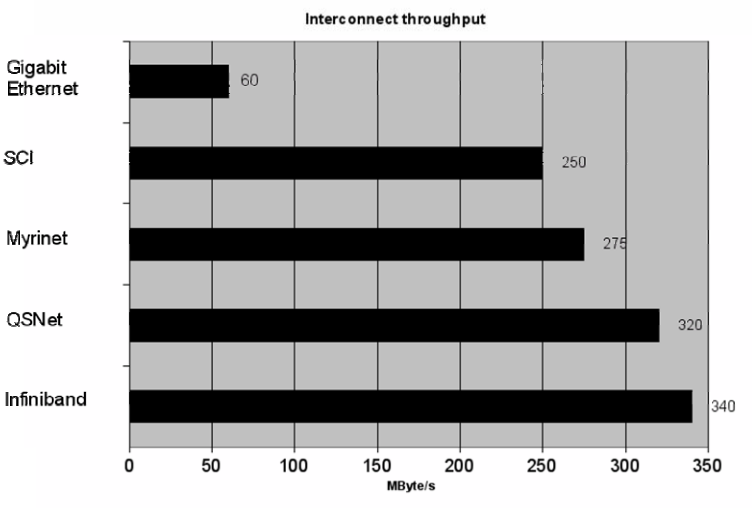


Picture 2. Price/performance ratio on the basis of SPECfp analysis

Today also new SPEChpc tests are available, that can give us an understanding of hpc computers performance for an applications from chemical, environment, seismic area and also OpenMP and MPI applications testing.

Price/performance analysis is made with a calculation of costs of all the main components of a system and its environment with a focus on a theoretical peak 300 GFlops performance, which is near 120 000 SPECfp. We have also take into consideration performance downsize for different platforms scaling on the basis of self made tests.

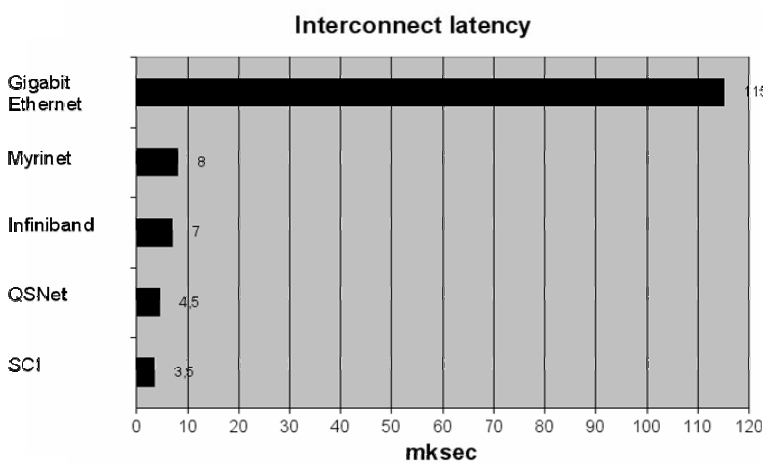
After the analysis, we choose an Itanium2 solution as the best scaling and best price/performance solution for floating point calculation intensive applications. But we understood that the selection of a newest Itanium2 architecture could cause problems with available 32-bit applications porting. So, we decided to build two systems. For SCIT-1 - a 32xCPU system we choose Xeon 2.67GHz platform and for SCIT-2 - a 64xCPU system we choose Itanium2 1.4GHz platform as the best one in 64-bit floating-point performer. It was also taken into account good perspective of Itanium2 architecture and its ability to operate faster with big precision operations and big memory. The other valued characteristics that cause better price/performance ratio of an Itanium2 systems is its best power/performance ratio between other well known general usage processors.



Picture 3. Interconnect throughput.

Design and a selection of internode communicational interfaces was done from the best performing ones. When making experiments with one of the software packages (Gromacs), we have found that a low latency is the most important issue for cluster scalability. We have seen from world published data and our own experiments that some of the tasks which don't scale more then 2-4 nodes on Gigabit Ethernet scales easily to 16 nodes on low-latency interconnect interfaces [3].

Understanding an importance of latency and throughput of an interface, we have made a price/performance analysis for interfaces available in Ukraine. The best one for 16x and 32x nodes' clusters we planned to build was SCI (Scalable Coherent Interface). Performance parameters of communicational interface received on 3rd quarter of the year 2004 for Intel Xeon platforms are shown on Picture 3 and Picture 4.



Picture 4. Interconnect latency

Today these pictures will look different (because of changes in platforms and interfaces itself), but price/performance leaders for latency intensive applications are SCI and QSNetII; for throughput intensive applications they are QSNetII and Infiniband. For small clusters an SCI is a preferable interface. But it has also one more useful feature. An SCI system network can be built on 2D mash topologies. Such architecture gives an ability to transfer data into two ways simultaneously. But to receive an advantage from this technology, software should be written with an understanding of this ability.

It is known that performance and intelligence are the most important factors promoting the development of modern universal high-performance computers. The first factor forced a development of parallel architectures. The rational base of this development is universal microprocessors, connected into cluster system architectures. The second factor becomes clear when the notion of **machine intellect** (MI) is used. The concept of MI is introduced by V.M.Glushkov. MI defines "internal computer intelligence" and the term "intellectualisation" is used to define an increase of machine intellect. During the last 5-6 years, V.M.Glushkov Institute of Cybernetics NAS of Ukraine carries out the research aimed at the development of cluster based, **knowledge-oriented architectures** called *intelligent solving machines* (ISM). ISM implementing high- and super-high-level languages (HLL and SHLL) and effective operation with large-size data- and knowledge bases. They operate as with traditional computation tasks (mathematical physics, modeling of complex objects and processes, etc.) as **artificial intelligence** (AI) tasks (knowledge engineering, pattern recognition, diagnosis, forecasting) [4].

Large-size complex data- and knowledge bases in these clusters are displayed as **oriented graphs** of an arbitrary complexity – trees, semantic networks, time constrained, etc. In ISM computers it is possible to build graphs with millions nodes and to represent various knowledge domains. It is also important that the developed architecture can be easily integrated with distributed database architectures, which are developed in Glushkov Institute of Cybernetics NAS Ukraine. This database architecture makes search processes and data processing much faster than solutions with traditional architectures.

The intellectual part of the cluster systems developed together with distributed databases is an advantage of this solution as compared with the systems developed in the other sites of the world.

Hardware and software of the systems developed. Today the following SCIT (**s**uper**c**omputer for **i**nformational **t**echnologies) supercomputers are built in the institute (Picture 5):

Picture 5. Photo of SCIT clusters.

SCIT-1 – 32xCPU, 16xNodes cluster on the basis of Intel Xeon 2.67GHz 32-bit processors. They are oriented to operate with 64-bit and 128-bit data. The peak performance of SCIT-1 is 170 GFlops with an ability to be upgraded to 0,5-1 TFlops (right on a photo).

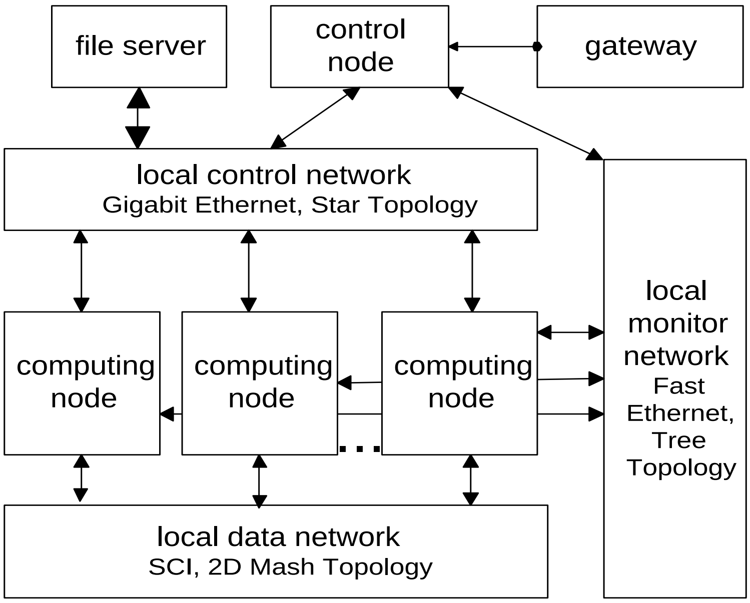
SCIT-2 – 64xCPU, 32xNodes cluster on the basis of Intel Itanium2 1.4GHz 64-bit processors. They are oriented to operate with 128-bit and 256-bit data. The peak performance of SCIT-2 is 358 GFlops with an ability to be upgraded to 2,0-2,5 TFlops. The storage system has capacity of 1 TByte and ability to be upgraded to 10-15 TBytes (left on the photo).

Each of two clusters is an array of computing nodes, connected together with three networks. The first one is a system network, based on SCI interface. The second one is a file data network, based on Gigabit Ethernet interface. The third one is a management network, based on Fast Ethernet interface. A general block-scheme of the SCIT supercomputer is shown on the Picture 6.

A local data network is based on SCI and is used for a high-performance low-latency inter-node communication during a calculation process. A local data network is built as 2D mash. For 16x node cluster it is configured as 4x4 or 2x8 2D mash. For 32x node cluster it is configured as 4x8 or 2x16 2D mash. On data transfers based on MPI the throughput for Xeon E7501 platforms is 250 MB/s, for Itanium2 8870 platforms – 355 MB/s.

A local control network is based on Gigabit Ethernet and is used to handle all cluster-computing nodes and to transfer data files between nodes and file server.

A local monitor network is used for service information transfer and monitoring of all the cluster system.



Picture 6. SCIT cluster structure.

On a table 1 performance parameters of SCIT-1 and SCIT-2 systems are described.

Performance characteristics of developed systems SCIT-1 and SCIT-2 are on the one stage with world best systems. They are also one of the best systems in a world mathematical supercomputing construction.

The creation of cluster systems SCIT-1 and SCIT-2 and their integration and finally launch was made due to a fruitful cooperation of Glushkov Institute of Cybernetics NAS of Ukraine with USTAR scientific and manufacturing company (based in Kiev) and Intel corporation (International). The partners of the institute delivered a technical support and consulting of a project.

Table 1. 64-bit performance parameters of SCIT-1 and SCIT-2 systems.

	SCIT-1	SCIT-2
1 Processors	P-IV Xeon 2,67 GHz	Itanium2 1,4 GHz
2 Peak performance of a single processor		
Integer operations per second, 10^9IPS	1,34	5,6
Floating point operations per second, GFLOPS	5,34	5,6
Node system bus performance, GB/s	4,2	6,4
3 Total peak performance of a system		
Integer operations per second, 10^9IPS	43	358
Floating point operations per second, GFLOPS	170	358
Total system bus performance, GB/s	67,2	204,8
4 Linpack performance of a system, GFLOPS	112,5	280

System Level Software

Components of system level software of a cluster support all stages of user-level parallel software development. They also provide execution of users processes of substantial processing on a solving field. They run on all the nodes of a cluster and a control node as well. Operating system used are **ALT Linux** for SCIT-1 and **Red Hat Enterprise Linux AS** for SCIT-2. Message Passing Interface (MPI) over SCI is used for programming in a message-passing model. In addition, system level software includes optimized compilers of C, C++, Fortran languages for parallel programming, fast Math libraries, etc.

Application Level Software

The powerful hardware, system level, service and specific cluster software integrated in a system is a strong ground for an application level software development. It gives an ability to solve new extra large tasks in a fields of science, economy, ecology, agriculture, technology, defense, space industry, etc.

Due to successful implementations of SCIT systems for the several months after the system was installed a lot of applications were developed and deployed on a supercomputer in Glushkov Institute of Cybernetics NAS of Ukraine.

The software packages for the following tasks were developed:

- soil ecology problems solution;
- seismic data processing;
- dynamical travelling salesman in a real time;
- modeling a structural-technological changes in a developing economy;
- a search for an optimal service center placement;
- construction of an interference-tolerant code;
- risk classification and evaluation decisions;
- data clusterization with genetics algorithms;
- decomposition, calculation, verification and solving of a theorems;
- software component for linear algebra;
- low-energy orbit selection;
- software package for a natural and technogenic processes analysis.

Conclusion

The supercomputer cluster project, as a first stage of a national supercomputer resources development, made a great impact on an intellectualisation of information technologies in Ukraine. The next stage will be devoted to improvement of performance characteristics of supercomputers designed and their software. This should allow extending an amount of large complex tasks that would be solved on the systems.

Bibliography

1. Koval V.N., Savyak V.V., Sergienko I.V., "Tendencies of modern supercomputer systems development", Control Systems and Computers, Vol.6, November-December 2004, 31-44 pp (In Russian).
2. Koval V.N., Savyak V.V., "Multiprocessor cluster systems: planning and implementation", "Nauka osvita", Artificial Intellect, Vol.3, 2004, 117-126 pp (In Russian).
3. www.gromacs.org
4. Koval V., Bulavenko O, Rabinovich Z Parallel Architectures and Their Development on the Basis of Intelligent Solving Machines // Proc. Int. Conf. on Parallel Computing in Electrical Engineering. — Warsaw (Poland).- 2002.- P.21-26

Authors' Information

Valeriy N. Koval – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: icdepval@ln.ua

Sergey G. Ryabchun – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: sr@emt.com.ua

Volodymyr V. Savyak – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: Volodymyr.Savyak@ustar.kiev.ua

Ivan V. Sergienko – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine

Anatoliy A. Yakuba – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova, 40, Kiev, 03680 MCP, Ukraine; e-mail: ayacuba@voliacable.com

NEW KNOWLEDGE OBTAINING IN STRUCTURAL-PREDICATE MODELS OF KNOWLEDGE

Valeriy Koval, Yuriy Kuk

Abstract: An effective mathematical method of new knowledge obtaining on the structure of complex objects with required properties is developed. The method comprehensively takes into account information on the properties and relations of primary objects, composing the complex objects. It is based on measurement of distances between the predicate groups with some interpretation of them. The optimal measure for measurement of these distances with the maximal discernibleness of different groups of predicates is constructed. The method is tested on solution of the problem of obtaining of new compound with electro-optical properties.

Keywords: New knowledge, Predicates, Complex objects, Primary objects, Maximal discernibleness.

ACM Classification Keywords: I.2.4 Artificial Intelligence: knowledge representation formalisms and methods.

Introduction

The present work deals with further development of methods of practical extraction of knowledge from experimental data. Its purpose is development of an effective mathematical method for obtaining of new knowledge on the structure of complex objects with certain properties. The work is focused on solution of an important applied problem - designing of structure of compounds with the needed properties.

In our previous works [1] - [2] in order to obtain new knowledge in form of production rules, a concept of variable predicate, able to accept a number of values - so-called predicate constants, predicates in the conventional sense

- as well as a concept of distance between the predicates were used. Both these concepts were further developed in the present work. However unlike the above mentioned works the present article considers predicates with the subject domains consisting of the objects, possessing an internal structure are considered, that is objects from subject domains of the predicates are considered as complex while earlier they were considered as integral. We will refer to components of complex object as primary objects [3], and to the predicates designating properties and the relations of primary objects as primary predicates. Normally, some information is also known on properties and relations of primary objects being parts of complex objects, which should be used in the procedures of new knowledge obtaining on the structure of complex objects possessing certain properties. The procedure of such knowledge obtaining proposed in the work is based on measurement of the distances between the groups of properties and relations between the primary objects, or in terms of logic, between the groups of predicates with some interpretation of them. The measure introduced by us in the works [1] - [2] for measurement of the predicate affinity degree, cannot be directly transferred on the groups of predicates. Therefore, in the present work a special measure, an optimal one in terms of maximal discernibleness of different groups of predicates for measurement of distances between them is introduced.

1. Structural-predicate Model of Knowledge

It is convenient to represent knowledge of complex objects in the form which we have named *the structural - predicate model of knowledge*. It is a further generalization of the structural - attributive model of knowledge [3]-[4]. The generalization is in that their relations, rather than just the properties of the objects are also considered. For example, the two-place predicate «difference of melting temperatures of two substances is more than Δ » describes a certain relation between two objects.

The structural - predicate model of knowledge (SPMK) is a four-layer graph of a pyramidal network, separate layers of which form its nodes. For clearness fig. 1 represents SPMK, containing knowledge on properties of chemical compounds with various types of crystal lattice structure: such as LiCaAlF6 (L-structure of the lattice), Na2SiF6 (N-structure of the lattice), Trirutile (T-structure of the lattice). Let us designate as P , A , S , V following sets of SPMK nodes. The first layer P corresponds to the predicate constants (values of variable predicates), designating properties and relations of the primary objects. Let the elements P be *primary predicates*. On fig. 1 primary variable predicates are: Tm - melting point, So - standard entropy for corresponding simple oxides, H - standard enthalpy formations for corresponding simple oxides, Rs - radius of ions, C - isobaric thermal capacity. On fig. 1 each of these predicates takes 2 values, and predicate constants corresponding to them are designated with figures 1 and 2.

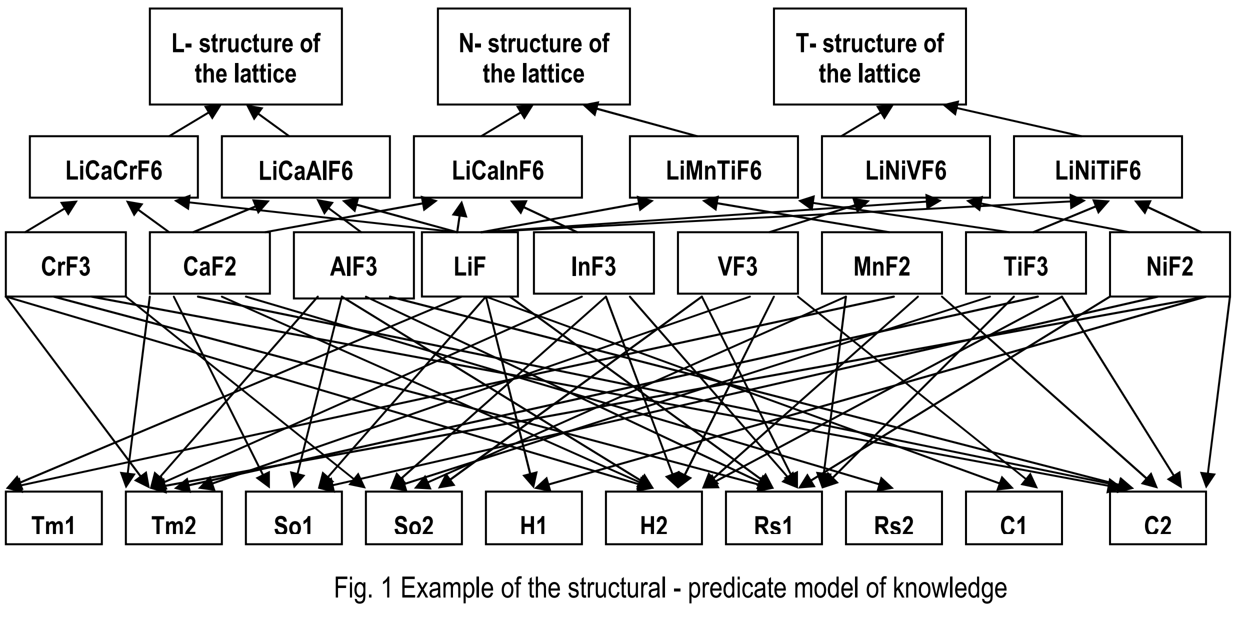


Fig. 1 Example of the structural - predicate model of knowledge

The second layer A corresponds to the names of the primary objects. They compose subject domains of primary predicates at their interpretation. The third layer S corresponds to the names of the compound objects, the fourth layer V - to the values of the variable predicates designating properties and relations of the compound objects. Let us call elements V *predicates of compound objects*. Their subject domains are compound objects. On fig. 1 predicates of compound objects are 3 values of the variable predicate «to have a certain type of the crystal lattice». Arches of the bottom and top layers connect the nodes representing objects to the nodes representing predicate constants and are directed from the primary and compound objects to the predicate constants. They are used in interpretation of the predicates. Let ω designate multiplicity of some predicate constant. Then the presence of the ω arches originating from ω objects and converging in a certain predicate corresponds to the logical value of the predicate "true" in substitution of these objects into the predicate, and to the value "false" in substitution of the object into the predicate in case of absence of the arch, connecting the object with the predicate. Thus, in substitution of the objects A and S from which the arrows go to the predicate constants of the sets P and V, two sets of knowledge R_1 and R_2 in form of true statements on properties and relations of the primary and compound objects are formed instead of the arguments of these predicate constants.

The arches of the middle layer connect the nodes corresponding to the primary objects to the nodes representing the compounds. The primary elements, from which the arches originate, are parts of those complex objects in which these arches terminate.

2. Measure of Affinity of the Predicate Groups

Let's consider a problem of construction of a measure for measurement of degree of affinity of variable predicate groups. This measure should possess the following natural property: the distance between the groups of predicates, measured with it, will not be zero, when these groups of predicates are different. It follows, that it will possess a property of maximal discernibleness of different groups of predicates. Let us construct a measure satisfying this property.

Let's designate N - number of primary predicates in structural - predicate model, M - number of predicates of compound objects, $n(k)$ - number of the primary objects, which are included in complex object s_k. With symbols $p_1, \cdots p_N$ we designate primary variable predicates of the model. In case of numerical values of variable predicates we will base on the following rule: indexes for their predicate constants are selected in the way that their sequence order corresponds to the sequence order of numerical values of the variable predicates. Thus when dividing the interval of variation of numerical values of predicates into segments (quantization), the predicate constant indexes, corresponding to them, will coincide with numbers of these segments.

Definition 1. *Let us understand as the label x_{ik} of a primary variable predicate p_i for complex object s_k* an index of the predicate constant of the predicate p_i which takes the logical value "True" when the primary objects included into s_k and connected by arches with this predicate constant are substituted into it as arguments.

Let us call a vector, the elements of which are labels for complex object s_k of all primary predicates which are included into the structural - predicate model of knowledge, a *distribution of labels* $x_k = (x_{1k}, x_{2k}, \cdots, x_{Nk})$ *of primary predicates for compound* s_k. Let us name the vector $h^{(1)} = (\bar{x}_1^{(1)}, \bar{x}_2^{(1)}, \cdots, \bar{x}_N^{(1)})$, the coordinates of which are equal to average values of components of vectors of labels in this group, a *typical label distribution* for the complex object group $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$. Let us name the vector $\tilde{x}_k = (x_{1k} - \bar{x}_1^1, x_{2k} - \bar{x}_2^1, \cdots, x_{Nk} - \bar{x}_N^1)$ *a centralized vector of labels of primary predicates for the complex object* s_k belonging to the group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$.

Primary predicate label distributions for a certain group of complex objects represent a set of points in the space R_N. If there are two groups of complex objects, we will obtain two such sets of points, which are intermixed in a random way. Following problem emerges: it is required to find the characteristics of these sets, which will allow measuring the degree of affinity of the groups with them. A following solution suggests itself: to take as characteristics of each set of points corresponding points with the average coordinates which represent the typical distribution of the labels for corresponding groups, and to take as a measure of affinity for these groups an Euclidean length of the distance between them. However, this decision is not the optimal one. Indeed, let us consider a following example. Let points of distributions of the labels of the primary predicates for both groups be located on two perpendicular straight lines symmetrically to the their intersection point, with the points of every group lying on the separate line. It can be easily seen, that typical distributions of labels for both groups will coincide, and, hence, the distance between them is zero although the groups of predicates themselves are different. The non-optimal solution above can be seen from another point, allowing to find the optimal one. Let us draw a direct line through two points which are in R_N typical distributions of labels of the primary predicates for complex objects of both groups, and project on it the sets of points for both groups. It is easy to see that average values of projections for both sets of points, the so-called centers of projections $\bar{z}^1$ and $\bar{z}^2$ of corresponding groups, coincide with the points representing typical distributions of labels. As the measure should possess the property of maximal discernibleness of different groups of predicates a following optimization problem arises: to construct in R_N the straight line c, not necessarily going through typical distributions labels, such that the distance between the centers of projections of both sets of points is the maximal one. The criterion of optimization of this problem is: $\bar{z}^1 - \bar{z}^2 \to \max$. The distance $\bar{z}^1 - \bar{z}^2$ obtained in result of optimization should be taken as a measure of affinity of complex objects for both groups.

3. Mathematical Apparatus for Construction of the Measure

It follows from the previous section, that in order to find an optimal solution of the problem of construction of the measure for degree of affinity of groups of variable predicates it is necessary to construct some auxiliary straight line c in space R_N and to project distributions of the labels for both groups of predicates on it. As a result, we will receive two overlapping sets of points on the straight line. As the choice of direction of the line c influences the distances between the projections of distributions of the labels and, hence, on their affinity the straight line should be chosen in the way that the projections of distributions of the labels from different groups of the complex objects are removed from each other as far as possible. Such choice of direction of the straight line will allow distinguishing different groups of complex objects in the optimum way.

Let us name the straight line c, on which distributions of the labels of the primary predicates for complex objects are projected *a projective* line.

Let's cite without proofs a number of auxiliary statements needed to find the required projective line.

Lemma 1. Projection of distributions of the labels $x_k = (x_{1k}, x_{2k}, \cdots, x_{Nk})$ of primary predicates for complex object s_k on the projective line c, going through the point of origin in the space R_N, is defined by the formula $\Pr_c x_1 = c_1 x_{11} + c_2 x_{21} + \cdots + c_N x_{N1}$, where $(c_1, c_2, \cdots, c_N)$ are cosines of the angles, formed by a straight line with the coordinate axis.

Definition 2. *Let us call the total distance* $D_1(z)$ concerning any point z of projections of distributions of labels of primary predicates for group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$.

$D_1(z) = \sum_{\nu=1}^{K} \left\| z_\nu^{(1)} - z \right\|$, where $z_1^{(1)} = c_1 x_{11}^{(1)} + \cdots + c_N x_{N1}^{(1)}, \ldots, z_K^{(1)} = c_1 x_{1K}^{(1)} + \cdots + c_N x_{NK}^{(1)}$.

A scattering in relation to an arbitrary point z of projections of distributions of the labels of primary predicates for group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$

Let us call an average value of projections of distributions of labels of primary predicates for the group G_1: $\bar{z}^{(1)} = \frac{1}{K}\sum_{\nu=1}^{K} z_\nu^{(1)}$ *a center of projections* for the group of complex objects $G_1 = \{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\}$

Lemma 2. *Scattering in relation of an arbitrary* point z of projections of distributions of the labels of primary predicates for group of complex objects $G_1 = \{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\}$ is minimal, when z is equal to the center of their projections: $z = \bar{z}^{(1)}$, thus $D_1(z) = D_1(\bar{z}) = \sum_{\nu=1}^{K} (z_\nu^{(1)} - \bar{z}^{(1)})^2$.

Let $G_1 = \{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\}$ and $G_2 = \{s_1^{(2)}, s_2^{(2)}, \cdots, s_L^{(2)}\}$ - two groups of the complex objects consisting correspondingly from K and L complex objects. For each complex object of these groups, on the basis of the structural - predicate model of knowledge we will construct distribution of labels of its primary predicates. We will obtain $K + L$ vectors, which in the space R_N will be represented with two sets of the vectors: X_1 - the set of vectors $x_1^{(1)}$ $x_2^{(1)}$, ..., $x_K^{(1)}$ and X_2 - the set of vectors $x_1^{(2)}$ $x_2^{(2)}$, ... $x_L^{(2)}$.

Let us project sets X_1 and X_2 on a projective line c. On the basis of lemma 1, we will find the values of projections:

$z_1^{(1)} = \Pr_c x_1^{(1)} = c_1 x_{11}^{(1)} + c_2 x_{21}^{(1)} + \cdots + c_N x_{N1}^{(1)}$, ..., $z_1^{(2)} = \Pr_c x_1^{(2)} = c_1 x_{11}^{(2)} + c_2 x_{21}^{(2)} + \cdots + c_N x_{N1}^{(2)}$,

$z_2^{(2)} = \Pr_c x_2^{(2)} = c_1 x_{12}^{(2)} + c_2 x_{22}^{(2)} + \cdots + c_N x_{N2}^{(2)}$, ... $z_L^{(2)} = \Pr_c x_L^{(2)} = c_1 x_{1L}^{(2)} + c_2 x_{2L}^{(2)} + \cdots + c_N x_{NL}^{(2)}$.

Let's designate sets of projections X_1 and X_2 as Z_1 and Z_2 accordingly, and their centers:

$$\bar{z}^{(1)} = \frac{1}{K}(z_1^{(1)} + z_2^{(1)} + \cdots + z_K^{(1)}), \ \bar{z}^{(2)} = \frac{1}{L}(z_1^{(2)} + z_2^{(2)} + \cdots + z_L^{(2)}).$$

Definition 3. Let us call scattering in relation to an arbitrary point z of projections of distributions of labels of primary predicates for the combined group of the complex objects $G = G_1 \bigcup G_2$ *a total scattering* of both groups $D(z)$.

It is obvious, that it is equal $D(z) = \sum_{\nu=1}^{K} \left\| z_\nu^{(1)} - z \right\| + \sum_{\nu=1}^{L} \left\| z_\nu^{(2)} - z \right\|$.

Let call *a common center* of the combined set of projections $Z = Z_1 \bigcup Z_2$ the value $\bar{z} = \frac{1}{K+L}(z_1^{(1)} + \cdots + z_K^{(1)} + z_1^{(2)} + \cdots + z_L^{(2)})$.

Let's cite without proof following theorem on scattering of projections of distributions of primary predicate labels.

Theorem 1. Total scattering $\overline{D} = \overline{D}(\bar{z})$ in relation to the common center $\bar{z}$ of projections of distributions of the labels of primary predicates of the combined group of complex objects $G = G_1 \bigcup G_2$ is calculated according to the formula $\overline{D} = \overline{D_1} + \overline{D_2} + \hat{D}_1 + \hat{D}_2$, where

$$\overline{D_1} = \sum_{\nu=1}^{K} (z_\nu^{(1)} - \bar{z}^{(1)})^2, \ \overline{D_2} = \sum_{\nu=1}^{L} (z_\nu^{(2)} - \bar{z}^{(2)})^2, \ \hat{D}_1 = K(\bar{z}^1 - \bar{z})^2, \ \hat{D}_2 = L(\bar{z}^{(2)} - \bar{z})^2.$$

It follows from theorem 1 that in order to maximize the measure of discernibleness of both groups of predicates – the distance $\overline{z}^1 - \overline{z}^2$ - it is necessary to maximize the sum $\hat{D}_1 = K(\overline{z}^1 - \overline{z})^2$ and $\hat{D}_2 = L(\overline{z}^2 - \overline{z})^2$. Let us call the sum $D_1 + D_2$ *a complete discernibleness* .

Let us obtain expressions for complete discernibleness and the sums of scattering of label distribution projections of both groups of predicates, which are used for the further calculations.

The vector of the difference of typical distributions of labels for the groups of complex objects G_1 and G_2 let us designate as $h = h^{(1)} - h^{(2)} == (\overline{x}_1^{(1)} - \overline{x}_1^{(2)}, \overline{x}_2^{(1)} - \overline{x}_2^{(2)}, \cdots, \overline{x}_N^{(1)} - \overline{x}_N^{(2)})$.

Let's construct a square matrix $H = h^T h$ where the top index T designates operation of transposing. Dimension H is $N \times N$. It looks like

$$H = \begin{pmatrix} (\overline{x}_1^{(1)} - \overline{x}_1^{(2)})^2 & (\overline{x}_1^{(1)} - \overline{x}_1^{(2)})(\overline{x}_2^{(1)} - \overline{x}_2^{(2)}) & \cdots & (\overline{x}_1^{(1)} - \overline{x}_1^{(2)})(\overline{x}_N^{(1)} - \overline{x}_N^{(2)}) \\ (\overline{x}_2^{(1)} - \overline{x}_2^{(2)})(\overline{x}_1^{(1)} - \overline{x}_1^{(2)}) & (\overline{x}_2^{(1)} - \overline{x}_2^{(2)})^2 & \cdots & (\overline{x}_2^{(1)} - \overline{x}_2^{(2)})(\overline{x}_N^{(1)} - \overline{x}_N^{(2)}) \\ \cdots & \cdots & \cdots & \cdots \\ (\overline{x}_N^{(1)} - \overline{x}_N^{(2)})(\overline{x}_1^{(1)} - \overline{x}_1^{(2)}) & (\overline{x}_N^{(1)} - \overline{x}_N^{(2)})(\overline{x}_2^{(1)} - \overline{x}_2^{(2)}) & \cdots & (\overline{x}_N^{(1)} - \overline{x}_N^{(2)})^2 \end{pmatrix}.$$

Let's consider a matrix H' with elements $h'(\nu,\mu) = \dfrac{KL}{K+L} h(\nu,\mu)$ and designate $A^{(1)}$ and $A^{(2)}$ matrixes the columns of which will consist of components of *centralized* primary predicate label distributions for corresponding groups of complex objects. They contain N lines and accordingly K and L columns.

$$A^{(1)} = \begin{pmatrix} x_{11}^{(1)} - \overline{x}_1^{(1)} & \cdots & x_{1K}^{(1)} - \overline{x}_K^{(1)} \\ \cdots & \cdots & \cdots \\ x_{N1}^{1} - \overline{x}_1^{1} & \cdots & x_{NK}^{(1)} - \overline{x}_K^{(1)} \end{pmatrix}, \quad A^{(2)} = \begin{pmatrix} x_{11}^{(2)} - \overline{x}_1^{(2)} & \cdots & x_{1L}^{(2)} - \overline{x}_{lK}^{(2)} \\ \cdots & \cdots & \cdots \\ x_{N1}^{(2)} - \overline{x}_1^{(2)} & \cdots & x_{NL}^{(2)} - \overline{x}_L^{(2)} \end{pmatrix}$$

Let's consider matrixes $B^{(1)} = A^{(1)} A^{(1)^T}$ and $B^{(2)} = A^{(2)} A^{(2)^T}$. They look like:

$$B^{(1)} = \begin{pmatrix} \sum_{\nu=1}^{K} (x_{1\nu}^{(1)} - \overline{x}_1^{(1)})^2 & \sum_{\nu=1}^{K} (x_{1\nu}^{(1)} - \overline{x}_1^{(1)})(x_{2\nu}^{(1)} - \overline{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{K} (x_{1\nu}^{(1)} - \overline{x}_1^{(1)})(x_{N\nu}^{(1)} - \overline{x}_N^{(1)}) \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{\nu=1}^{K} (x_{N\nu}^{(1)} - \overline{x}_N^{(1)})(x_{1\nu}^{(1)} - \overline{x}_1^{(1)}) & \sum_{\nu=1}^{K} (x_{N\nu}^{(1)} - \overline{x}_N^{(1)})(x_{2\nu}^{(1)} - \overline{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{K} (x_{N\nu}^{(1)} - \overline{x}_N^{(1)})^2 \end{pmatrix},$$

$$B^{(2)} = \begin{pmatrix} \sum_{\nu=1}^{L} (x_{1\nu}^{(1)} - \overline{x}_1^{(1)})^2 & \sum_{\nu=1}^{L} (x_{1\nu}^{(1)} - \overline{x}_1^{(1)})(x_{2\nu}^{(1)} - \overline{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{L} (x_{1\nu}^{(1)} - \overline{x}_1^{(1)})(x_{N\nu}^{(1)} - \overline{x}_N^{(1)}) \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{\nu=1}^{L} (x_{N\nu}^{(1)} - \overline{x}_N^{(1)})(x_{1\nu}^{(1)} - \overline{x}_1^{(1)}) & \sum_{\nu=1}^{L} (x_{N\nu}^{(1)} - \overline{x}_N^{(1)})(x_{2\nu}^{(1)} - \overline{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{L} (x_{N\nu}^{(1)} - \overline{x}_N^{(1)})^2 \end{pmatrix}$$

In following theorems, formulas for calculation of the complete discernibleness and the sum $\overline{D}_1 + \overline{D}_2$ of scatterings are given. The theorems are cited without proof.

Theorem 2. The complete discernibleness is equal to:

$$\hat{D}_1 + \hat{D}_2 = \frac{KL}{K+L} \sum_{\nu=1}^{N} \sum_{\mu=1}^{N} c_\nu c_\mu h(\nu,\mu),$$

where $h(\nu,\mu)$ is an element of the matrix H situated in ν line and μ column.

Theorem 3. The sum of scattering of $\overline{D}_1 + \overline{D}_2$ primary predicate label distribution projections is $\overline{D}_1 + \overline{D}_2 = \sum_{\nu=1}^{N} \sum_{\mu=1}^{N} c_\nu c_\mu b(\nu, \mu)$, where $b(\nu, \mu)$ are elements of the matrix $B = B^{(1)} + B^{(2)}$.

We cite without proof the basic theorem, which allows distinguishing groups of predicates in the way that complete discernibleness is as great as possible.

Theorem 4. Complete discernibleness $\hat{D}_1 + \hat{D}_2$ reaches its maximum at the fixed value of the sum of scatterings of groups when values of the cosines $(c_1, c_2, \cdots, c_N)$ of the pointing angles formed by the projective line with coordinate axis are the components of an eigen vector W for a nonzero eigen value of the matrix $B^{-1}H'$.

6. Stages of Development of the Structure of Complex Objects

Let's consider stages of solution of the problem on development of the structure of the compounds with set properties [3]. For clearness we will accompany this solution with an example based on the data, cited in [4]. Let it be required to develop new compounds possessing electro-optical properties. It is known, that the fluoride crystals, which have crystal structures of types LiCaAlF6 and Na2SiF6, possess electro-optical properties, and the fluoride crystals with the structure such as Trirutile do not possess these properties. For the sake of simplicity, we will limit ourselves to consideration of compounds with structures such as Na2SiF6 and Trirutile. On the first stage of development SPMK, describing the properties of fluoride compounds, is constructed. The fragment of such model is shown in fig. 1. On the second stage we single out in the SPMK a set of predicate constants of compound objects V^+, to which the required properties of projected compound correspond, and the set of predicate constants V^-, to which undesirable properties of projected compound correspond. In our example V^+ is a predicate «to have a structure such as Na2SiF6», and V^- is a predicate «to have a structure such as Trirutile». On the third stage we single out in the SPMK the set of nodes of the group G_1 corresponding to known compounds which have connections with predicates of set V^+, and have no connections with predicates of set V^-, and set of the nodes G_2 corresponding to known compounds which have connections with predicates of set V^-, and have no connections with predicates of set V^+. Let the first group include 10 compounds which are listed in the first column of the table 1, and the second group include 17 compounds listed in the first column of table 2.

Table 1

LiMgAlF6	MgF2	AlF3	1536	13,68	268,7	0,72	14,72	1545	15,8	361	0,39	17,95
LiMnAlF6	MnF2	AlF3	1133	22,25	202,4	0,83	16,24	1545	15,8	361	0,39	17,95
LiCaInF6	CaF2	InF3	1691	16,36	291,8	1	16,02	1445	33,5	250	0,8	15,93
LiMnTiF6	MnF2	TiF3	1133	22,25	202,4	0,83	16,24	1500	21,1	342	0,67	15,93
LiMnVF6	MnF2	VF3	1133	22,25	202,4	0,83	16,24	1679	23,1	271	0,64	21,62
LiMnCrF6	MnF2	CrF3	1133	22,25	202,4	0,83	16,24	1677	22,5	277	0,61	18,82
LiMnRhIF6	MnF2	RhF3	1133	22,25	202,4	0,83	16,24	1460	26	175	0,66	15,93
LiFeGaF6	FeF2	GaF3	1375	20,79	158	0,78	16,28	1225	28	255	0,62	15,93
LiCoInF6	CoF2	InF3	1400	19,59	159,1	0,745	16,44	1445	33,5	250	0,8	15,93
LiNiInF6	NiF2	InF3	1430	17,6	157,2	0,69	15,31	1445	33,5	250	0,8	15,93
		h1	1309	19,92	204,7	0,808	15,99	1496	25,30	279,2	0,64	17,33

Table 2

LiMgCrF6	MgF2	CrF3	1536	13,68	268,7	0,72	14,72	1677	22,5	277	0,615	18,82
LiMgGaF6	MgF2	GaF3	1536	13,68	268,7	0,72	14,72	1225	28	255	0,62	15,93
LiMgRhF6	MgF2	Rh3	1536	13,68	268,7	0,72	14,72	1460	26	175	0,665	15,93
LiNiTiF6	NiF2	TiF3	1430	17,6	157,2	0,69	15,31	1500	21,1	342,2	0,67	15,93
LiNiVF6	NiF2	VF3	1430	17,6	157,2	0,69	15,31	1679	23,1	271	0,64	21,62
LiCoCrF6	CoF2	CrF3	1400	19,59	159,1	0,745	16,44	1677	22,5	277	0,615	18,82
LiCuCrF6	CuF2	CrF3	1043	16,4	128,5	0,73	16,8	1677	22,5	277	0,615	18,82
LiZnCrF6	ZnF2	CrF3	1148	17,61	183	0,74	15,69	1677	22,5	277	0,615	18,82
LiNiFeF6	NiF2	FeF3	1430	17,6	157,2	0,69	15,31	1300	25	239	0,645	15,93
LiNiCoF6	NiF2	CoF3	1430	17,6	157,2	0,69	15,31	1230	27	187,2	0,61	15,93
LiZnCoF6	ZnF2	CoF3	1148	17,61	183	0,74	15,69	1230	27	187,2	0,61	15,93
LiCoGaF6	CoF2	GaF3	1400	19,59	159,1	0,745	16,44	1225	28	255	0,62	15,93
LiNiGaF6	NiF2	GaF3	1430	17,6	157,2	0,69	15,31	1225	28	255	0,62	15,93
LiCuRhF6	CuF2	RhF3	1043	16,4	128,5	0,73	16,8	1460	26	175	0,665	15,93
LiZnRhF6	ZnF2	RhF3	1148	17,61	183	0,74	15,69	1460	26	175	0,665	15,93
LiMgVF6	MgF2	VF3	1536	13,68	268,7	0,72	14,72	1679	23,1	271	0,64	21,62
LiFeCrF6	FeF2	CrF3	1375	20,79	158	0,78	16,28	1677	22,5	277	0,615	18,82
		h2	1352,8	16,96	184,9	0,722	15,60	1474	24,7	245,4	0,63	17,45

On the fourth stage we single out in the SPMK the set of the nodes corresponding to the primary objects for compounds of groups G_1 and G_2, and the set of the nodes corresponding to the primary predicates to which arrows from these primary objects approach. The primary objects are represented in the 2-nd and 3-rd columns in tables 1 and 2. The primary object LiF is included into all compounds, therefore its primary properties do not influence belonging of the compound to a certain group and consequently it is not taken into further consideration. On the fifth stage we find distributions of primary predicate labels for G_1 and G_2 groups of compounds. Each of primary variable predicates takes countable set of values - predicate constants. As their labels we took numerical values of properties of the primary objects, which correspond to them. In the example following 5 primary variable predicates were considered: Tm - melting point, So - standard entropy for corresponding simple oxides, H - standard enthalpy formations for corresponding simple oxides, Rs - radius of the ions, C - isobaric thermal capacity. Their values are represented in columns 4-8 for the primary element of the 2-nd column and in columns 9-13 for the primary element of the 3-rd column, thus the labels of predicates are chosen in the way that they coincide with these values.

On the sixth stage typical distributions of labels h_1 and h_2 are calculated through finding the average for the values in columns 4-13 of every table. h_1 and h_2 are represented in the last lines of tab. 1 and 2. Further there are centralized distributions of labels by way of subtraction of the obtained average values of each column from the actual values of their cells. As a result in the numerical cells of both tables we will obtain the values of the transposed matrixes $A^{(1)T}$ and $A^{(2)T}$. On the seventh stage of development matrixes $B^{(1)} = A^{(1)}A^{(1)^T}$ $B^{(2)} = A^{(2)}A^{(2)^T}$ $B = B^{(1)} + B^{(2)}$, an inverse matrix B^{-1}, and matrixes $H = (h^{(1)} - h^{(2)})^T (h^{(1)} - h^{(2)})$ and $H' = H * \frac{KL}{K+L}$, where $K = 10$, $L = 17$ are calculated. On the eighth stage we find an eigen vector W for a nonzero eigen value of the matrix $B^{-1}H'$ (for example, with Matlab software). For the example considered an eigen vector is $W = (c_1, c_2, \cdots, c_{10}) =$ (-0.0002 0.0359 0.0017 0.6283 -0.0276 0.0004 0.0363 0.0015 -0.7754 -0.0242). Cosines of the angles formed by the optimal projective line c with quadrantal angles are proportional to the values of this vector; thus, the coefficient of proportionality does not play any part. On the ninth stage, projections of typical distributions of labels on this line are found: $\bar{z}^{(1)} = W * h_1^T$ and $\bar{z}^{(2)} = W * h_2^T$. We have $\bar{z}^{(1)} = 1.889$, $\bar{z}^{(2)} = 1.6201$. Their common center is $\bar{z} = 0.5(\bar{z}^{(1)} - \bar{z}^{(2)}) = 1.7545$. On the tenth stage we select in the SPMK primary objects for projected compound in the following way. The objects are selected with connections to the primary predicates, with which the primary objects of group of compounds G_1 also have connections, and with no connections to the primary predicates, with which the primary objects of group of compounds G_2 have connections, thus possible restrictions on the structure of compounds are taken into consideration. Let us assume that the compounds represented in table 3 have been selected.

Table 3

LiMgInF6	MgF2	InF3	1536	13,68	268,7	0,72	14,72	1445	33,5	250	0,8	15,93
LiMnFeF6	MnF2	FeF3	1133	22,25	202,4	0,83	16,24	1300	25	239	0,645	15,93
LiMnGaF6	MnF2	GaF3	1133	22,25	202,4	0,83	16,24	1225	28	255	0,62	15,93
LiMnInF6	MnF2	InF3	1133	22,25	202,4	0,83	16,24	1445	33,5	250	0,8	15,93
LiZnInF6	ZnF2	InF3	1148	17,61	183	0,74	15,69	1445	33,5	250	0,8	15,93
LiCdInF6	CdF2	InF3	1345	20	167,4	0,95	15,93	1445	33,5	250	0,8	15,93
LiMgTiF6	MgF2	TiF3	1536	13,68	268,7	0,72	14,72	1500	21,1	342,2	0,67	15,93
LiMgFeF6	MgF2	FeF3	1536	13,68	268,7	0,72	14,72	1300	25	239	0,645	15,93
LiMgCoF6	MgF2	CoF3	1536	13,68	268,7	0,72	14,72	1230	27	187,2	0,61	15,93
LiFeTiF6	FeF2	TiF3	1375	20,79	158	0,78	16,28	1500	21,1	342,2	0,67	15,93
LiCoTiF6	CoF2	TiF3	1400	19,59	159,1	0,745	16,44	1500	21,1	342,2	0,67	15,93
LiZnTiF6	ZnF2	TiF3	1148	17,61	183	0,74	15,69	1500	21,1	342,2	0,67	15,93
LiZnVF6	ZnF2	VF3	1148	17,61	183	0,74	15,69	1679	23,18	271	0,64	21,62
LiNiCrF6	NiF2	CrF3	1430	17,6	157,2	0,69	15,31	1677	22,5	277	0,615	18,82
LiFeFeF6	FeF2	FeF3	1375	20,79	158	0,78	16,28	1300	25	239	0,645	15,93
LiCoFeF6	CoF2	FeF3	1400	19,59	159,1	0,745	16,44	1300	25	239	0,645	15,93
LiCuFeF6	CuF2	FeF3	1043	16,4	128,5	0,73	16,8	1300	25	239	0,645	15,93
LiZnFeF6	ZnF2	FeF3	1148	17,61	183	0,74	15,69	1300	25	239	0,645	15,93
LiCuCoF6	CuF2	CoF3	1043	16,4	128,5	0,73	16,8	1230	27	187,2	0,61	15,93
LiCoRhF6	CoF2	RhF3	1400	19,59	159,1	0,745	16,44	1460	26	175	0,665	15,93
LiNiRhF6	NiF2	RhF3	1430	17,6	157,2	0,69	15,31	1460	26	175	0,665	15,93
LiCuGaF6	CuF2	GaF3	1043	16,4	128,5	0,73	16,8	1225	28	255	0,62	15,93

To check the correctness of selection, the projection $z = c_1 x_1^{(3)} + c_2 x_2^{(3)} + \cdots + c_N x_N^{(3)}$ of distribution of labels for each selected connection to the projective straight line is calculated. If $\left|\bar{z}^{(1)} - z\right| < \left|\bar{z}^{(2)} - z\right|$, than the selection is considered the correct one. For compounds of table 3 from top to down in succession we find: z equally 1.8395, 1.9060, 2.0089, 2.1339, 1.8838, 2.0870, 1.6272, 1.6115, 1.5784, 1.7448, 1.6329, 1.6715, 1.6521, 1.6135, 1.7291, 1.6173, 1.5107, 1.6558, 1.4776, 1.5440, 1.4934, 1.6136. As $\bar{z}^{(1)} = 1.889$, $\bar{z}^{(2)} = 1.6201$ only the first 6 compounds according to this technique were selected correctly, and others erroneously. The example considered allows also to check correctness of the technique itself as lattice structure of the compounds of table 3 is known from the beginning: the first 6 compounds have structure of Na2SiF6 crystal lattice type, and all subsequent chemical compounds have structure of Trirutile crystal lattice type. Thus, we receive 100 % of correct answers that proves the technique while the work [4] obtains 86,4 % correct answers for the same group of chemical compounds.

Conclusion

The work considers complex objects with internal structure. A structural - predicate model of knowledge, which is a generalization of the structural - attributive model of knowledge is proposed. A method of obtaining of new knowledge on the structure of complex objects with required properties based on measurement of distances between the groups of the predicates with some interpretation of them is developed in the work. An optimal measure for measurement of these distances with the maximal discernibility of different groups of predicates is constructed. The stages of solution of the problem of complex object development are considered.

Bibliography

1. V.N. Koval, Yu.V. Kuk. Distances between predicates in by-analogy reasoning systems, “Information Theories and Applications”, International Journal, vol. 10, N 1, p. 15-22, Sofia, 2003.
2. V.N. Koval, Yu.V. Kuk. Finding Unknown Rules of an Environment by Intelligent Goal-Oriented Systems, “Information Theories and Applications”, International Journal, vol. 17, N 3, p. 127-138, Sofia, 2001.
3. Гладун В.П. Партнерство с компьютером. Человеко-машинные целеустремленные системы. – Киев: «Port-Royal», 2000. –128 с.
4. Величко В.Ю. Розв’язання дослідницьких задач в дискретних середовищах методами виведення за аналогією. – Киев: Кандидатская диссертация. – 2003. – 150 с.

Authors' Information

Valeriy Koval – The Institute of Cybernetics, Head of Department, address: 40, Prospect Glushkova, Kiev, Ukraine; 03680 e-mail: icdepval@ln.ua

Yuriy Kuk – The Institute of Cybernetics, senior scientific researcher, 40, Prospect Glushkova, Kiev, Ukraine; 03680; e-mail: vkyk@svitonline.com .

CLUSTER MANAGEMENT PROCESSES ORGANIZATION AND HANDLING

Valeriy Koval, Sergey Ryabchun, Volodymyr Savyak, Anatoliy Yakuba

Abstract: The paper describes cluster management software and hardware of SCIT supercomputer clusters built in Glushkov Institute of Cybernetics NAS of Ukraine. The paper shows the performance results received on systems that were built and the specific means used to fulfil the goal of performance increase. It should be useful for those scientists and engineers that are practically engaged in a cluster supercomputer systems design, integration and services.

Keywords: cluster, computer system management, computer architecture.

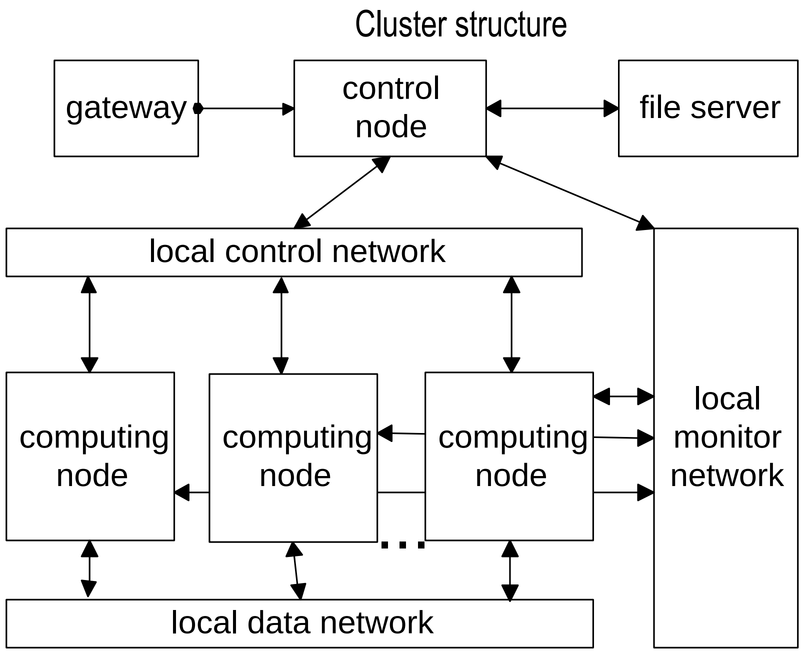
ACM Classification Keywords: C.1.4 Parallel Architectures; C.2.4 Distributed systems; D.4.7 Organization and Design

1. Cluster Complex Architecture

Basis cluster architecture is the array of servers (contains computing nodes and the control node), are connected among themselves by several local computer networks - a high-speed network of data exchange between computing nodes, a network of dynamic management of a server array and a network for cluster nodes monitoring. User access to cluster as a whole can cope by the access server - a gateway on which check of the rights of access of users to cluster and preliminary preparation of tasks for execution is realized. File services are given user tasks by a file server through the cluster control node. A file server in a system provides data access on file level protocols, like Network File System (NFS). A file server is connected directly to a local data network via high throughput channel. In some cases, the gateway and/or file server functions may be carried out on the control node.

Cluster computing node is a server, more often dual-processor, for direct execution of one user task in one-program mode. Computing nodes are dynamically united through a network in a resource for a specific task, simultaneously on cluster some problems may be executed, depending on amount of free computing nodes.

The control node of cluster is a server on which are carried out compilation of tasks, assignment of cluster resources (computing modules - cluster nodes, processors) to the user task, global management of processes activated on nodes during task execution, granting to task needed services of a file server.



2. Dynamic Management with Cluster Nodes

The role of the dynamic management is to manage access to computing nodes and to provide a dynamic reconfiguration of a system. Dynamic management of a cluster system is mostly determined by the used logical systems of a parallel programming (LSPP) (i.e. their architecture and communication libraries). But it can also be influenced by nodes interconnect architecture, rather, a data communication network (means to connect the cluster nodes among themselves and with cluster control node).

A basis of a dynamic cluster reconfiguration under a user task is defined by the list of cluster resources allocated to the task (nodes, processors). After the resources are reconfigured, the system provides a corresponding handling of a user task only within the framework of the appointed resources.

The element of this list of cluster resources is assigning to task the name of node and quantity of processors, which are active in the node. A node always is appointed entirely, whereas the request of a task always specifies necessary amount of processors.

The cluster resources handling system estimates real presence of resources and "collects" the number of processors necessary to a task from the pool of really active nodes at the moment of free nodes request. Processors are allocated always in the cluster node staff, i.e. it is impossible to allocate in one node on one processor to the different tasks, processors of node unused in a task always should stand idle.

In the cluster, where the communication network is based on the switch (Gigabit Ethernet, Infiniband), any of nodes accessible to a task can cope irrespective of other nodes in this configuration up to full restart. Mutual influence of cluster nodes upon serviceability of a communication network does not exist as a whole - it is provided with the switch.

For a network on basis SCI cards the opportunity of a direct handling of the cluster node within the framework of allocated cluster resources is sharply limited, as the communication network "rises" entirely and serviceability of separate node can depend on serviceability of connections with the next nodes essentially [1].

Though at application 2D-and 3D-topology, it is possible the dynamic change of routing that supposes detour short, but defective connection due to working, but longer, connections through other nodes. However if several nodes die, then a general cluster performance is going down up to transition to a disabled condition. On the other hand, when using a central switch (which is not mirrored), the switch causes a death of all the system.

An opportunity of reconfiguration depends also on a usage of local disk memory of the node. For a cluster systems with a distributed storage based on a local node's hard drives there is a problem found with an execution of user tasks in a background batch mode. When a repeated return to a computing process for the task execution is required, it is necessary to receive the same cluster resources for a task that was provided in a previous stage of the task execution (it implicitly demands long reservation of disk resources on all cluster nodes, appointed to a task).

Reduction of negative influence of this restriction is possible only at refusal from the local disk resource for background tasks for the benefit of network file systems (for example NFS) or the general file systems oriented on cluster application (GFS) [2]. This allows do not care about granting the same cluster resources for the task being executed in a background batch mode.

After task is finished, all allocated resources should be returned in a pool of free resources. Rational use of this pool assumes a regular check of resources' state. The system diagnosis and makes a conclusion about an unavailable resources in an emergency configuration to exclude their incorrect usage. This part of a management system is one of the most important parts of all the cluster management software.

3. Management of Cluster Accessibility

There are several approaches known in a field of cluster resources access management. All of them are based on a standard user authentication on a stage of a system user login. After login is made there are following general ways possible:

1. A user receive an access to all cluster nodes, assigned as a resource to one's queued task, i.e. the task is executed on behalf of the user and a user has a full control over the behaviour of nodes, usage of node own resources (main memory, exchanges with a file server and other nodes, employment of the processor) is given to this user.
2. A user receives an access to an interface of a task status control and management of a task execution. Thus, a user has no real access to cluster nodes allocated.

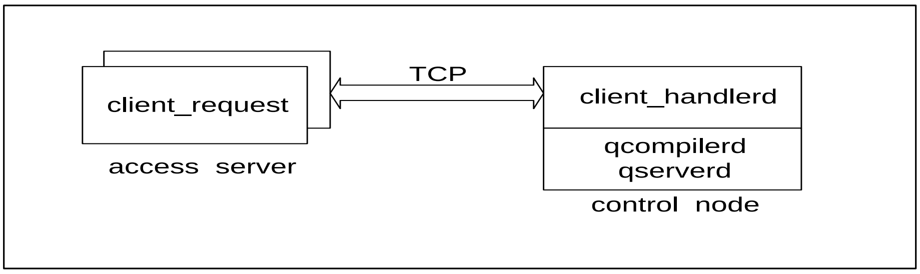
At the first approach the list of users is exported to all cluster parts or the real system user is dynamically created for the period of a task execution. The control over access to the system variables, data and command files of cluster management, nodes essentially becomes complicated, as for communication network SCI this control should be more rigid, than for cluster on the basis of the switch. On the other hand, granting to the user the full access to node allows going to the manual management of task execution up to loading into local disk memory of a node. One of the examples of the mentioned approach implementation is MBC-1000M (Moscow) system [3].

In our opinion, the second approach, despite of considerably big system costs on the organization and support of user work, is represented to more reliable in preservation of integrity of the system software, its functioning and cluster security from the non-authorized access. In this case all works on task execution on nodes are carried out by the specialized pseudo-users existed only on cluster nodes. On behalf of these pseudo-users, the task is executed. For integrity of the approach, an every LSPP has the unique specialized pseudo-user; i.e. the policy of safety does not permit a real user, except for repair managers, to log in into cluster nodes. Such a system provides greater security and reliability of a cluster.

Absence of direct access of the user to cluster nodes is compensated by presence of a specific user interface. An interface allows users to operate a task execution, task queues, to load the data for a task, to supervise a condition of the nodes, which are included in a resource of a task, etc. A program of the user interface cooperates with a demon started on the control node and carrying out all necessary user work. The cluster administrator has the possibility to execute any of these functions.

3. Task Processes Handling

Users, as with the remote access as taking place in a corporate local network, get access only to a gateway - access server, the last holds all user catalogues exported from a cluster file server and support user preparations of the tasks for execution. The subsystem of service of users and their tasks has client-server architecture: the client part settles down on a gateway, the server part - on the control node, connection between these parts is organized under TCP- protocol.



Requests from the user are transferred to the control node and executed by a demon **client_handlerd,** at this one control node can serve a little clusters with identical architecture. The demon **client_handlerd** carries out the requested action and returns result of performance to the user.

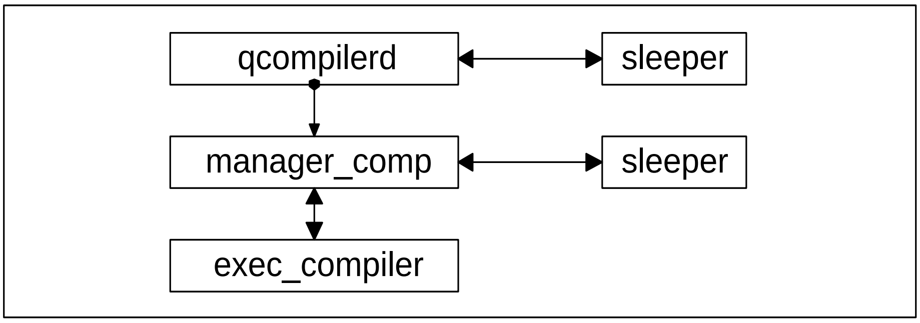
One of such actions is the definition of necessity to compile task and queue it up for compilation with the subsequent placing (at the absence of compiling mistakes) in the execution queue. Each of these queues is served by the demon, correspondingly, **qcompilerd** and **qserverd,** their activities on the control node; their Status may be change only by the cluster administrator. In the same way the user receives data about queued tasks, on cluster congestion, presence of free resources, etc.

The **qcompilerd** functions are:

5 Search of a task (without the control of parameters of task execution);

6 Creation of working structure where this task is compiled;

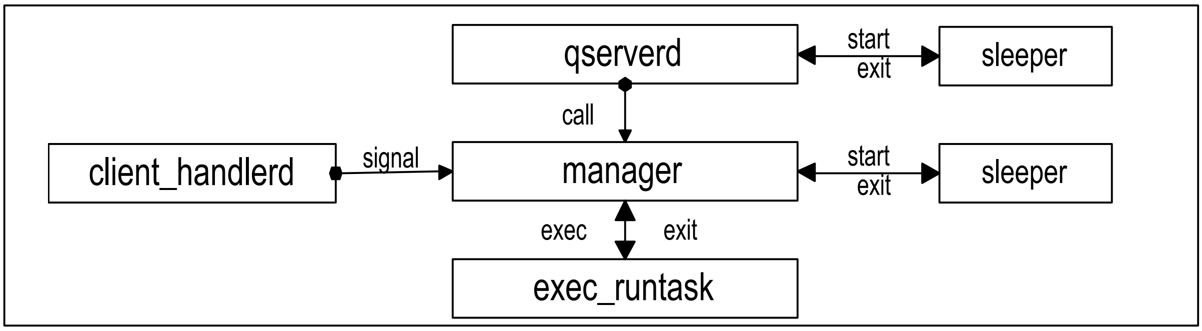
7 Start of the compilation manager, monitoring the specified task, and return to search of other task to compile.



The manager of compilation, in turn, starts as independent process a command file of compilation in a mode *chroot*, expects the end of compilation and returns after that results of compilation in the user individual catalogue.

The **qserverd** functions are:

- Search of a task (with the control of parameters of task execution);
- Updating or creation of working structure of a task for execution;
- Assignment of resources to a task;
- Export of an environment, start of the manager of execution of a specific target and return to search another task for execution.



The manager of execution **(manager)**, in turn, starts as independent process the command file of execution **(exec_runtask)** in a mode *chroot*, expects the end of execution **exec_runtask** or the user signal about task execution stopping - through a demon **client_handlerd -** and returns after that results of execution to the user individual catalogue.

4. Cluster Management System (Base Functions)

Management system – cluster control facilities, used both the system administrator, and various software systems over the operating system, having for an object "continuous" monitoring of computing process, the equipment and the software. It contains, at least, three obligatory parts:

- A direct control of computing process and functioning of the cluster equipment;
- Management of service means of a task stream processing and user works with cluster;
- Monitoring cluster infrastructure (system of power supplies and cooling, a cluster configuration and availability of the cluster components through its communication networks).

The management system may be resident on one of control nodes with an opportunity to change this place to another, and may be distributed among them is depends only on the rules of functioning of managing means.

Obligatory functions of a management system are:

- Management of start, stop and restart all cluster equipment, and its separate nodes and also active means of the cluster system software, in particular, means of a task stream processing;
- Monitor service of the system administrator needs with results of the analysis of a cluster status, its configuration and availability of its nodes;
- Management reconfiguration of node connections if it allows the accepted circuit of a configuration;
- User authentication at its local or remote login to the cluster, support of its functioning during task preparation, granting of help services both online and offline;
- Support of service means of the user interface at compilation, assembly and task execution, under the control of intermediate results over long task execution, on preservation of results of the task running, maintenance of user tasks with services of a file server and DBMS on it;
- Support of a message exchange between the system administrator and users;
- Remote user maintenance with means **upload/download** to transfer the data between its local client computer and the cluster client individual catalogue.

5. Support of the User Computing Process Means

Cluster oriented tasks should use the communication libraries, more often implementing the MPI interface. In this interface the task will start on the zero allocated node with the indication of necessary processors quantity, names of a task code file and some other parameters. For example, **mpirun-np 16 /test/test2**, where **mpirun -** standard command for task start, **np -** required number of processors, **/test/test2 -** a path to the task code file.

Implicitly in this start rights of the owner of the catalogue from which start is carried out, and rights of the owner of a code file are taken into account also. The coordination of these rights and maintenance of start correctness, and also a correctness of access to the data, dissipated upon cluster file system, are assigned to service means.

Compilation of a task is made on behalf of the pseudo-user, representing chosen LSPP, on the control node without attraction of cluster resources with the subsequent transferring the compiled task to queue for execution on cluster nodes under the control of the same pseudo-user determined as the only thing for ordered LSPP.

Client-server means of user's interaction are included into means of support of the user computing process with control facilities tasks also. The accepted principle is the user alienation from executed tasks, that is client, placed in cluster environment, get access only to the gateway – access server physically separated by network addresses from the control node and other cluster nodes, and working areas of the tasks started on execution are placed on the control node. Functioning service means, **client-request** on the access server and **client_handlerd** on the control node, having established connection among them, support it activity till the moment of the termination of concrete user request.

The direct task start is connected to significant inconveniences by the rights of access. More effectively to add additional interfacing means to start the task on allocated cluster resources. These interfacing means should coordinate correctly rights of access during start, estimate and prepare for real use the list of cluster resources, check their sufficiency and, maybe, real availability. As unification of LSPP is absent, these means are individually adjusted on each type of LSPP through environment variables of execution PATH, LD_LIBRARY_PATH and specific ones for concrete LSPP.

Cluster tasks, as a matter of fact, are tasks with great volumes of calculations and consequently, the period of the maximal uninterrupted execution cannot be uncertain, that is why the monopolization of cluster as a whole or only some its parts under one task is incorrect, long on time of the task running should represent a chain of consecutive starts and breaks of the execution (i.e. a set of quantums to run the task), alternated by the idle periods waiting the reception of quantum. Service that means to support the execution of such tasks should provide a correctness of the termination of concrete quantum, preservation of the intermediate data and renewal the execution in the other quantum.

One more service means, facilitated work of users, may be the debugger of cluster tasks, it allows with cluster resources limited from above receiving reports as task executions on the concrete processor, as characteristics of data exchanges between cooperating processors. The attitude to such debuggers dual, rough debugging on them goes conveniently enough and naturally, exact debugging is usually connected to searches of opportunities of increase of task productivity, searches of memory "leakage" and adjustment of a task for the big number of processors, that just and cannot really be supported by noncommercial cluster debuggers.

6. System Means for Increasing the Cluster Performance

Among many means to improve the quality of cluster functioning, it is possible to discuss the basic:

- To carry out hardware improvements in a communication network of nodes, in particular, using network adapters SCI-technology instead of switch oriented Gigabit Ethernet, making up the connections on the basis of 2D-topology (or 3D-topology) and choosing the optimal variant of node switching (i.e., for 16-node cluster with processors Xeon only transition from the network based on switch with Gigabit Ethernet to a network based on SCI gives almost 30 % a gain of performance in Linpack test, and replacement of switching 2x8 nodes on switching 4x4 nodes gives a gain on 4-6 %).
- To maximize the using of node main memory due to exact selection of the used software. So, use only a necessary minimum of demons on node allows to achieve employment of all 12-16MB on the unloaded node.
- To use architecturally – optimized libraries and the compilers giving the most effective codes, in particular, Intel compilers for languages C and Fortran or family compilers GCC, use library MKL (Intel Math Kernel Labs) instead of library ATLAS.

Total results of consecutive changes for 16-node cluster with processors Xeon 2.66 GHz at 2 processors and main memory 1 GByte on node (that gives peak performance in 166 Gflops) are resulted in table 1.

The analysis of table 1 shows, that obligatory elements of cluster adjustment, needed for the maximal productivity, should be - "thin" adjustment of a node main memory for system using, installation, adjustment and use of the richest noncommercial libraries, even for rather weak communication network on Gigabit Ethernet. In case of replacement of switch oriented weak network by more powerful (in particular, by SCI as with Infiniband

[4] we did not have experiments) yet it is necessary to choose rational configuration of data connections, recommended the vendor firms, and to use communication library Scali, instead of MPICH-SCI.

Table 1

Changes in structure and the system software	*The measured maximal performance in Linpack test (Gflops)*	*Ratio max/peak performance (%%)*
Initial configuration: Communication network =Gigabit Ethernet, Accessible MM = 0.83 GByte, Compiler = GNU, Library = ATLAS	71	43
Communication network =SCI, Switching = 2x8, Communication library = MPICH-SCI	94	57
Accessible MM = 0.99 Gbyte	99	60
Switching = 4x4	104	63
Library = MKL, Communication library = SCALI	112	67

One more factor influencing the common cluster performance is a rational choice of structure of file system. Generally, when installation of commercial OS Red Hat Cluster Suite which contains cluster oriented file system Global File System is not supposed, and there is a local system of a data storage based on a RAID-array in the structure of control node entering or served by the specialized server, and local disk memory on cluster nodes is absent, the most effective means may appear export of references to contents of a RAID-array to all points of the cluster where work with files is supposed. Thus even for the user individual catalogues which formally should be on a gateway – access server, their physical accommodation in disk memory of the gateway is not supposed, they only there are exported from a file server by the references.

Similar by results of the decision can be offered for access to files in an executing task - despite of accommodation of the big data files in the individual catalogue of the user, direct access to which to the absolute address from node is impossible, and copying of data files in working structure of task execution is comprehensible only to the small sizes of files (for example, tens Mbytes), indirect addressing through tables of address transformation will provide access to the data of great volume without their moving to working task structures.

The reference to databases, which are stored in the same RAID-array, actually does not differ from described. Unfortunately, experiments in this direction just begin, as well as authentic results are absent.

Bibliography

1. http://www.scali.com
2. http://www.redhat.com/software/rha/gfs
3. http://parallel.ru/computers/reviews/MVS1000M.html (In Russian)
4. http://www.mellanox.com

Authors' Information

Valeriy N. Koval – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: icdepval@ln.ua

Sergey G. Ryabchun – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: sr@emt.com.ua

Volodymyr V. Savyak – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: Volodymyr.Savyak@ustar.kiev.ua

Anatoliy A. Yakuba – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: ayacuba@voliacable.com

AN ANALYSIS OF SOME RELATIONS AMONG DOMAIN ONTOLOGIES[1]

Alexander Kleshchev, Irene Artemjeva

Abstract: Building domain ontologies and applying them to different objectives, researchers faced the fact that many ontologies are associated with one another by one or another relations. Therefore, the problem arose to study relations among different ontologies of the same domains as well as of different ones. A formalization of a relation among domain ontologies is the analogous mathematical relation among mathematical models of these ontologies. The article considers the case when domain ontology model is represented by logical relationship system. Relations among domain ontologies give a possibility to reuse one ontology model when another ontology models are worked out and when new intellectual computer system for same or different domain is worked out.

Keywords: Mathematical model of domain ontology, ontologies representing the same conceptualisation, resemblance between ontologies, simplification of ontologies, composition of ontologies, intellectual task solver.

ACM Classification Keywords: I.2.4 Knowledge Representation Formalisms and Methods, F4.1. Mathematical Logic

Introduction

Building domain ontologies and applying them to different objectives, researchers faced the fact that many ontologies are associated with one another by one or another relations. Therefore, the problem arose to study relations among different ontologies of the same domains as well as of different ones. Although, as noted in [van Heijst et al, 1996], the field is still in its infancy and many questions are unsolved or even unaddressed (for example, how can ontologies be compared and integrated?), by now there has been some information in professional literature related to this problem. Many works studying this problem considered relations among ontologies within the context of ontology integration.

In [Gangemi et al, 1999] ontology integration is defined as the construction of an ontology C that formally specifies the union of the vocabularies of two other ontologies A and B. Three aspects of an ontology are taken into account: (a) the intended models of the conceptualisations of its vocabulary, (b) the domain of interest of such models, i.e. the topic of the ontology, and (c) the namespace of the ontology. The most interesting case is when A and B are supposed to commit to the conceptualization of the same domain of interest or of two overlapping domains. In particular, A and B may be:

Alternative ontologies: `The intended models of the conceptualizations of A and B are different (they partially overlap or are completely disjoint) while the domain of interest is (mostly) the same. This is a typical case that requires integration: different descriptions of the same topic are to be integrated.

Truly overlapping ontologies: Both the intended models of the conceptualisations of A and B and their domains of interest have a substantial overlap. This is another frequent case of required integration: descriptions of strongly related topics are to be integrated.

Equivalent ontologies with vocabulary mismatches: The intended models of the conceptualisations of A and B are the same, as well as the domain of interest, but the namespaces of A and B are overlapping or disjoint. This is the case of equivalent theories with alternative vocabularies.

Overlapping ontologies with disjoint domains: The intended models of the conceptualizations of A and B overlap while the domain of interest are disjoint. This concerns overlapping theories with different extensions. Actually,

1 This paper was made according to the program of fundamental scientific research of the Presidium of the Russian Academy of Sciences «Mathematical simulation and intellectual systems», the project "Theoretical foundation of the intellectual systems based on ontologies for intellectual support of scientific researches".

it is often the case that some fragments from an ontology A can be reused as components in another ontology B that models a different topic.

Homonymically overlapping ontologies: The intended models of the conceptualizations of A and B do not overlap, but A and B overlap. This is the case of two unrelated ontologies with a vocabulary intersection that – if presented – generates polysemy: this is one of the reasons to maintain ontology modules.

To be sure that A and B can be integrated at some level, C has to commit to both A's and B's conceptualizations. In other words, the intention of the concepts in A and B should be mapped to the intention of C's concepts. The authors call this approach principled conceptual integration.

As noted in [Gangemi et al, 1996], the ontological integration envisaged is at a deeper level than representational integration. In fact, the representational integration concerns heterogeneity of formal languages, or heterogeneity of data base schemata. Ontological integration concerns the heterogeneity among conceptualizations.

In [Guarino, 1998] it is noted that information integration is a major application area for ontologies. As well known, even if two systems adopt the same vocabulary, there is no guarantee that they can agree on a certain piece of information unless they commit to the same conceptualization. Assuming that each system has its own conceptualization, a necessary condition to make an agreement possible is that the intended models of the original conceptualizations overlap. Supposing now that these two sets of intended models are approximated by two different ontologies, it may be the case that the two ontologies overlap while the intended models do not. Hence, it seems more convenient to agree on a single top-level ontology rather than relying on agreements based on the intersection of different ontologies.

In [Sowa] ontology integration is defined as the process of finding commonalities between two different ontologies A and B and deriving a new ontology C that facilitates interoperability between computer systems that are based on the A and B ontologies. The new ontology C may replace A or B, or it may be used only as an intermediary between a system based on A and a system based on B. Depending on the amount of change necessary to derive C from A and B, different levels of integration can be distinguished: alignment, partial compatibility, and unification.

Alignment is a mapping of concepts and relations between two ontologies A and B that preserves the partial ordering by subtypes in both A and B. If an alignment maps a concept or relation x in ontology A to a concept or relation y in ontology B, then x and y are said to be equivalent. The mapping may be partial: there could be many concepts in A or B that have no equivalents in the other ontology. Before two ontologies A and B can be aligned, it may be necessary to introduce new subtypes or supertypes of concepts or relations in either A or B in order to provide suitable targets for alignment. No other changes to the axioms, definitions, proofs, or computations in either A or B are made during the process of alignment. Alignment does not depend on the choice of names in either ontology. For example, an alignment of a Japanese ontology to an English ontology might map the Japanese concept Go to the English concept Five. Meanwhile, the English concept for the verb go would not have any association with the Japanese concept Go.

Partial compatibility is an alignment of two ontologies A and B that supports equivalent inferences and computations on all equivalent concepts and relations. If A and B are partially compatible, then any inference or computation that can be expressed in one ontology using only the aligned concepts and relations can be translated to an equivalent inference or computation in the other ontology.

Refinement is an alignment of every category of an ontology A to some category of another ontology B, which is called a refinement of A. Every category in A must correspond to an equivalent category in B, but some primitives of A might be equivalent to non-primitives in B. Refinement defines a partial ordering of ontologies: if B is a refinement of A, and C is a refinement of B, then C is a refinement of A; if two ontologies are refinements of each other, then they must be isomorphic.

Unification is a one-to-one alignment of all concepts and relations in two ontologies that allows any inference or computation expressed in the one to be mapped to an equivalent inference or computation in the other. The usual way of unifying two ontologies is to refine each of them to more detailed ontologies whose categories are one-to-one equivalent.

Alignment is the weakest form of integration: it requires minimal change, but it can only support limited kinds of interoperability. It is useful for classification and information retrieval, but it does not support deep inferences and computations. Partial compatibility requires more changes in order to support more extensive interoperability, even though there may be some concepts or relations in one system or the other that could create obstacles to full interoperability. Unification or total compatibility may require extensive changes or major reorganizations of A and B, but it can result in the most complete interoperability: everything that can be done with one can be done in an exactly equivalent way with the other.

In [Wielinga et at, 1994] more general and more special ontologies are considered. Ontologies can have a recursive structure, meaning that ontology expresses a viewpoint on another ontology. Such a viewpoint entails a reformulation and/or reinterpretation on other ontology. This multi-level organization raises research questions such as the required expressiveness of the mapping formalisms for expressing viewpoints between ontologies. At least two different mapping operations can be identified. The first one is the mapping of terminology in one formalism onto the terminology of another formalism. The other one is the adding of supplementary commitments to one ontology by the mapping of the terms of the ontology onto the terms of the other ontology that takes additional commitments. The first terminology mapping will occur frequently. Since the ontology describes the meaning of the domain theory, for which it is a meta-model, without commitment to the language, in which this meaning is expressed, it will be confronted with meta-models, which partially convey the same meaning, but with different terminology. In this case merging of the two ontologies, or translation of the one ontology into the other is simply a mapping of terminology (e.g. boat in one ontology can be mapped on ship in another ontology if they refer to the same type of object in the universe of discourse (note that the knowledge bases described by these ontologies, even when they describe the same object in the real world, may be totally different!)). The second type of mapping occurs when it is necessary to provide an interpretation of underlying ontology or to provide a more specific interpretation that takes additional commitments. If the more restrictive ontology is already available (such as, sometimes, the ontology of a task or of a method) than it is necessary to map this ontology on the more general one. An example of this type of mapping occurs when there exists a model of the problem-solving task, that should be accomplished, and an existing ontology of the domain of the application. In this case, it is necessary to map terminology from the task (e.g. hypothesis) on terminology of the domain ontology. A simple mapping will not always be possible. Sometimes the ontology - introducing the additional commitments - needs to be constructed. This will often be the case with domain-model oriented ontologies.

In [Laresgoiti et al] and [Schreiber et al] a combination of ontologies is introduced. An example of some artifact such as a ship is considered. One can define multiple viewpoints on a ship. Well-known examples of such viewpoints are the physical structures (what are the parts of a ship?) and the functional structure (how can a ship be decomposed in terms of functional properties?). Although these two viewpoints often partially overlap, they constitute two distinct ways of “looking” at a ship. The purpose of ontology is to make those viewpoints explicit. For a design application such as CAD application, one would typically need a combined physical/functional viewpoint: a combination of two ontologies. For a simulation application (e.g. modeling the behavior of a ship), one would need an additional behavioral viewpoint. Many other viewpoints exist such as the process type in the artifact (heat, flow, energy, ...). Each ontology introduces a number of specific conceptualizations, that allow an application developer to describe, for example, a heat exchange process.

In [Studer et al, 1998] constructing ontologies from reusable ontologies is considered. Assuming that the world is full of well-designed modular ontologies, constructing a new ontology is a matter of assembling existing ones. There are several ways to combine ontologies. In [Studer et al, 1998] the most frequently occurring ones are only given. The simplest way to combine ontologies is through inclusion. Inclusion of one ontology into another has the effect that the composed ontology consists of the union of the two ontologies (their classes, relations, axioms). In other words, the starting ontology is extended with the included ontology. Conflicts between names have to be resolved. Another way to combine ontologies is by restriction. This means that the added ontology only is applied on a restricted subset of what it was originally designed for. The last way to assemble ontologies that is discussed in [Studer et al, 1998] is polymorphic refinement, known from object-oriented approaches.

It is possible to make some conclusions from this overview.

Many authors consider supporting interoperability as a main objective of ontology integration. But if this objective is reached, then it is not clear, what properties integrated ontologies and the result of their integration will have. Before studying these relations and building their formal models, it seems necessary to declare the fundamental properties, that all the relations among ontologies will have.

Consideration of overlapping but different conceptualizations as a necessary condition for possibility of ontology integration seems slightly speculative. If a conceptualization is adequate [Kleshchev et al, 2000a], then it must include the domain reality. In this case, the reality must be a subset of the intersection of these conceptualizations. But the conceptualization that is their intersection is adequate, too. And any top-level conceptualization is worse (wider) than initial ones and especially than their intersection.

Vocabularies (concept systems) are only external structures, by which sets of intended situations, sets of intended knowledge systems and correspondences between them are expressed. Thus, it is unlikely that the union of the vocabularies can be considered as a principal property of ontology integration.

In the same way a mapping of concepts between two ontologies can be but one of ways to determine relations between ontologies. This way cannot be always applied to do this. If there is a mapping between concepts of two ontologies, then this fact alone does not allow us yet to call corresponding concepts as equivalent. The notion of equivalence is defined in mathematics as reflexive, symmetric and transitive relation.

When defining relations among ontologies, any references to properties of inferences or computations cannot be considered as admissible because they darken rather than clarify the meaning of introduced relations. The condition that all the inferences or computations are equivalent cannot be verified.

Properties of Relations among Domain Ontologies

Any domain is characterized by its reality, i.e. by the set of all the possible situations that have ever taken place in the past, are taking place now and will take place in the future [Kleshchev et al, 2000a]. Since the reality is known only partially, the domain knowledge system gives a more comprehensive idea of it. The knowledge system determines the set of situations admitted by the system, i.e. of such situations that are considered as possible in the reality by this knowledge system. So an observer comes across only situations of the reality, but a person possessing a knowledge system is able to imagine situations admitted by the knowledge system. Where does he or she take these imaginary situations from? They are determined by a conceptualization, that can be imagined as the implicitly given set of all the intended situations, i.e. all the situations which can be imagined within the framework of this conceptualization. In this case, the set of the situations admitted by a knowledge system is a subset of the set of all the intended situations.

An investigation of a domain, i.e. of its reality, is aimed at obtaining such a knowledge system that admits the set of situations being as near to the reality as possible. So the set of the situations admitted by a knowledge system is considered as an approximation of the reality, and the investigation of the domain is aimed at obtaining the best (the most adequate) approximation of its reality. This investigation perpetually gives birth to new knowledge systems instead of outdated ones. Where does these knowledge systems come from? They are determined by a conceptualization, too. So a conceptualization can be imagined also as the implicitly given set of all the intended knowledge systems, i.e. of such knowledge systems that can be formed within the framework of the concept system introduced by the conceptualization.

Ontology of a domain is an explicit representation of a conceptualization of the domain. Since the ontology can represent the conceptualization imprecisely, it determines two external approximations both for the set of all the intended situations and for the set of all the intended knowledge systems.

A relation among knowledge systems of the same or different domains is a relation defined on the sets of the situations admitted by these knowledge systems. If this relation takes place among these knowledge systems, and another, more adequate, knowledge system is found instead of one of them, then, in the general case, this relation does not have to take place among the renewed collection of knowledge systems. But from practical needs, it is quite desirable to have a possibility to determine with what other knowledge systems the new knowledge system is in the same relation.

A relation among ontologies of the same or different domains is a relation defined on the sets of all the intended knowledge systems of these ontologies (i.e. a subset of the Cartesian product of these sets) possessing the property that only the tuples consisting of knowledge systems belong to the relation that are in the analogous relation. Thus, if relations among ontologies are determined, then it determines the analogous relation among all the intended knowledge systems of these ontologies. In this article the relations possessing this property are considered only.

A formalization of a relation among domain ontologies is the analogous mathematical relation among mathematical models of these ontologies. The article considers the case when domain ontology model is represented by logical relationship system [Kleshchev, 2000a, 200b].

Ontologies Representing the Same Conceptualization

Domain ontology is a collection of agreements. It defines domain terms, determines their interpretations, contains statements that restrict the meaning of these terms and also gives interpretations for these statements. These agreements are the result of understanding among some members of the community working in this domain [Kleshchev et al, 2000a]. Different members of this community can advance different ontologies of this domain. The question arises: do these ontologies represent the same conceptualization or different ones? Let us discuss this question on the assumption that the models of these ontologies have the form of unenriched logical relationship systems [Kleshchev et al, 2000b].

If a conceptualization is considered as a set of all the intended situations, then two ontologies can represent the same conceptualization only when the sets of terms for situation description in these ontologies are the same. If a conceptualization is considered as a set of all the intended knowledge systems, then two ontologies can represent the same conceptualization only when the sets of terms for knowledge description in these ontologies are the same (they can be empty sets).

Let us consider the case when two different ontologies have the same sets of terms for situation description as well as the same sets of terms for knowledge description. In this case, to be different, these ontologies must have different sets of ontological agreements. Two points of view are possible on the condition under that these ontologies represent the same conceptualization: (1) when both the sets of intended situations and the sets of intended knowledge systems determined by these ontologies are the same; (2) when, following the definitions of the previous section, the sets of intended knowledge systems determined by these ontologies are the same, and for any knowledge system the sets of situations admitted by this knowledge system in these two ontologies are also the same.

The models of these ontologies have the same sets of unknowns and the same sets of parameters but different sets of logical relationships. Formalization of the conditions above means that:

1. the sets of logical relationships for the models of these ontologies are equivalent as applied logical theories (two applied logical theories are equivalent, if they have the same set of models [Kleshchev et al, 2000b]);
2. the models of this domain determined by the models of these ontologies for the same knowledge model have the same models of the reality, i.e. the models of these ontologies are equivalent as unenriched logical relationship systems [Kleshchev et al, 2000b].

It is easily seen that both these conditions are equivalent. Thus, equivalent transformations of the logical relationship set for a domain ontology model (as an applied logical theory) lead to a model of another ontology representing the same conceptualization. These transformations can be, for example, transformation of an applied logical theory to a disjunctive normal form, a conjunctive normal form and so on.

Now let us consider the case when two ontologies of the same domain have the same sets of terms for situation description but different sets of terms for knowledge description. In this case, following the previous section, we can consider these ontologies as representing the same conceptualization, if there is a one-to-one correspondence between their knowledge system sets, and for any corresponding knowledge systems the sets of the situations admitted by these knowledge systems are the same. When passing to models, it means that the models of these ontologies are equivalent [Kleshchev et al, 2000b].

Now let us consider the case when two ontologies of the same domain have different sets of terms for situation description but the same sets of terms for knowledge description. In this case, following the previous section, we can consider these ontologies as representing the same conceptualization if for any knowledge system there is a one-to-one correspondence between the sets of the situations admitted by this knowledge system in both these ontologies. When passing to models, it means that the models of these ontologies have the same sets of all possible enrichments and are isomorphic [Kleshchev et al, 2000b].

Resemblance between Ontologies

In the case when both terms for situation description and terms for knowledge description are different in two ontologies, it is possible to speak of resemblance between these ontologies only (of the same or different domains).

Two knowledge systems related to different ontologies (of the same or different domains) can be considered as resembled if there is a one-to-one correspondence between the sets of situations admitted by these knowledge systems. So two ontologies of the same or different domains can be considered as resembled if there is such a one-to-one correspondence between their sets of intended knowledge systems that any corresponding knowledge systems are resembled. It means that the models of these ontologies are isomorphic [Kleshchev et al, 2000b].

If terms of an ontology are substituted by different terms (by abstract designations), then, as a result, a resembled ontology will be obtained. The resemblance between ontologies is a relation of equivalence. It is reflexive, symmetric and transitive.

Simplification (Coarsening) of Ontologies

Comparing different ontologies of the same domain, one can sometimes say that one of these ontologies is a simplification (coarsening) of another. In the same way considering ontologies of different domains, one can sometimes say that an ontology of one of these domains resembles a simplified ontology of another domain. The availability of more simple and more complex ontologies of the same domain can be important to develop knowledge based systems for specialists of different qualifications (for example, medical systems for physicians of high qualification and for doctor’s assistants).

One can say that a knowledge system related to an ontology is more simple than a knowledge system related to another ontology (of the same or different domains) if for every situation admitted by the second knowledge system (of the more complex ontology) the only situation admitted by the first knowledge system (of the more simple ontology) can be set as corresponding. Then one can consider an ontology as more simple than another ontology (of the same or different domains) if for every knowledge system of the second ontology the only more simple knowledge system of the first ontology can be set as corresponding. It means that a model of the first ontology is a homomorphic image of the second ontology [Kleshchev et al, 2000b].

A domain model <O1, k2> is a simplification (coarsening) of a domain model <O1, k1> if the enriched logical relationship system <O1, k2> is a homomorphic image of the system <O1, k1>. A coarsened model of medical diagnostics can be obtained, for example, by elimination of a few signs.

The simplification determines a partial order of ontologies. If B is more simple than A, and C is more simple than B, then C is more simple than A. If one ontology is simpler than another, and the second ontology is simpler than the first ontology, then they resemble one another.

Composition of Ontologies

When we speak about complex domains, we must usually bear in mind that these domains include knowledge from other different domains. Thus, when knowledge and reality of complex domains are described, concepts related to other domains are used. These other domains are components of the complex domain. Ontologies of complex domains are built from components, which are ontologies of other domains.

We can consider that a (starting) knowledge system related to a complex domain consists of knowledge systems (components) related to other domains if every component is more simple than the starting knowledge system, and the transfer from any situation admitted by the starting knowledge system to corresponding situations admitted by components takes place without the loss of information. The latter statement means that for any two different situations admitted by the starting knowledge system the two sets consisting of the situations corresponding to these two situations and admitted by all the components are different. In this case a starting ontology of a complex domain can be considered as consisting of components which are ontologies of other domains if every component is more simple than the starting ontology, every knowledge system of the starting ontology consists of knowledge systems of components, and the transfer from any knowledge system of the starting ontology to corresponding knowledge systems of components takes place without loss of information. The latter statement means that for any two different knowledge systems of the starting ontology the two sets consisting of the knowledge systems of components corresponding to these two knowledge systems are different. It follows from these definitions that every model of a starting ontology for a complex domain is the product of ontology models that are components [Kleshchev et al, 2000b].

Using Relations among Domain Ontologies for Working out Intellectual Solvers for Applied tasks

At present time, a demand arises to develop program systems for different domains having means for adaptation of problem solving methods to alteration of knowledge in these domains. Such program systems are called the intellectual solvers for applied problems. The base of developing an intellectual solver is domain ontology. Intellectual problem solvers on domain should permit experts and specialists to form and edit ontology and knowledge on domain and to get the programs for solving applied problems in this domain.

If there are alternative points of view on the same domain then we can speak about equivalence or resemblance between different ontologies of the domain. Recognition of the equivalence between alternative points of view on the same domain can give a possibility to solve tasks arising within the framework of a point of view using methods worked out within the framework of another point of view. Recognition of a resemblance between ontologies of the same domain can give a possibility to solve the tasks described within the framework of one concept system by methods developed within the framework of the other concept system. Recognition of a resemblance between ontologies of different domains can give a possibility to solve the tasks of one domain by reasoning using analogy in the case when methods for solving analogous tasks of the other domain have been developed.

A mathematical specification of an applied task can contain a domain model, input and output data of the task, task conditions (a set of formulas), and also criterion of selecting solutions. All the components of the applied task specification are represented in terms of the domain model. If every value of input data is replaced by a variable (different variables correspond to different values) in the task specification then the mathematical specification of the task will be transformed into a mathematical specification of a class of applied tasks. These variables will be called variables of the class of applied tasks. There is a one-to-one correspondence between the set of tasks belonging to the class and the set of all the admissible substitutions of values instead of these variables. To get the mathematical specification of an applied task belonging to a class it is necessary to replace all the variables of the class by values of input data.

If the domain model is replaced by the domain ontology model and knowledge base of the domain are considered as another set of input data of all the tasks of the class then the mathematical specification of the class of applied tasks will be transformed into the mathematical specification of the class of applied tasks corresponding to the domain ontology. There is a one-to-one correspondence between the set of tasks belonging to the class and the Cartesian product of the set of all the admissible substitutions of values instead of variables of the class of the tasks by the set of all the possible knowledge bases for the domain ontology model. To get the mathematical specification of an applied task belonging to a class of tasks corresponding the domain ontology it is necessary to replace all the variables of the class by values of input data and to enrich the domain ontology model by an appropriate knowledge base.

Finally, if domain terms in the mathematical specification of the class of applied tasks corresponding to a domain ontology are replaced by abstract designations then this mathematical specification of the class will be transformed into a mathematical task. The transformation of a mathematical specification of a class of applied tasks corresponding to a domain ontology into a mathematical task is important because different classes of applied tasks corresponding to ontologies of different domains, generally speaking, can be reduced to the same mathematical task.

If intellectual solver can solve mathematical tasks then it can be used for any domain which ontology model is isomorphic or equivalent to ontology model from mathematical task specification.

Let's consider a set of mathematical specifications of applied tasks such that every specification contains the same domain model. Such a set will be called an applied multitask. Just as an applied task was transformed into a class of applied tasks, the latter was transformed into a class of applied tasks corresponding to a domain ontology, and the latter was transformed into a mathematical task, so an applied multitask can be transformed into a class of applied multitasks, the latter can be transformed into a class of applied multitasks corresponding to a domain ontology, and the latter can be transformed into a mathematical multitask. An intellectual solver is intended for solving applied multitasks of a class of applied multitasks or for solving applied multitasks of a class of applied multitasks corresponding to a domain ontology.

The availability of simpler and more complex ontologies of the same domain can be important to develop intellectual solvers for specialists of different qualifications. As this takes place, working out methods for solving tasks based on a more simple ontology can be a simplification of methods for solving the corresponding tasks based on a more complex ontology. The same can also take place for ontologies of different domains.

The same methods often can be used for solving a few tasks and subtasks. Abstraction of applied tasks to mathematical ones gives a possibility of reusing methods for their solving. If different applied tasks can be reduced to the same mathematical task then a method for solving the mathematical task can be used for solving these applied tasks too. A decomposition of a mathematical task into mathematical subtasks in working out a method for solving the mathematical task gives an additional possibility for reusing methods. In this case, the same mathematical subtasks can be components of decompositions of different mathematical tasks and methods for solving these subtasks can be components of methods for solving different mathematical tasks.

Ontologies of complex domains are built from components, which are ontologies of other domains. The fact that an ontology of a complex domain is a composition of other domain ontologies can be used to work out methods for solving tasks in the complex domain. These tasks can be divided into subtasks corresponding to tasks for components of the ontology. If methods for solving these tasks have been already known, working out a method for solving the whole task may be considerably simplified.

Conclusions

In this article, general properties of relations among domain ontologies have been considered. Examples of these relations can be the relation between ontologies representing the same domain conceptualization, the relation of resemblance between ontologies, the relation “to be more simple or more complex” and the relation among an ontology consisting of components, which are other ontologies, and these components. A formalization of these relations has been suggested. This formalization preserves the properties above. These results show that the definitions of an ontology and its model given in [Kleshchev et al, 2000a] allow us to recognize these relations among ontologies. Relations among domain ontologies give a possibility to reuse one ontology model when another ontology models are worked out and when new intellectual computer system for same or different domain is worked out.

References

[van Heijst et al, 1996] van Heijst G., Schreiber A.T., and Wielinga B.J. Using Explicit Ontologies in KBS Development. In International Journal of Human and Computer Studies, 1996, 46 (2-3): 183-292.

[Gangemi et al, 1999] Gangemi A., Pisanelli D.M. and Steve G. An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies // In Data & knowledge Engineering, Vol. 31, N 2, 1999, pp. 183–220

[Gangemi et al, 1996] A. Gangemi, G.Steve, F. Giacomelli. ONIONS: An Ontological Methodology for Taxonomic Knowledge Integration. In P. van der Vet (ed.) Proceedings of the Workshop on Ontological Engineering, ECAI96, 1996.

[Guarino, 1998] Guarino N. Formal Ontology and Information systems. In Proceeding of International Conference on Formal Ontology in Information Systems (FOIS’98), N. Guarino (ed.), Trento, Italy, June 6-8, 1998. Amsterdam, IOS Press, pp. 3-15.

[Kleshchev et al, 2000a] Kleshchev A.S., Artemjeva I.L. Mathematical Models of Domain Ontologies. Technical Report 18-2000. Vladivostok, 2000. 43 p. (available in http://iacp.dvo.ru/es/)

[Kleshchev et al, 2000b] Kleshchev A.S., Artemjeva I.L. Unenriched Logical Relationship Systems. Technical Report 1-2000. Vladivostok, 2000. 43 p. (available in http://iacp.dvo.ru/es/)

[Laresgoiti et al] L. Laresgoiti, A. Anjewierden, A. Bernaras, J. Corera, A.Th.Schreiber, B.J.Wielinga. Ontologies as Vehicles for Reuse: a Mini–expirement. Available from http://ksi.cpsc.ucalgary.ca/KAW/KAW96/laresgoiti/k.html

[Schreiber et al] G.Schreiber, W.Jansweijer, B.Wielinga. Framework & Formalism for expressing Ontologies (version 2). Technical Report, University of Amsterdam, DO1b2. http://www.swi.psy.uva.nl/projects/NewKACTUS/Reports.html

[Sowa] Sowa J., Knowledge Representation: Logical, Philosophical and Computational Foundations. In http://www.bestweb.net/sowa/ontology/gloss.htm

[Studer et al, 1998] R Studer, V.R. Benjamins, D. Fensel. Knowledge Engineering: Principles and methods. In Data & Knowledge Engineering 25, 1998, p 161–197

[Wielinga et at, 1994] Wielinga, B., Schreiber A.T., Jansweijer W., Anjewierden A. and van Harmelen F. Framework and Formalism for Expressing Ontologies (version 1). ESPRIT Project 8145 KACTUS, Free University of Amsterdam deliverable, DO1b.1, 1994. http://www.swi.psy.uva.nl/projects/NewKACTUS/Reports.html

Author's Information

Alexander S. Kleshchev – kleschev@iacp.dvo.ru

Irene L. Artemjeva – artemeva@iacp.dvo.ru

Institute for Automation & Control Processes, Far Eastern Branch of the Russian Academy of Sciences

5 Radio Street, Vladivostok, Russia

METHODS OF ADAPTIVE EXTRACTION AND ANALYSIS OF KNOWLEDGE FOR KNOWLEDGE-BASE CONSTRUCTION AND FAST DECISION MAKING

Alexander Kuzemin, Darya Fastova, Igor Yanchevsky

Abstract: An approach for knowledge extraction from the information arriving to the knowledge base input and also new knowledge distribution over knowledge subsets already present in the knowledge base is developed. It is also necessary to realize the knowledge transform into parameters (data) of the model for the following decision-making on the given subset. It is assumed to realize the decision-making with the fuzzy sets' apparatus.

ACM Classification Keywords: I.2.5 - Expert Systems; I.2.6 - Knowledge acquisition

Introduction

The problem of knowledge representation in the process of the expert system (ES) design is the central one as the knowledge base (KB) main function implementation, i.e. new knowledge gaining, depends on its successful solution. Starting from this the structure and form of models and methods for knowledge representation making the decisive action on the ES efficiency and external information

In the majority experts' opinion the expert system power is defined by the volume of the knowledge the given system offers. Despite the fact that a lot of instrumental means helpful in gaining knowledge has appeared recently this problem still remains poorly defined and laborious one. Knowledge gaining is inseparably connected with the process of their check-out consisting in detection of insufficient knowledge and their introduction to the system, the KB check on non-inconsistency and completeness, check of the managing mechanism, the ES analysis and modification.

The process of compatible ES development implies creation of the specialized instrumental systems. Such systems support execution of the life cycle main stages, they commonly fix presentation of the used information, the knowledge presentation language, the knowledge interpreter (display) and a set of software instruments intended for a number of problems solution. However, these systems are oriented to the support of a user from the knowledge engineer class [2,3,4] and not information carriers (experts). Thus, they do not take into account a modern approach to creation of means for information processing (MIP), which consists in exclusion of a knowledge engineer, from this process as a redundant mediator.

By gaining knowledge is meant the process of gaining knowledge from experts or some other knowledge sources and their transmission to the ES whose efficiency completely depends on the gained knowledge quality and correctness of their presentation. The complexity is stipulated by a great volume of knowledge used by the expert, which are not always realized by him as such or formulated dimly.

In terms of the ES interaction with an expert and a knowledge engineer in the process of knowledge gaining it is possible to single out the following main phases:

1. The preliminary phase. Knowledge engineer obtains from an expert or some other knowledge source a general information about software (the main concepts, relations, sub-problems, data structure) and forms a general concept of the ES design principles on their basis. Starting from the aims of the ES being created the corresponding instruments is chosen: envelopes, empty ES, development medium, knowledge presentation language.
2. The initial phase. Knowledge engineer fills the system with knowledge about presentation, i.e. the values defining the organization, structure and means of the KB presentation.
3. The accumulation phase. Gaining of the main knowledge on software envisages the following:
   - detection of the knowledge incompleteness or discrepancy;
   - extraction of new knowledge controlling a definite discrepancy, incompleteness or incorrectness of the KB;
   - transformation of new knowledge into the form intelligible to the ES;
   - correction of the KB.

The process of gaining knowledge reduces to the following problems' solution:

1. definition of the necessity to modify or widen knowledge; if there is no such a necessity then the process of gaining knowledge then the process of gaining knowledge is finished;
2. extraction of new knowledge on software;
3. transformation of knowledge into the form intelligible to the ES;
4. modification of the knowledge system.

There exist several types of models of knowledge gaining (according to the degree of knowledge gaining process automation degree and degree of their independence from an expert):

1. Gaining knowledge only with the knowledge engineer help. At the early stages of work (interaction of the knowledge engineer with a PC) the above problems solution through their assimilation with the software expert help and the following development of the system only by the knowledge engineer is assumed. The shortages of this method are as follows:

- insufficient knowledge of software doesn't make it possible to ensure the completeness and consistency of the gained knowledge;
- absence of braking down of the system components to the KB and mechanism of input doesn't make it possible to preserve once achieved consistency at inevitable modifications of the system.

2. Gaining knowledge at the cost of organizing interaction of the expert and ES directly either with the knowledge engineer help at the expense of knowledge separation from the programs and knowledge presentation in the KB simple information structures. Such a model assumes that the first and the second problems of gaining knowledge are solved by the expert with the help of the knowledge engineer, the third problem is solved only by the knowledge engineer and the fourth one – by the ES. A great labor consuming can be assigned to shortages of the given approach as of four problems of knowledge gaining of this model only one problem is automated. Incorporation into ES of the intelligent editor possessing the dialog capabilities and an extensive meta-knowledge (knowledge of the KB structure) decreases significantly the labor consuming of knowledge gaining process and reduces to minimum participation of the knowledge engineer participation in this process. When using the intelligent editor, the expert solves the first and the second problems of knowledge gaining, the third and the fourth problems are performed with the ES. This model is widely used in practice.

3. The complete automation of knowledge gaining four problems based on application of the inductive generalizing programs. It is assumed that in the KS the concrete facts on software are kept in an explicit form. The Inductive program analyses the input data from some field of the expertise and forms automatically meaningful deviations and rules describing software. The shortage of the model is that the inductive programs are capable to generate knowledge only on the basis of the structured rules.

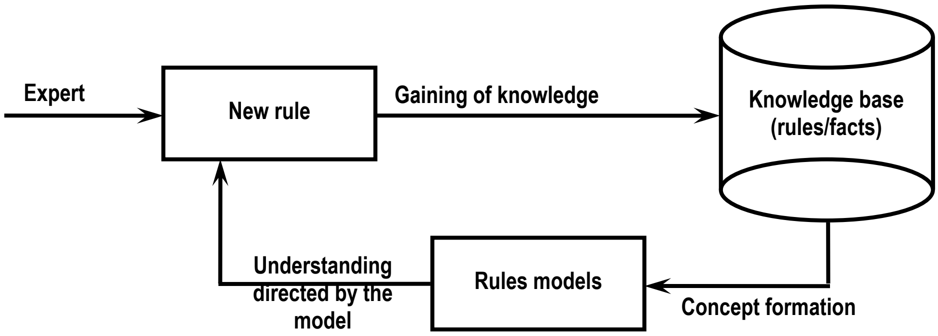


Fig. 1 – Rules derivation

Thus, to automate the knowledge gaining process the **problem of gaining knowledge from the expert should be solved** with the aim to develop further the means for inductive rules generation for the concrete software.

To solve such a problem it is expedient to define a priori some subset of information specifying the initial features to software. This assists software partitioning into parts with the aim of the expert's interrogation context directional narrowing.

ES successively inquires of the expert for examples of the object falling into every class called by him. Every example is analyzed for the elucidation of the following questions:

- attributes of the object in the example;
- values of these attributes in other objects belonging to the identical class;
- manifestation of these values in the objects of other classes.

The process of software rules derivation joins understanding directed by the model and the concept formation. The text of the new rule performs here the role of the signal, the program of knowledge derivation stands here for the signal processor; the new rule expressed in an internal representation is the interpretation and the model is the rules model.

Gaining of knowledge by this scheme represents the process with a feedback: the existing models of the rules acts on the process of new rules derivation and the new rules added to the KB stipulate the rules models re-calculation.

The main idea is based on the assumption that knowledge of software can be described at some generalized level in the form of the typical rules (rules models) which define the course of the knowledge gaining process. The set of the rules models is the KB model. The rules models are formed by the system automatically on the basis of the KB analysis, i.e. the system uses the rules models for limitation of the input message interpretation and the initial rule is used then for the rules model reshaping.

The availability of the rules models allows the system not only to derive new rules but also to give advice for modification of the rules introduced by the expert. Advice is given on the basis of comparison of the new rule with a model corresponding to it. If incomplete but only a partial comparison takes place then the system offers the expert to refine the rule according to those aspects of the model which turned out to be not represented in the new rule.

As the models of the rules are constructed by the system itself on the basis of the current content of KB a number of important conclusions is observed:

1. the models are created automatically and the expert doesn't take care of their creation and even doesn't know about their presence;
2. the KB models (i.e. a set of rules models) accumulate the whole previous experience (previously introduced models are used for current models shaping);

3. the KB model evolves with the KB growth and represents variations in the base content due to modification of the rules' model when inputting every rule.

A set of all rules' models can be regarded as an abstract image of KB. The system when analyzing the KB model specifies content and limits of the accumulated knowledge. For every model, it is possible to introduce the measure characterizing this model power. It is considered that the higher the model power the higher its definiteness index and the greater number of rules serving the basis for this model creation. A small power value for some model can indirectly point to insufficiency of the knowledge on some fragments of software that should be taken into account when adding to the ES rules models.

The shortages of the offered model:

- the presence of some minimum volume of information in the KB (critical mass of knowledge), it is impossible to ensure the acceptable context for gaining knowledge without it;
- derivation only one rule at the current moment;
- ES functioning is adequate if a number of detected errors is small or their number is great but they depend only slightly on each other.

Prospects of creation of the ES assisting in the process of decision-making depends on such a source as the capability for understanding and reasoning through analogy. The key to the analogy realization with a computer consists in presentation of a new information in a structured form.

From the knowledge organization standpoint, it is expedient to classify knowledge in two aspects: by the concept levels and activity levels. Each of the aspects forms their own hierarchy.

To ensure that the ES could control the solution search process, gain new knowledge it should be able not only to use its knowledge but also to understand and study them. If knowledge on software is related to the zero concept level then meta-knowledge will get to the first level. This first level contains knowledge about the means of zero level knowledge presentation. It is precisely meta-knowledge which play an essential role in consultation control, the ES actions explanation and knowledge gaining as meta-knowledge are invariant to software.

It is possible to single out the second level of the concept containing knowledge about meta-knowledge i.e. knowledge about presentation of the fundamental concepts of the first level etc. Separation of knowledge by the levels of representation widens the ES application field.

The number of comprehension levels in many instances depends on the specificity of the problems being solved, knowledge volume and means of representation. As a rule no less than three comprehension levels are singled out, they represent the general, logical and physical organization of knowledge. Introduction of several comprehension levels ensures an additional degree of ES flexibility as it makes it possible to perform changes at one level without referring to another ones. Variations at one comprehension level can only result in variations at the same level necessary to ensure matching of data and program structures.

Separation of the comprehension levels for considering knowledge with different degree of detail.

One of the main problems in terms of mathematical simulation in the process of creating the information systems for situational management is the knowledge presentation means, the decision is made in the concrete situation on its basis. Difficulty in knowledge presentation consists in realization of the information fragments transition into terms of the databases (DB) and knowledge bases (KB) structures. In terms of facts and processes resulting in changes in DB one should consider semantics and syntax of such presentation. By syntax is meant a set of rules for joining symbols into logically concrete expressions, and by semantics is meant the method of interpretation of expressions formed as a result of the concrete realizations of syntax rules.

Urgency of the problem is stipulated by the necessity of the information intelligent system development on the BD and KB basis which will be able to shape and add the database and knowledge base itself and perform logical derivation for the further decision-making on subsets.

Aim of the work consists in reduction of the time needed for decision-making on some definite subset at the cost of derivation and analysis of a new knowledge arriving to the KB input.

To gain the specified aim it is necessary to solve the **following problems:**

– to realize predicate inquiries construction and their modification which are the apparatus for description and investigation of the processes of updating and modification of databases and knowledge bases,

- to define rules of logical derivation based on databases an knowledge bases,
- to calculate the shortest way between knowledge needed for referring the new knowledge to the nearest subset.

By the database is meant a set of facts. The main ideas of such an approach are considered in the frameworks of the SQL concrete realizations or WWW and WEB realizations.

But in the above realizations the function of inquiries logical substantiation is displaced to the database user, in Prolog-program the knowledge base construction by the engineer is envisaged and constant support of the user during the logical derivation session is applied.

1. KB description

All decisions in the universe of discourse are made on the basis of experienced experts' analysis conclusions. The information system knowledge base is considered according to [1,2] as a set of information essences of atomic predicates from some information space $\Re$. All changes taking place in the knowledge base are considered as a result of the modifying predicate inquiries Q_m . The basis of the predicate inquiries taken alone is a set of the modifying predicate rules:

$$Q_m \leftrightarrow (K_B)^{<<} \left\| \begin{matrix} K_{B-(X)} \\ K_{B+(X)} \end{matrix} \right. << \qquad (1)$$

where $X \in \Re$, $K_{B+(X)}$ means that the atomic predicate o should be incorporated into the knowledge base K_B ; K_B means that X should be eliminated from the knowledge base; $(K_B)^{<<}$ means the knowledge base modification at the level of the predicate rules logical coherence as a consequence of the rules incorporation and elimination operations use; $K_{B\pm(X)}$ means the possibility of modification not only of the knowledge base but security of the user as well on the basis of descriptors; $<<$ is considered as a complex arrow, its features are studied with the categories theory.

2. Knowledge Gaining

Knowledge can be presented as production rules of the type [3]:

«if $X_1 \& ... \& X_K$, then $X_{K+1} \& ... \& X_{K+L}$ »,

where $X_1 ... X_K$, $X_{K+1} ... X_{K+L}$ - some predicates.

Definition 1. The set $W = \Pi_1 \times \Pi_2 \times ... \times \Pi_{K+L}$ is called the contents of the knowledge «if $X_1 \& ... \& X_K$, then $X_{K+1} \& ... \& X_{K+L}$ ». The random element of this set is called the element of the knowledge contents.

The set $W_1 = \Pi_1 \times \Pi_2 \times ... \times \Pi_K$ is called the contents of knowledge condition. The random element of this set is called the element of the knowledge condition contents.

The set $W_2 = \Pi_{K+1} \times \Pi_{K+2} \times ... \times \Pi_L$ is called the contents of the knowledge result. The random element of this set is called the element of the knowledge result.

Definition 2. By the probability p_i of the knowledge contents element $w_i \in W$ is meant the event probability, which consists in that all predicate constants being a part of w_i will take the logical meaning «and» when substituting the objects from the truth domain of variable predicates making part of this knowledge instead of arguments.

The element of the knowledge content w_j represents a vector, predicates variables making part of this knowledge are its components. Vector $z_j = z_j(1) ,..., z_j(K+L)$ from R^{K+L} can be put in conformity with the knowledge content element w_j .

The function of knowledge distribution is the function of $K+L$ arguments: $F(y)=F(y(1),y(2),...,y(K+L))$, with the domain of definition R^{K+L} and assuming the values in the space R^1. It is defined by the formula $F(y)=\sum_{z_j \le y} p_j$, where z_j - image of the knowledge contents element w_j in R^{K+L}. Expression $z_j < y$ is understood as a fulfilling of the conditions: $z_j(i) < y(i),\ i=1,\ ...,\ K+L$.

Definition 3. By distance between the comparable knowledge 3H1 and 3H2 is meant Hellinguer distance $d(G,\ Q)$ between two probability distributions of their contents elements $G=\{p_{11},\ p_{12},\ ...,\ p_{1r},\ ...\}$ and $Q=\{p_{21},\ p_{22},...,p_{2r},...\}$, which is calculated by the following formula:

$$d(G,Q)=\sum_j (\sqrt{p_{1j}}-\sqrt{p_{2j}})^2 \qquad (2)$$

Calculating the distance between knowledge, the problem of a new gained knowledge distribution is solved. The distance between a new knowledge and all knowledge in the knowledge base and then a new knowledge is placed into the subset containing such a knowledge for which the distance assumes the least value.

3. Decision-making

Despite the fact that decision-making is realized in the chosen knowledge subset for the complex systems and processes the adequate mathematical description of decision-making is absent or represent rather cumbersome mathematical constructions whose optimization and practical use in the real time are impossible. This problem can be solved with using algorithms built with the models simulating the process of decision-making by an experienced expert. The theory of fuzzy sets can be used as a mathematical apparatus for a great number of decision-making models. When choosing decisions in the situation centers (SC) the aim of the designing stages consists in the choice of the design version or the parameter value from the rather small specified set defined, as it was mentioned above, with the formula (2). To simulate the decision-making process it is offered to use the decision-making models based on the fuzzy rule modus ponens, fuzzy inductive scheme of derivation and fuzzy expert information of the second type. In this case, the following inductive output scheme will be used [4]:

$$\frac{\tilde{L}^{(2)}=\begin{cases}\langle If\ \ \tilde{B}_1\ ,then\ \tilde{A}_1\rangle ; \\ \langle If\ \ \tilde{B}_2\ ,then\ \tilde{A}_2\rangle ; \\ \dots\dots\dots\dots \\ \langle If\ \ \tilde{B}_m\ ,then\ \tilde{A}_m\rangle ; \\ A'-true\end{cases}}{B'-true} \qquad (3)$$

Here the clear statements A' and B' have the following form:

$$A' : \langle \beta_W\ is\ \ w' \rangle\ ;\quad B' : \langle \beta_V\ is\ \ v' \rangle$$

$$w'=(x,\ y,\ z...)\in X\times Y\times Z\times ...,\ \ v'\in V\ .$$

In the given derivation scheme the statements concerning the values of the input parameters are the premise for the scheme itself (statement A') and the result inside the system $\tilde{L}^{(2)}$ of the statement (statement A_j). The statements about the values of the output parameters are the result for the derivation scheme (3) (statement B') but it is the premise inside the scheme $\tilde{L}^{(2)}$ (statement $\tilde{B}_j$). That is why to choose the values of the output parameter V based on the rule modus ponens it is necessary to transform the derivation scheme (3) to the form:

$$\frac{\begin{matrix}\tilde{L}^{(1)} \\ A' - true\end{matrix}}{B' - true}.$$

To do this it is offered to transform the second type statements system into the first type system equivalent to it using the contraposition rule according to which the statements “IF A, THEN B” and “IF¬B, THEN¬A” are equivalent for the random expressions, i.e.

$$\langle If\ A\ then\ B\rangle \equiv \langle If\ \neg B, then\ \neg A\rangle,$$

Here expressions $\neg A$ and $\neg B$ are negative expressions A and B.

Applying the contraposition rule to the expressions $\tilde{L}_j^{(2)}, j = 1,...,m$ of the second type system we will obtain

$$\langle If\ \tilde{B}_j\ then\ \tilde{A}_j\rangle \equiv \langle If\ \neg\tilde{A}_j, then\ \neg\tilde{B}_j\rangle,$$

where the statements $\neg\tilde{A}$ and $\neg\tilde{B}_j$ can be regarded as statements $\langle \beta_W\ is\ \alpha_k\rangle$ and $\langle \beta_V\ is\ \alpha_l\rangle$, $k = W_j^*$ and $l = V_j^*$ where the values α_k and α_l are defined by the belonging functions μ_k and μ_l, being the complement to μ_n and μ_p, where $n = W_j$ and $p = V_j$:

$$\mu_k(w) = 1 - \mu_n(w), \forall w \in W = X \times Y \times Z \times ...$$

$$\mu_l(v) = 1 - \mu_p(v), \forall v \in V.$$

Conclusions

Scientific novelty of the developed approach consists in that the information intelligent system based on DB and KB makes it possible to shape and complement the database and knowledge base. After that, it in its own transforms the analyzed knowledge into parameters of the model for making decisions on a definite subset.

Practical value – in the average the given approach shortens the time needed for making decision on the subsets for 20%. This result was obtained at the cost of using Hellinguer distance at the stage of new knowledge comparison and classification.

Bibliography

1. Sheketa V.I. Modification Predicate Applications // Programming Problems. Institute of Program Systems of NAS of Ukraine. – 2004. - №2-3. – 339-343pp. //Special Issue. Proceedings of the 4-th SPC “UkrProg2004”, July 1-3, 2004.– Kyiv, Cybernetics Center of NAS of Ukraine.
2. Sheketa V.I. Predicate Queries Modification, as a Tool to Work with the Knowledge Bases of Oil and Gas Subject Domain //Proceedings of 6-th International Scientific Conference “Modern Problems of Radio Engineering, Telecommunications and Computer Science – TCSET’2004”. – Lviv-Slavsko. – 2004. – February 24-28. – P.315-319.
3. V.N. Koval, Yu.V. Kuk. Knowledge Gaining and Analysis //Artificial Intelligence. – 2004. - №3 – 293-304pp.
4. N.G.. Malyshev, L.S. Bershtein, A.V. Bozheniuk. Fuzzy Models for Expert Systems in SAPR. M.: Energoatomizdat, 1991, 136 p.

Authors' Information

Kuzemin Alexander Yakovlevich, Cand. Techn. Sciences, Prof. Department of Information Sciences, Kharkov National University of Radio Electronics. Address: KNURE, 14, Lenin Ave., Kharkov. Tel.: (057) 702-15-15. Scientific interests: systematic approach in designing and programming

Fastova Darya Vladimirivna, post-graduate, Department of Information Sciences Kharkov National University of Radio Electronics. Address: KNURE, 14, Lenin Ave., Kharkov. Tel.: 702-15-15. Scientific interests: information systems design, systems analysis.

Yanchevsky Igor Vladimirovich, General director of JSC PJ “CT and SA”, competitor of the Department of Information Sciences, Kharkov National University of Radio Electronics.

TABLE OF CONTENTS OF VOLUME 12, NUMBER 1