INFORMATION SCIENCE & COMPUTING

International Book Series Number 13

Intelligent Information and Engineering Systems

Supplement to
International Journal "Information Technologies and Knowledge" Volume 3 / 2009

I T H E A

RZESZOW – SOFIA, 2009

Galina Setlak, Krassimir Markov (ed.)

Intelligent Information and Engineering Systems

International Book Series "INFORMATION SCIENCE & COMPUTING", Number 13

Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 3 / 2009

Institute of Information Theories and Applications FOI ITHEA

Rzeszow University of Technology, Poland

Rzeszow, Poland - Sofia, Bulgaria, 2009

This issue contains a collection of papers that concern the problems of intelligent and engineering systems.

Papers in this issue are selected from the International Conference INFOS-2 2009, Krynica, Poland, a part of the Joint International Events of Informatics "ITA 2009", Autumn Session.

International Book Series "INFORMATION SCIENCE & COMPUTING", Number 13
Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 3, 2009

Edited by **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria,
and **Rzeszow University of Technology, Poland,** in collaboration with
- **V.M.Glushkov Institute of Cybernetics of NAS**, Ukraine,
- **Institute of Mathematics and Informatics, BAS**, Bulgaria,
- **Institute of Information Technologies, BAS**, Bulgaria.

Publishers: Institute of Information Theories and Applications FOI ITHEA, Sofia, 1000, P.O.B. 775, Bulgaria.
Rzeszow University of Technology, Poland
Издатели: Институт по информационни теории и приложения ФОИ ИТЕА, София, 1000, п.к. 775, България
Ржешов Технологичен Университет, Полша.
www.ithea.org, www.foibg.com, e-mail: info@foibg.com

General Sponsor: **Consortium FOI Bulgaria** (www.foibg.com).

Printed in Poland and Bulgaria

Copyright © 2009 All rights reserved
© 2009 Institute of Information Theories and Applications FOI ITHEA - Publisher
© 2009 Rzeszow University of Technology, Poland - Publisher
© 2009 Galina Setlak, Krassimir Markov – Editors
© 2009 For all authors in the issue.

ISSN 1313-0455 (printed) **ISSN 1313-048X (online)** **ISSN 1313-0501 (CD/DVD)**

PREFACE

The scope of the International Book Series “Information Science and Computing” (**IBS ISC**) covers the area of Informatics and Computer Science. It is aimed to support growing collaboration between scientists from all over the world. IBS ISC is official publisher of the works of the members of the ITHEA International Scientific Society.

The official languages of the IBS ISC are English and Russian.

IBS ISC welcomes scientific papers and books connected with any information theory or its application.

IBS ISC rules for preparing the manuscripts are compulsory.

The rules for the papers and books for IBS ISC are given on www.foibg.com/ibsisc .

The camera-ready copyes of the papers should be received by ITHEA Submition System http://ita.ithea.org .

The camera-ready copyes of the books should be received by e-mail: info@foibg.com .

Responsibility for papers and books published in IBS ISC belongs to authors.

This issue contains a collection of papers that concern the problems of intelligent information and engineering systems. Papers are peer reviewed and are selected from the International Conference INFOS-2, Krynica, Poland, a part of the Joint International Events of Informatics "ITA 2009" – autumn session.

ITA 2009 has been organized by

ITHEA International Scientific Society

in collaboration with:

- Rzeszow University of Technology (Poland)
- Institute of Information Theories and Applications FOI ITHEA
- Dorodnicyn Computing Centre of the Russian Academy of Sciences
- International Journal “Information Theories and Applications”
- International Journal "Information Technologies and Knowledge"
- Association of Developers and Users of Intelligent Systems (Ukraine)
- Association for Development of the Information Society (Bulgaria)
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Information Technologies, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)
- Taras Shevchenko National University of Kiev (Ukraine)
- Universidad Politecnica de Madrid (Spain)
- BenGurion University (Israel)
- University of Calgary (Canada)
- University of Hasselt (Belgium)
- Kharkiv National University of Radio Electronics (Ukraine)
- Astrakhan State Technical University (Russia)
- Varna Free University “Chernorizets Hrabar” (Bulgaria)
- National Laboratory of Computer Virology, BAS (Bulgaria)
- Uzhgorod National University (Ukraine)

The main ITA 2009 events were:

KDS	XVth International Conference "Knowledge - Dialogue – Solution"
i.Tech	Seventh International Conference "Information Research and Applications"
MeL	Fourth International Conference "Modern (e-) Learning"
INFOS	Second International Conference "Intelligent Information and Engineering Systems"
CFDM	International Conference "Classification, Forecasting, Data Mining"
GIT	Seventh International Workshop on General Information Theory
ISSI	Third International Summer School on Informatics

More information about ITA 2009 International Conferences is given at the www.ithea.org .

The great success of ITHEA International Journals, International Book Series and International Conferences belongs to the whole of the ITHEA International Scientific Society.

We express our thanks to all authors, editors and collaborators who had developed and supported the International Book Series "Information Science and Computing".

General Sponsor of IBS ISC is the **Consortium FOI Bulgaria** (www.foibg.com).

Rzeszow-Sofia, September 2009 *G.Setlak, Kr. Markov*

TABLE OF CONTENTS

Industrial Control and Monitoring

Intelligent Agents and Multi-Agent Systems

Intelligent Manufacturing Systems

Business Intelligence Systems

Medical and Diagnostic Systems

Computer Engineering, Mechanical Engineering

INDEX OF AUTHORS

General Information Theory

THEORY OF INFOS

Krassimir Markov, Krassimira Ivanova, Ilia Mitov

Abstract: *Theory of Infos is a part of the General Information Theory. Main features of the Theory of Infos are outlined in the paper. Further investigation is discussed.*

Keywords: *General Information Theory, Theory of Infos.*

ACM Classification Keywords: *A.1 Introductory and survey*

Introduction

The genesis of the concept of **Infos** started from the understanding that the concept **"Information Subject"** is perceived as human characteristic. It is clear that in the nature there exist many creatures which may be classified to this category. To exclude the misunderstandings we decide to introduce new word to denote all possessors of the characteristics of the Information Subject.

This word is **"*Infos*"**.

There exist several common theoretical information paradigms in the Information Sciences. May be, the most popular is the approach based on the generalization of the Shannon's Information Theory [Shannon, 1949], [Lu, 1999]. Another approach is the attempt to be synthesized in a common structure the existing mathematical theories, which are applicable for explanation of the information phenomena [Cooman et al, 1995].

Besides of this, we need to point the diligence of the many researchers to give formal or not formal definitions of the concept "information". Unfortunately, although they are quite attractive in some cases, these definitions did not bring to any serious theoretical results [Abdeev, 1994], [Bangov, 1995], [Markov P., 2002], [Tomov, 1991], [Elstner, 1993].

The development of the General Information Theory (GIT) had started in the period 1977-1980. The first publication, which represents some elements of GIT, was published in 1984 [Markov, 1984]. In 1988, the not formal definition of the concept of Information was published in [Markov, 1988]. It became as a fundamental definition for the GIT [Markov et al, 1993], [Markov et al, 2003a]. The translation of the philosophical theory into the formal one is a good approach for verification of the scientific ideas [Markov et al, 2003b], [Markov et al, 2004].

The General Information Theory is based only on primary consideration of the world as variety of entities, which are formed by relationships between entities that form lower levels. The fundamental notion of the GIT is the concept "Information". All other concepts are defined based on this definition.

The GIT is build by three specialized theories:

- Theory of Information,
- Theory of Infos,
- Theory of Inforaction.

This article presents the main features of the Theory of Infos.

Entity

In our examination, we consider *the real world* as a space of *entities.* The entities are built by other entities, connected with *relationships*. The entities and relationships between them form the internal *structure* of the entity they build. To create the entity of a certain structural level of the world, it is necessary to have:

- the entities of the lower structural level;
- established forming relationship.

The entity can dialectically be considered as a relationship between its entities of all internal structural levels.

The forming relationship has a representative significance for the entity. The destruction of this essential relationship causes its disintegration. The establishment of forming relationship between already existing entities has a determine significance for the emerging of the new entity.

The forming relationship is the reason for *the emergence* of individual properties, which distinguish the new entity from the forming ones. ***The relationships form and present the entity***.

Impact, Interaction, Reflection

Building the relationship between the entities is a result of the ***contact*** among them. During the contact, one entity ***impacts*** on the other entity and vice versa. In some cases the opposite impact may not exist, but, in general, the contact may be considered as two mutually opposite impacts which occur in the same time.

The set of contacts between entities forms their ***interaction***. The interaction is a specific ***interactive relationship*** between entities which take part in it.

The contacts of the given structural level are processes of interaction of the entities on the lower levels.

During the establishing of the contact, the impact of an entity changes temporally or permanently the internal structure of the impacted entity. In other words, the realization of the relationships between entities changes, temporary or permanently, their internal structure at one or at few levels.

The internal change in the entity, which is due to impact of the other entity we denote with the notion "***direct reflection***".

Every entity has its own level of sensibility. This means that the internal changes occur when the external influence is over the boundary of the sensibility of the entity.

The ***"reflection impulse"*** for given entity is the amount of the external influence needed for transition from one state to the reflection one.

The entities of the world interact continuously. It is possible, after one interaction may be realized another. In this case, the changes received by any entity, during the first interaction, may be reflected by the new entity.

This means the ***secondary (transitive external) reflection*** exists.

The chain of the transitive reflections is not limited. In general, the concept “transitive impact” (respectively “transitive reflection”) of the first entity on the third entity through the second one will denote every chain of impacts (reflections) which start from first entity and ends in the third entity, and include the second entity in any internal place of the chain.

One special case is the ***external transitive self-reflection*** where the entity reflects its own relationships as a secondary reflection during any external interaction.

Some entities have an opportunity of ***internal self-reflection***. The internal self-reflection is possible only for very high levels of organization of the entities, i.e. for entities with very large and complicated structure. The self-reflection (self-change) of the entity leads to the creating of new relationships and corresponding entities in it.

Of course, the internal self-reflection is a result of the interaction provided between entities in the lower levels of the structure of the entity. Such kind of entities has relatively free sub-entities with own behavior in the frame of self-preservation of the whole entity. As a result of the self-reflection, some relationships and corresponding sub-entities are created or changed in the entity.

The combination of the internal and external self-reflection is possible.

Finally let remark that the reflection could not be detected by the entity that contains it. This is dialectical behavior of the reflection - it is only an internal change caused by the interaction.

Information

The real world contains unlimited number of entities. When an entity contacts another, there exists a great possibility to join third entity in this process. It is clear; the third entity may contact and reflect each of others as well as the process of realization of the interaction between them — the process of realization of the contact is a specific (temporal) forming relationship between entities and during the process of establishing the contact the entities form new (temporal) entity which in the same moment may be reflected by the third entity. So, the third entity may reflect any vestiges of this interaction from both first and second entities.

In the special case when the third entity contains reflections of the first entity received by both two different ways:

1. by transitive impact of the first entity on the third one through the second entity,
2. by impact of the first entity on the third one which is different from the transitive one, i.e. it can be direct impact or transitive impact through another entity (-ies)

then the third entity became as an external relationship between first entity and its reflection in the second entity – it became as ***"reflection evidence"*** of this relationship.

The first entity is called ***reflection source;*** the second entity is called ***reflection recipient***; and the third entity is called ***reflection evidence***.

In this special case, when there exist the triple

"(source, recipient: evidence)",

the reflection of the first entity in the second is called ***information*** in the second for the first entity.

Let point one very important case of the real world - simultaneous contacts of the three entities. Every one of them may be source, recipient and evidence in the same time. There exist six cases which represent the simultaneous contacts of three entities. Therefore, the entities A, B and C may be in the next six reflection relations: (A, B: C); (B, C: A); (C, A: B); (A, C: B); (C, B: A); (B, A: C).

All reflection relations are equivalent from point of view of the interrelations between reflection source, reflection recipient and reflection evidence. Because of this we will discuss only the case (A, B: C).

For practical needs, it is more convenient to follow the next consideration.

The reflection in the recipient represents both the relationships and the sub-entities of the source. From other point of view, the relationships build up and present the entities. Because of this, the reflected relationships are the essence of the reflection. In other words, iff there exist reflection evidence than the reflection of the forming relationship may be considered as "information" for reflected entity. Therefore, in the sense that the evidence exists to point what relationship (between what entities) is reflected and where it is done, we may say

"The information is reflected relationship".

So, the *reflection* of the first entity in the second one is "***information"*** for the first entity if there is corresponded *reflection evidence*. The generalization of this idea leads to assertion that ***every reflection can be considered as information, iff there exists corresponding reflection evidence.***

Activity and Information Expectation

Every forming relationship as well as every relationship unites the entities and this way it satisfies some theirs possibilities for building the relationship by establishing the contact. In other words, for creating the forming relationship we need:

- entities, from which the new entity is able to built;
- possibilities of the entities for establishing the contact by satisfying of which the forming relationship may be originated.

The forming relationship is the aggregate of the satisfied possibilities for establishing the contact.

It is clear that after establishing the relationship we may have any of two cases:

1. all possibilities of the entities for establishing the contact are satisfied by such possibilities of other entities;
2. there are any free possibilities after finishing the establishment of the new relationship - on the low levels of the entity or, if it is a new entity, on the level of the whole entity. Disintegration of the whole entity or any its part may generate any possibilities too.

In the second case, the entity has any ***"free valences"***, which needs to be satisfied by corresponded contacts with other entities. We may say the entity has ***activity*** generated by the free possibilities for establishing the contacts with the entities from the environment.

The process of interaction is satisfying the possibilities for contact of the entities. From point of view of the entity, the interaction may be external or internal.

During the interaction given entity may be destroyed partially or entirely and only several but not all parts of the destroyed entity may be integrated in the new entity. This means that there exist both constructive and destructive processes in the process of interaction between entities. The determination of the type of the interaction depends on the point of view of given entity. The interaction dialectically contains constructive and destructive sub-processes.

If the entity is a complex, it is possible for it to have an opportunity of self-reflection. In such case, it is able to reflect any reflection, which has been already reflected in it. In this case, because of the new internal changes (self-reflection) the entity may obtain any new ***"secondary activity"***.

The secondary activity is closely connected to the structural level of the entity, which correspond to the level of the self-reflection. This way the secondary activity may be satisfied by internal or external entity from point of view of the given entity. In other words, the ***resolving*** of the secondary activity may be ***internal*** or ***external***.

During the establishment of the information relationship it is possible to be generated any secondary free activity (possibilities on the low levels of the entity or on the level of the whole entity) which needs to be satisfied by corresponded contacts with other entities.

The secondary activity generated by the information relationship is called ***"information activity".***

On given level of complexity of the entities a new quality becomes — the existence of self-reflection and internal activity based on the main possibilities for contact of the sub-entities as well as based on the new (secondary) possibilities created after internal self-reflection.

The internal activity may be resolved by:

- the internal changes which lead to partial internal disintegration of the sub-entities and theirs a posterior internal integration in the new structures;
- the external influence on the environment.

The internal changes may lead to removing of some sub-entities if they have no possibilities for integration with the others, i.e. if they have no free valences to be resolved in the process of integration.

The external influence is the most important. The impact on the entities around the entity is the way to resolve its activity. The destroying of the external entities and including the appropriate theirs parts in itself is the main means to exist and satisfy the free valences.

One special kind of activity is the information one. We assume that the secondary activity needs to be resolved by relevant to the information valences corresponded opposite (information) valences which need to be of the same genesis, i.e. generated by any information relationship. So, not every entity may be used for resolving the secondary activity.

This way, the entity expects a special kind of (information) contacts and (information) interaction for resolving the information activity. Because of this the information activity is called ***"information expectation".***

The Information Witness

Let remember the special case from above when the third entity contains reflections of the first entity received by both two different ways:

1. by transitive impact of the first entity on the third one through the second entity,
2. by impact of the first entity on the third one which is different from the transitive one, i.e. it can be direct impact or transitive impact through another entity (-ies).

In this case the third entity became as an external relationship between first entity and its reflection in the second entity — it became as ***"reflection evidence"*** of this relationship.

In addition, if during establishing the information relationship ***i*** = (source, recipient: evidence) in the reflection evidence is generated information expectation (activity) it is called ***"information witness"*.**

As the information witness is more complex entity so the information relationship may be more complex. In addition, let remark that the complex reflection is time-depended process. In other hand, the memory and actual context determine the result of the complex reflection.

The Information is a Model

As Marx Wartofsky remarks, the concept "**model**" has been used for denotation of the very large class of phenomena: mechanical, theoretical, linguistic, etc. constructions. He gave a good definition of the model relation and made clear the main characteristics of the model [Wartofsky, 1979]. This definition is as follow:

The model relation is triple M:

$$M: (S, x, y)$$

where "S" is subject for whom "x" represents "y". In other words only in this relation and only for the subject "S" the entity "x" is a model of the entity "y".

As we point above, the interaction between two entities is a specific theirs relationship. If there exist information witness (**W**) of the interaction between two entities as well as of the existence of the information about the first entity in the second entity, **W** became as subject for whom the information in the second entity represents the first one. In other words, there exists relation

$$M: (\mathbf{W}_{BA}, I_{BA}, A),$$

where "A" and "B" are entities, and the $\mathbf{W}_{BA}$ is the information witness, which proofs that the assertion "$I_{BA} \subset B$ is information in B for A" is true.

In the relation ($\mathbf{W}_{BA}$, I_{BA}, A) the information I_{BA} is a model of A**.**

The Information Model

The entities of the world interact continuously in the time. It is possible, after any interaction one another may be realized. In this case, the changes received by any entity, during the first interaction, may be reflected by the new entity. This means the ***secondary (transitive, external) reflection*** exists. The chain of the transitive reflections is not limited.

Let A, B and C are entities; A and B interact and after that B interacts with C. If there exist the relations:

- M_{BA}: ($\mathbf{W}_{BA}$, I_{BA}, A), where $\mathbf{W}_{BA}$ is the information witness, which proofs that the assertion "$I_{BA} \subset B$ is information in B for A" is true,
- M_{CB}: ($\mathbf{W}_{CB}$, I_{CB}, B), where $\mathbf{W}_{CB}$ is the information witness, which proofs that the assertion "$I_{CB} \subset C$ is information in C for B" is true,

and if $M_{C(B)A}$: ($\mathbf{W}_{C(B)A}$, $I_{C(B)A}$, A), where $\mathbf{W}_{C(B)A}$ is the information witness, which proofs that the assertion "$I_{C(B)A} \subset C$ is information in C for information in B for A" is true.

In such case, from point of view of the $\mathbf{W}_{C(B)A}$ the information $I_{C(B)A}$ is a model of A. In other hand, because of transitive reflection, $I_{C(B)A}$ is created as reflection of the I_{BA} but not directly of A.

This means that **$I_{C(B)A}$ is a model of the information in B for A**.

In other words the $I_{C(B)A}$ is an ***information model*** in C for A**.**

The collecting of information models for given entity in one resulting entity may exist as a result of the process of interaction between entities. Such process is in the base of the ***Information modeling***.

If an information model **IM** contains information for (reflected from) the two source information models $\mathbf{IM}_1$ and $\mathbf{IM}_2$ than the source information models are "***similar***" in the sense of the model **IM**.

The similarity of the information models causes the establishing the relation of aggregation between them.

The relation of similarity aggregates the similar models in new ***internally determined information model*** in the memory of the information witness.

The aggregation may cause the generating the new information activity, which may be resolved not only in the environment around the information witness. The possibility of self-reflection may cause the generating the new information models in his memory without any external influence and so on.

This process of aggregation and generation of new models is not limited.

The (information) models internally generated via self-reflection are called "***mental (information) models***" of the information witness.

Resolving the Information Expectation

Because of the existing of the information expectation, i.e. the existing of the secondary information activity, the Information Witness "expects" to combine the information valences with any others.

The combining the valences of the information expectation with some others is called ***resolving the information expectation***.

Let "n" is the number of free valences in an information expectation. After the contact some of them are combined as well as the others are not. The new valences, which are generated by the contact, do not belong to the information expectation before contact. They may form new information expectation but the basis for our reasoning will be the starting information expectation.

The normalized by "*n*" number *D'* of the not combined valences is called ***degree of discrepancy*** (**D**) of the incoming reflection to the information expectation, i.e.

$$\mathbf{D} = \frac{D'}{n}$$

The normalized by "*n*" number *C*' of the combined valences is called ***degree of combining*** (**C**) of the incoming reflection to the information expectation, i.e.

$$\mathbf{C} = \frac{C'}{n}$$

There exists the equation: **C + D = 1.**

From point of view of given expectation for contact the number of free valences is fixed. After the contact, as a result of reflection, some of the free valences of the entity may be combined with any new (internal or external) valences. Of course, new free valences may occur. The number "*n*" varies in the process of interaction. Every contact may change it.

The more valences of the information expectation have been resolved, the more qualitative is the incoming information and vice versa.

The difference **A** between normalized number **C** of resolved valences and normalized number **D** of not resolved valences of the information expectation is called ***adequacy of the reflection to the information expectation***, i.e.

$$A = C - D$$

It is easy to see that the values of adequacy A are in the interval [-1,1].

The Infos

The resolving of the information activity is ***a goal*** of the information witness.

This goal may be achieved by the establishment and providing (information) contacts and interaction.

The entity, which has possibility for:

- ***(primary) activity*** for external interaction;
- ***information reflection and information memory***, i.e. possibility for collecting the information;
- ***information self-reflection***, i.e. possibility for generating "secondary information";
- ***information expectation*** i.e. the (secondary) information activity for internal or external contact
- ***information modeling and resolving the information expectation***

is called ***Infos***.

Conclusion

What gives us the concept “Infos”?

At the fist place, this is the common approach for investigating the natural and artificial information agents.

In other hand, this is the set of common characteristics which are basic for all entities, which we may classify to the category of the Infos.

And, at the end, this is a common philosophical basis for understanding the information subjects.

The development of the General Information Theory should not become by the single creative impulse. For a long period, the constructive activity of the many researchers is needed for establishing the new common paradigm.

We all need free scientific look at things, which will permit us to build the general theory without partiality, and aberrations taking in account all information paradigms already created and adopted.

Our main goal is to provoke the scientists to continue the research in this important area and to make the next step.

The concept **"Infos"** is basic for the General Information Theory [Markov et al. 2003a]. Its definition is only the starting point for further investigations and building the ***Infos Theory***.

The variety of types of Infoses in the real world needs to be investigated and classified in the future research. At the first step, we may propose that may be at least two main types of Infoses exist:

- ***infogens*** - the natural creatures;
- ***infotrons*** - the artificial creatures.

Also, the Infos Theory needs to give answers to many other very important questions, such as:

- What is the nature of the activity of the Infos?
- What is the difference between the living level of the Infos and the non-living one?
- Is it true that the boundary between non-living and living entities is self-reflection and internal activity for satisfying the secondary (information) possibilities for internal or external contact?

Etc.

It is impossible to answer to all questions in a single article. We may make only the next little step.

This was the aim of the present paper.

Bibliography

[Abdeev,1994] R.F.Abdeev. The Philosophy of the Information Civilization. Moscow, VLADOS, 1994. (in Russian)

[Bangov, 1995] I.Bangov. A Graph-Topological Model of Information. Int. J. Information Theories and Applications, 1995, v.3, n.6, pp.3-9.

[Burgin, 1997] M.S.Burgin. General Information Theory. http://www.math.ucla.edu/~mburgin/res/compsc/Site3GTI.htm

[Burgin, Markov, 1991] M.Burgin, Kr.Markov. Formal Definition of the Concept Materialization. Mathematics and Mathematical Education, BAS, Sofia, 1991, pp.175-179.

[Cooman et al, 1995] G.de Cooman, D.Ruan, E.Kerre, Eds. Foundations and Applications of Possibility Theory. World Scientific, Singapore, 1995.

[Elstner, 1993] D.Elstner. About Duality of the Information and Organization. Int. J. Information Theories and Applications, 1993, v.1, n.1, pp. 3-5. (in Russian)

[Gladun, 1994] V.P. Gladun. Processes of New Knowledge Formation. Pedagog 6, Sofia, 1994 (in Russian).

[Glushkov, 1986] V.M.Glushkov. Epistemological Nature of the Information Modeling. In: V.M.Glushkov. Cybernetics, Questions and Answers, (in Russian), Moscow, Science, 1986, pp. 33-41.

[Lu, 1999] C.-G.Lu. A Generalisation of Shannon's Information Theory. Int. J. of General Systems, 28:(6), 1999, pp.453-490.

[Markov, 1984] Kr. Markov. A Multi-domain Access Method. Proc. of Int. Conf. "Computer Based Scientific Research". Plovdiv, 1984. pp. 558-563.

[Markov, 1988] Kr.Markov. From the Past to the Future of Definition of the Concept of Information. Proceedings "Programming '88", BAS, Varna 1988, p.150. (In Bulgarian).

[Markov et al, 1993] Kr.Markov, Kr.Ivanova, I.Mitov. Basic Concepts of a General Information Theory. IJ Information Theories and Applications. FOI ITHEA, Sofia, 1993, Vol.1, No.10, pp.3-10

[Markov et al, 2003a] Kr.Markov, Kr.Ivanova, I.Mitov. General Information Theory. Basic Formulations. FOI-Commerce, Sofia, 2003.

[Markov et al, 2003b] K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova. The Information. IJ Information Theories and Applications, FOI ITHEA, Sofia, 2003, Vol.10, No.1, pp.5-9.

[Markov et al, 2004] K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova. Formal Definition of the Concept “INFOS”. Proceedings of the Second International Conference “Information Research, Applications and Education” i.TECH 2004, Varna, Bulgaria. Sofia, FOI-Commerce, 2004, pp. 71-74. Int. Journal “Information Theories and Applications”, 2004, Vol.11, No.1, pp.16-19

[Markov P., 2002] P.Markov. Think with Your Mind. Markov College. Sofia, 2002. (in Bulgarian)

[Pavlov, 1987] T.Pavlov. Collection of Selected Works, Vol. II. Theory of Reverberation. Science and Art, Sofia, 1987. (in Bulgarian).

[Rosenblueth, Wiener, 1945] A.Rosenblueth, N.Wiener. Role of Models in Science. Philosophy of Science, Vol.12, No.4, 1945.

[Shannon, 1949] C.E.Shannon. The Mathematical Theory of Communication. In: The Mathematical Theory of Communication. Ed. C.E.Shannon and W.Weaver. University of Illinois Press, Urbana, 1949.

[Tomov, 1991] K.Tomov. The Resomal-Isomorphic Principle. Arges, Sofia, 1991. (in Bulgarian)

[Wartofsky, 1979] M.W.Wartofsky. Models. Representation and the Scientific Understanding. D.Reidel Publishing Company, Dordrecht: Holland /Boston: USA, London: England/, 1979.

[Popper, 1968] K.R.Popper. Conjectures and Refutations: The Growth of Scientific Knowledge. Harper & Row, Publishers, New York and Evanston. 1968.

Authors' Information

Krassimir Markov - *Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; markov@foibg.com*

Krassimira Ivanova - *Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria, e-mail: ivanova@foibg.com*

Ilia Mitov - *Institute of Information Theories and Applications FOI ITHEA, P.O.Box: 775, Sofia-1090, Bulgaria; e-mail: mitov@foibg.com*

Hybrid Intelligent Systems

ADAPTIVE GUSTAFSON-KESSEL FUZZY CLUSTERING ALGORITHM BASED ON SELF-LEARNING SPIKING NEURAL NETWORK

Yevgeniy Bodyanskiy, Artem Dolotov, Iryna Pliss

***Abstract**: The Gustafson-Kessel fuzzy clustering algorithm is capable of detecting hyperellipsoidal clusters of different sizes and orientations by adjusting the covariance matrix of data, thus overcoming the drawbacks of conventional fuzzy c-means algorithm. In this paper, an adaptive version of the Gustafson-Kessel algorithm is proposed. The way to adjust the covariance matrix iteratively is introduced by applying the Sherman-Morrison matrix inversion procedure. The adaptive fuzzy clustering algorithm is implemented on the base of self-learning spiking neural network known as a realistic analog of biological neural systems that can perform fast data processing. Therefore, the proposed fuzzy spiking neural network that belongs to a new type of hybrid intelligent systems makes it possible both to perform fuzzy clustering tasks efficiently and to reduce data processing time considerably.*

***Keywords**: computational intelligence, hybrid intelligent system, fuzzy clustering, adaptive Gustafson-Kessel algorithm, self-learning spiking neural network, spiking neuron center, the temporal Hebbian learning.*

***ACM Classification Keywords**: I.2.6 [Artificial Intelligence]: Learning – Connectionism and neural nets; I.5.1 [Pattern Recognition]: Models – Fuzzy set, Neural nets; I.5.3 [Pattern Recognition]: Clustering – Algorithms.*

Introduction

Computational intelligence aims to achieve biologically and naturally plausible computing. Data clustering methods as a part of computational intelligent means has been successfully evolved in these directions over the several last decades.

On the one hand, data clustering methods progressed considerably toward more realistic approaches of clusters detection and separation. Unnatural 'all or none' membership restriction of hard clustering was overcome by fuzzy probabilistic clustering [Bezdek, 1981]. High degree of membership that fuzzy probabilistic clustering unnaturally produced for outliers that were equidistant from clusters centers was avoided by applying fuzzy possibilistic approach [Bezdek, 2005]. Adaptive versions of the mentioned clustering methods were proposed allowing of data processing in on-line mode [Park, 1994; Bodyanskiy, 2005]. Finally, the Gustafson-Kessel fuzzy clustering algorithm made it possible to detect clusters of different sizes and orientations [Gustafson, 1979; Krishnapuram, 1999].

On the other hand, data clustering means enriched by biologically plausible methods of data processing, namely, self-learning spiking neural networks [Natschlaeger, 1998; Bohte, 2002; Gerstner, 2002]. Being highly realistic models of biological neural systems, self-learning spiking neural networks appeared to be a considerably powerful and fast clustering tool of computational intelligence. Moreover, new type of hybrid intelligent systems

that combined capabilities of self-learning spiking neural networks and fuzzy probabilistic and possibilistic clustering approaches was shown to be successfully applied in various data clustering problems [Bodyanskiy, 2008a-f; Avdiyenko, 2008].

In this paper, adaptive version of the Gustafson-Kessel fuzzy clustering algorithms is introduced. It is shown how the proposed algorithm can be implemented on the base of self-learning spiking neural network.

Adaptive Version of the Gustafson-Kessel Fuzzy Clustering Algorithm

In the general case, the clustering quality criterion can be stated as following:

$$E(\mu_j(x(k)),v_j)=\sum_{k=1}^{N}\sum_{j=1}^{m}\mu_j^{\zeta}(x(k))\left\|x(k)-v_j\right\|_A^2,\ k=1,2,...,N,...;\ j=1,2,...,m\ , \tag{1}$$

where $\mu_j(x(k))\in[0,1]$ is membership level of the input pattern $x(k)$ to the j-th cluster, v_j is the center of the j-th cluster, N is the number of input patterns, m is the number of clusters, $\zeta\geq 0$ is the fuzzifier that determines boundary between clusters and controls the amount of fuzziness in the final partition, A is a norm matrix that defines distance measure. Under restriction

$$\sum_{j=1}^{m}\mu_j(x(k))=1, \tag{2}$$

minimization of (1) by applying the method of indefinite Lagrange multipliers leads us to the following solution:

$$\mu_j(x(k))=\frac{\left(\left\|x(k)-v_j\right\|_A^2\right)^{\frac{1}{1-\zeta}}}{\sum_{\iota=1}^{m}\left(\left\|x(k)-v_\iota\right\|_A^2\right)^{\frac{1}{1-\zeta}}}, \tag{3}$$

$$v_j=\frac{\sum_{k=1}^{N}\mu_j^{\zeta}(x(k))x(k)}{\sum_{k=1}^{N}\mu_j^{\zeta}(x(k))}. \tag{4}$$

In the case when norm matrix A is the identity matrix and $\zeta=2$, equations (3)-(4) present conventional fuzzy c-means algorithm. The Gustafson-Kessel algorithm uses the Mahalanobis distance measure, i.e. the inverse covariance matrix Σ_j^{-1} for the j-th cluster is used as a norm matrix in (3).

Minimization of (1) under restriction (2) based of the Arrow-Hurwicz-Uzawa gradient method [Arrow, 1958] produces the following recursive solution for the Mahalanobis distance measure:

$$\mu_j(x(k+1))=\frac{\left(\left(x(k+1)-v_j(k+1)\right)^T\Sigma_j^{-1}(k+1)\left(x(k+1)-v_j(k+1)\right)\right)^{\frac{1}{1-\zeta}}}{\sum_{\iota=1}^{m}\left(\left(x(k+1)-v_\iota(k+1)\right)^T\Sigma_\iota^{-1}(k+1)\left(x(k+1)-v_\iota(k+1)\right)\right)^{\frac{1}{1-\zeta}}}, \tag{5}$$

$$v_j(k+1)=v_j(k)+\eta_v(k)\mu_j^{\zeta}(x(k))\Sigma_j^{-1}(k+1)\left(x(k+1)-v_j(k)\right), \tag{6}$$

where $\eta_v(k)>0$ is a learning rate that defines to what extent information retrieved from new patterns will override old information .

Let us obtain recursive expression for the inverse covariance matrix. The covariance matrix for k input patterns is

$$\Sigma(k) = \frac{1}{k}\sum_{r=1}^{k}\left(x(r)-\bar{x}(k)\right)\left(x(r)-\bar{x}(k)\right)^T, \quad (7)$$

where $\bar{x}(k)$ is the average over k patterns. For a new incoming pattern $x(k+1)$, the covariance matrix can be written in form of

$$\Sigma(k+1) = \frac{1}{k+1}\left(k\Sigma(k) + \left(x(k+1)-\bar{x}(k+1)\right)\left(x(k+1)-\bar{x}(k+1)\right)^T\right). \quad (8)$$

By applying the Sherman-Morrison matrix inverse procedure

$$\left(B+ab^T\right)^{-1} = B^{-1} - \frac{B^{-1}ab^TB^{-1}}{1+b^TB^{-1}a} \quad (9)$$

to (8), the inverse covariance matrix can be expressed in recursive form:

$$\Sigma^{-1}(k+1) = \frac{k+1}{k}\left(\Sigma^{-1}(k) - \frac{\Sigma^{-1}(k)\left(x(k+1)-\bar{x}(k+1)\right)\left(x(k+1)-\bar{x}(k+1)\right)^T\Sigma^{-1}(k)}{k+\left(x(k+1)-\bar{x}(k+1)\right)^T\Sigma^{-1}(k)\left(x(k+1)-\bar{x}(k+1)\right)}\right). \quad (10)$$

Finally, the recursive average upon incoming pattern $x(k+1)$ is

$$\bar{x}(k+1) = \bar{x}(k) + \frac{1}{k+1}\left(x(k+1)-\bar{x}(k)\right). \quad (11)$$

In this paper, it is proposed to use centers of spiking neurons that are adjusted on each step of self-learning spiking neural network learning. This way, clusters centers adjustment (6) becomes unnecessary.

Self-Learning Spiking Neural Network

Architecture of self-learning spiking neural network is depicted on Figure 1. As illustrated, it is a heterogeneous two-layered feed-forward neural network with lateral connections in the second hidden layer [Bohte, 2002].

The first hidden layer performs pulse-position transformation of $n\times 1$-dimensional input patterns $x(k)$ (here, $k = 0,1,...,N$ is a pattern number) to the input vector of spikes $t^{[0]}(x(k))$ where each spike is defined by its firing time. The transformation is performed by population coding that implies that an input $x_i(k)$, $i = 1,2,...,n$, is processed at the same time by a pool of h receptive neurons RN_{li}, $l = 1,2,...,h$.

As a rule, a receptive neuron activation function is bell-shaped (Gaussian usually), and activation functions of the neurons within a pool are shifted, overlapped, and of different widths. Generally firing time of a spike emitted by receptive neuron lies in a certain interval $[0, t_{max}^{[0]}]$ referred to as a coding interval and is defined by the expression

$$t_{li}^{[0]}(x_i(k)) = \left\lfloor t_{max}^{[0]}\left(1-\psi\left(\left|x_i(k)-c_{li}^{[0]}\right|,\sigma_i\right)\right)\right\rfloor, \quad (12)$$

where $\lfloor\bullet\rfloor$ is the floor function, $\psi(\bullet,\bullet)$, $c_{li}^{[0]}$, and σ_i are the receptive neuron's activation function, center, and width respectively.

It is worth to note that a pool of receptive neurons RN_{li}, $l = 1,2,...,h$ can be treated as a certain linguistic variable of input data and each receptive neuron (more precisely, fuzzy receptive neuron) within the pool – as its linguistic term with membership function $\psi_{li}(x_i(k))$ [Bodyanskiy, 2008d]. Having any a priori knowledge of data structure, it is possible to adjust the first layer neurons activation functions as membership functions to fit the knowledge and thus, to get better clustering results.

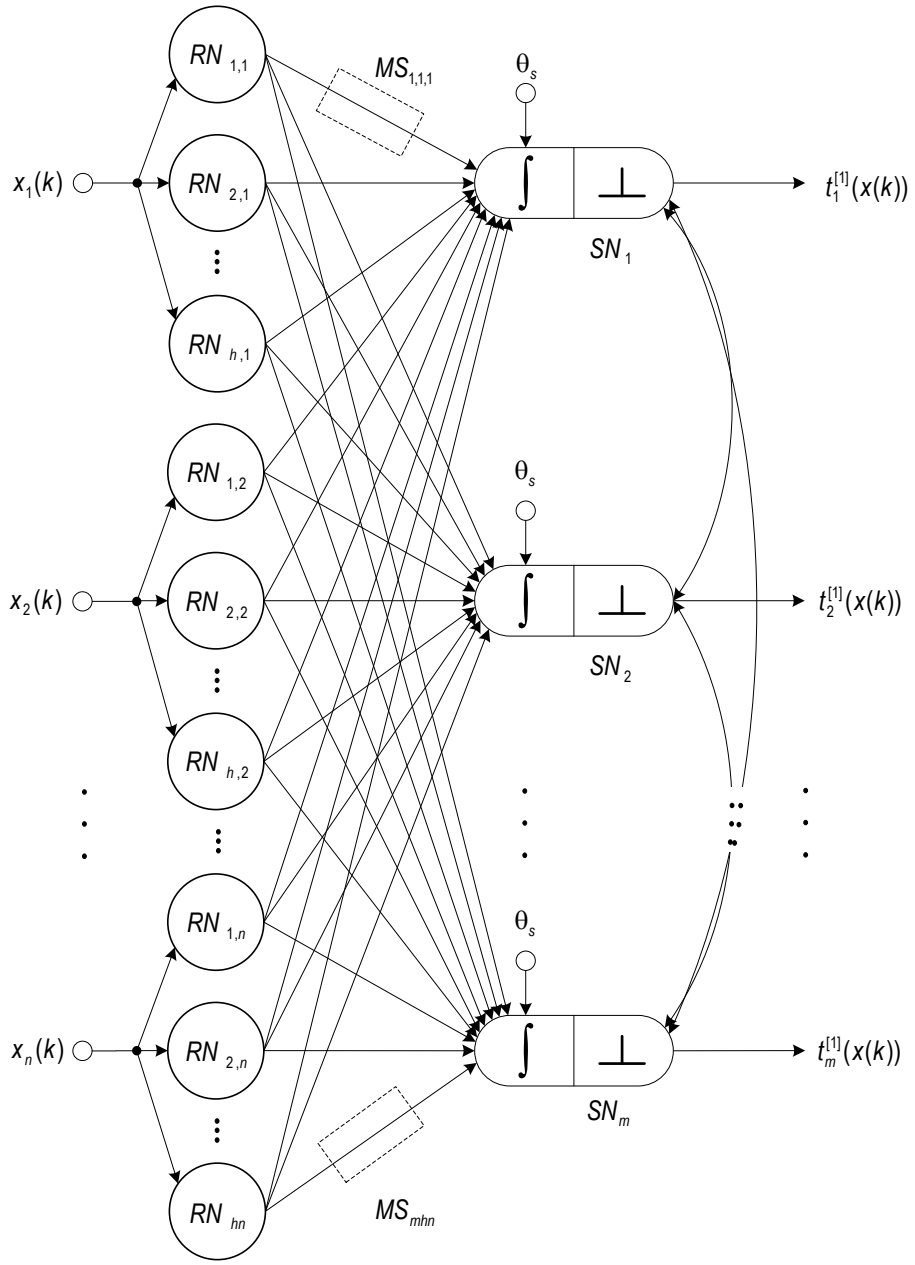


Figure 1. Self-learning spiking neural network architecture

The second hidden layer consists of m spiking neurons SN_j, $j=1,2,...,m$ (m – number of clusters). They are connected with neurons of the previous layer by multiple synapses. As shown on Figure 2, a multiple synapse MS_{jli} consists of a set of q subsynapses with different time delays d^p, $d^p - d^{p-1} > 0$, $d^q - d^1 > t_{\max}^{[0]}$, and varying weights w_{jli}^p (here $p=1,2,...,q$). It should be noted that number of subsynapses within a multiple synapse are fixed for the whole network. Having a spike $t_{li}^{[0]}(x_i(k))$ from the li-th receptive neuron, the p-th subsynapse of the j-th spiking neuron produces a delayed postsynaptic potential

$$u_{jli}^p(t) = w_{jli}^p \varepsilon_{jli}^p(t) = w_{jli}^p \varepsilon(t - (t_{li}^{[0]}(x_i(k)) + d^p)), \tag{13}$$

where $\varepsilon(\bullet)$ is a spike-response function usually described by the expression

$$\varepsilon(t) = \frac{t}{\tau} e^{1-\frac{t}{\tau}} H(t), \tag{14}$$

τ is the membrane potential decay time constant whose value can be obtained empirically, $H(\bullet)$ is the Heaviside step function. Output of the multiple synapse MS_{jli} forms total postsynaptic potential

$$u_{jli}(t) = \sum_{p=1}^{q} u_{jli}^{p}(t). \tag{15}$$

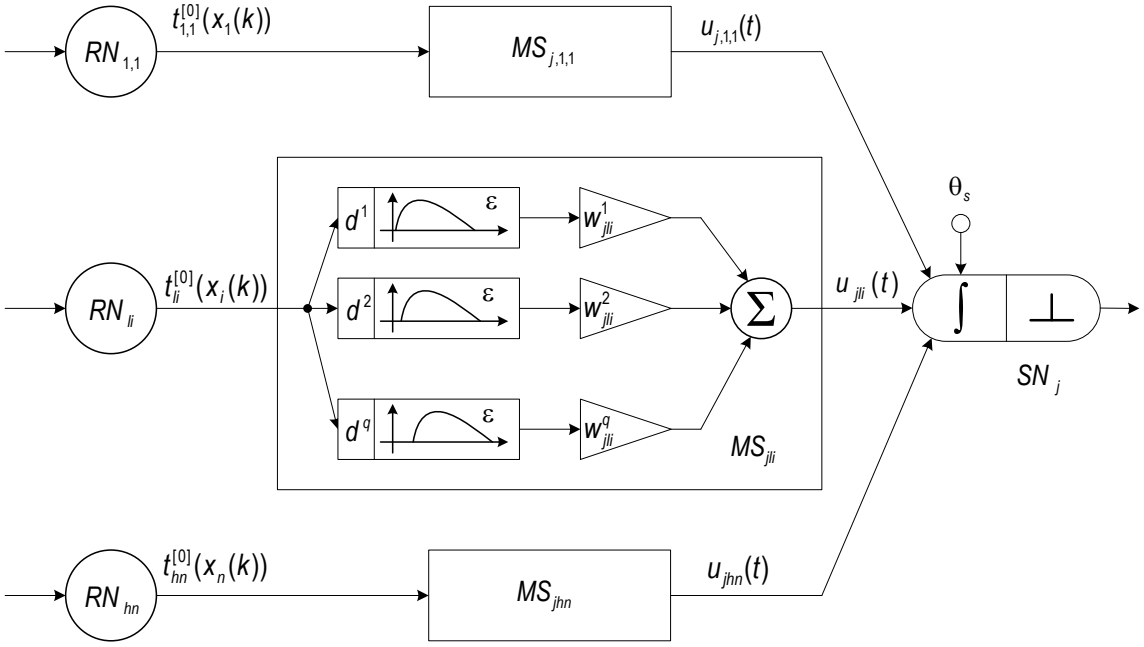


Figure 2. Multiple synapse

Each incoming total postsynaptic potential contributes to membrane potential of spiking neuron SN_j as follows:

$$u_j(t) = \sum_{i=1}^{n} \sum_{l=1}^{h} u_{jli}(t). \tag{16}$$

Spiking neuron SN_j generates at most one outgoing spike $t_j^{[1]}(x(k))$ during a simulation interval (the presentation of an input pattern $x(k)$), and fires at the instant the membrane potential reaches firing threshold θ_s. After neuron firing, the membrane potential is reset to the rest potential (0 usually) until the next input pattern is presented.

The spiking neuron described above corresponds to the integrate-and-fire model that is the one of the well-known models of a biological neuron [Gerstner, 2002].

Spiking neurons are linked with lateral inhibitory connections that disable all other neurons to fire after the first one has fired. In other words, lateral connections leads spiking neurons to fire according to the 'winner-takes-all' rule. After learning stage, each spiking neuron firing time reflects the distance of the input pattern to the neuron center. Thus, pattern supplied to the learned spiking neural network fires the spiking neuron whose center is the closest to it. This way self-learning spiking neural network performs clusters separation.

The purpose of the learning algorithm of self-learning spiking neural network is to move the center of the neuron-winner closer to input pattern by adjusting the neuron synaptic weights. The weights are adjusted according to the temporal Hebbian rule in the following way: weights of those subsynapses which contributed to the neuron's firing

are strengthened, whereas weights of subsynapses which did not contribute are weakened. As a rule, the temporal Hebbian rule used in spiking neural networks learning has the following form [Berredo, 2005]:

$$w_{jli}^{p}(K+1) = w_{jli}^{p}(K) + \eta_w(K)\left((1+\beta)\exp\left(\frac{\left(t_{li}^{[0]}(x_i(k)) + d^p - t_j^{[1]}(x(k)) - \alpha\right)^2}{2(1-\kappa)} \right) - \beta \right), \quad (17)$$

$$\kappa = 1 - \frac{\nu^2}{2\ln\left(\frac{\beta}{1+\beta}\right)}, \quad (18)$$

where K is the epoch number, $\eta_w(\bullet)$ is the learning rate, α, β, ν are shape parameters of the learning function.

One can readily see that the unsupervised learning of the spiking neurons layer is identical to the one of self-organizing maps (SOM).

Spiking Neuron Center

A spiking neuron center is considered to represent a spiking neuron response to the input pattern in a convenient way. In the general case, it is considered to possess the following property: the closer input pattern is to the neuron's center, the earlier output spike fires. Apparently, the earliest time of spiking neuron response can appears when all postsynaptic potentials reach the neuron soma simultaneously, i. e. the firing time of outgoing spike depends on synchronization degree of incoming spikes. It is worth to note here that synchronization phenomena is of primary importance in nature [Pikovsky, 2001].

In this work, spiking neuron center calculation is required to obtain the difference of incoming pattern and cluster center (represented by the neuron center) $x(k+1) - v_j(k+1)$ that is used in (5). Since the neuron center represents temporal aspect of the network functioning, we pursue the temporal difference of the input vector of spikes and the center $t^{[0]}(x(k)) - c_j^{[1]}$ ($c_j^{[1]}$ is the center of spiking neuron SN_j). Apparently, original form of pattern $x(k)$ should be replaced with time-pulsed form $t^{[0]}(x(k))$ in (5), (10), and (11) in this case.

There exist several ways to define spiking neuron center formally. The ones that were successfully used in fuzzy clustering are the Natschlaeger-Ruf center [Natschlaeger, 1998, Bodyanskiy, 2008c] and the Goren center [Goren, 2001, Avdiyenko, 2008]. Both of them requires calculation of the mean weighted synaptic delay d_{jli}. In this paper, it is proposed to calculate it in a general way, by using the quasi-arithmetic mean [Kolmogorov, 1985]:

$$d_{jli} = f^{-1}\left(\frac{\sum_{p=1}^{q} w_{jli}^{p} f(d^p)}{\sum_{p=1}^{q} w_{jli}^{p}} \right), \quad (19)$$

where $f(\bullet)$ is a strictly monotonic function.

T. Natschlaeger and B. Ruf proposed to calculate the center of spiking neuron SN_j as follows:

$$c_{jli}^{[1]} = d_{jli} - \min\{d_{jab} \mid 1 \le a \le h, 1 \le b \le n\}, \quad (20)$$

where $c_{jli}^{[1]}$ is a component of vector $c_j^{[1]} = (c_{j,1,1}^{[1]}, c_{j,1,2}^{[1]}, ..., c_{jh,1}^{[1]}, c_{jh,s}^{[1]}, ..., c_{jhn}^{[1]})^T$. Transforming the input vector of spikes as follows:

$$\rho_{li}(x_i(k)) = \max\{t_{ab}^{[0]}(x_b(k)) \mid 1 \le a \le h, 1 \le b \le n\} - t_{li}^{[0]}(x_i(k)), \quad (21)$$

the difference in (5) will be

$$x(k+1) - v_j(k+1) = \rho(x(k)) - c_j^{[1]}. \quad (22)$$

Goren center does not require the input vector of spikes transformation:

$$c_{jli}^{[1]} = \max\{d_{jab} \mid 1 \le a \le h, 1 \le b \le n\} - d_{jli}, \quad (23)$$

$$x(k+1) - v_j(k+1) = t^{[0]}(x(k)) - c_j^{[1]}. \quad (24)$$

It is notable that the Natschlaeger-Ruf center takes into account only the synchronization degree, whereas the Goren centers incorporates also temporal distance of spikes and the origin of coding interval. Anyway, even the Goren center does not reflect all peculiarities of spiking neuron firing. For example, it does not consider how value of firing threshold θ_s influences on the neuron firing time. Pursuing a center definition that would describe spiking neuron firing more precisely is a purpose of the further work.

In summary, the whole learning process of the proposed hybrid system can be formulated. Upon supplying a new pattern, self-learning spiking neural network adjusts the neurons centers using learning algorithm (17), (18) (as a rule, a few epochs are sufficient). Then a new fuzzy partitioning is performed according to (5), (10), and (11) using either Natschlaeger-Ruf (20)-(22) or Goren centers (23),(24). The simulation experiments show good results of the proposed hybrid system using in various data clustering problems solving.

Conclusion

The adaptive version of the Gustafson-Kessel fuzzy clustering algorithm is proposed. The new hybrid system is designed on the based of self-learning spiking neural network. It utilizes temporal difference of incoming vector of spikes and spiking neurons centers to perform fuzzy partitioning. The obtained hybrid system allows of combining tractability of the Gustafson-Kessel algorithm and velocity capabilities of spiking neural network and processes data in on-line mode.

Bibliography

[Arrow, 1958] K.J.Arrow, L.Hurwitz, H.Uzava. Studies in Linear and Nonlinear Programming. Stanford University Press, Stanford, CA, 1958.

[Avdiyenko, 2008] L.Avdiyenko, A.Dolotov, I.Pliss. Self-learning spiking neural network with synapse fuzzy inference based on Goren centers. Proc. Int. Sci. Conf. “Intellectual Systems for Decision Making and the Problems of Computational Intelligence – 2008” (May 19-23, 2008, Yevpatoria, Ukraine), Kherson National Technical University, Vol. 3, Part 1, Kherson, 2008, P. 11-14 (in Russian).

[Berredo, 2005] R.C.De Berredo. A review of spiking neuron models and applications. M. Sc. Dissertation, University of Minas Gerais, 2005.

[Bezdek, 1981] J.C.Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, 1981.

[Bezdek, 2005] J.C.Bezdek, J.Keller, R.Krishnapuram, N.R.Pal. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Springer, New York, 2005.

[Bodyanskiy, 2005] Ye.Bodyanskiy. Computational intelligence techniques for data analysis. Lecture Notes in Informatics, Vol. 72, 2005, P. 15-36.

[Bodyanskiy, 2008a] Ye.Bodyanskiy, A.Dolotov. A self-learning spiking neural network for fuzzy clustering task. Sci. Proc. of Riga Technical University: Information Technology and Management Science, 36, 2008, P. 27-33.

[Bodyanskiy, 2008b] Ye.Bodyanskiy, A.Dolotov. Image processing using self-learning fuzzy spiking neural network in the presence of overlapping classes. Proc. 11th Int. Biennial Baltic Electronics Conf. "BEC 2008" (October 6-8, 2008, Tallinn, Estonia),Tallinn University of Technology, Tallinn, 2008, P. 213-216.

[Bodyanskiy, 2008c] Ye.Bodyanskiy, A.Dolotov. Self-learning spiking neural network with synapse fuzzy inference based on Natschlaeger-Ruf centers. Proc. 10th Int. Sci. and Tech. Conf. "System Analysis and Information Technologies" (May 20-24, 2008, Kyiv, Ukraine), National Technical University of Ukraine "KPI", Kiyv, 2008, P. 173 (in Ukrainian).

[Bodyanskiy, 2008d] Ye.Bodyanskiy, A.Dolotov, I.Pliss. Fuzzy receptive neurons using in self-learning spiking neural network. Proc. Int. Sci. and Tech. Conf. "Automation: Problems, Ideas, Solutions" (September 8-12, 2008, Sevastopol, Ukraine), Sevastopol National Technical University, Sevastopol, 2008, P. 12-14 (in Russian).

[Bodyanskiy, 2008f] Ye.Bodyanskiy, A.Dolotov, I.Pliss, Ye.Viktorov. Fuzzy possibilistic clustering using self-learning spiking neural network. Wissenschaftliche Berichte der Hochschule Zittau/Goerlitz, Heft 100, Nr. 2360-2395, 2008, S. 53-60.

[Bohte, 2002] S.M.Bohte, J.S.Kok, H.La Poutre. Unsupervised clustering with spiking neurons by sparse temporal coding and multi-layer RBF networks. IEEE Trans. on Neural Networks, 13, 2002, P. 426-435.

[Gerstner, 2002] W.Gerstner, W.M.Kistler. Spiking Neuron Models. The Cambridge University Press, Cambridge, 2002.

[Goren, 2001] Ya.Goren. Clustering with spiking neurons. 2001 http://cis.cs.technion.ac.il/Done_Projects/Projects_done/SpikingNeurons/RBF.doc.

[Gustafson, 1979] D.E.Gustafson, W.C.Kessel. Fuzzy clustering with a fuzzy covariance matrix. Proc. IEEE Conf. Decision Contr. San Diego, CA, 1979, P. 761-766.

[Kolmogorov, 1985] A.N.Kolmogorov. Mathematics and Mechanics. Nauka, Moscow, 1985 (in Russian).

[Krishnapuram, 1999] R. Krishnapuram, J.Kim. A note on the Gustafson-Kessel and adaptive fuzzy clustering algorithms. IEEE Trans. on Fuzzy Systems, Vol. 7, 4, 1999, P. 453-461.

[Natschlaeger, 1998] T.Natschlaeger, B.Ruf. Spatial and temporal pattern analysis via spiking neurons. Network: Computations in Neural Systems, 9, 1998, P. 319-332.

[Park, 1994] D.C.Park, I.Dagher. Gradient based fuzzy c-means (GBFCM) algorithm. Proc. IEEE Int. Conf. on Neural Networks. IEEE Press, Orlando, FL, 1994, P. 1626-1631.

[Pikovsky, 2001] A.Pikovsky, M.Rosenblum, J.Kurths. Synchronization: A Universal Concept in Nonlinear Sciences. The Cambridge University Press, New York, 2001.

Authors' Information

Yevgeniy Bodyanskiy *– Professor, Dr.-Ing. habil., Scientific Head of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine; e-mail: bodya@kture.kharkov.ua.*

Artem Dolotov *– Ph. D. student, Junior Researcher of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine; e-mail: dolotov.artem@gmail.com.*

Iryna Pliss *– Ph. D., Leading Researcher of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine; e-mail: pliss@kture.kharkov.ua.*

THE CASCADE GROWING NEURAL NETWORK USING QUADRATIC NEURONS AND ITS LEARNING ALGORITHMS FOR ON-LINE INFORMATION PROCESSING

Yevgeniy Bodyanskiy, Yevgen Viktorov, Iryna Pliss

***Abstract**: New non-conventional system of the computational intelligence is proposed. It has growing structure similar to the Cascade-Correlation Learning Architecture designed by S. E. Fahlman and C. Lebiere but differs from it in type of artificial neurons. Quadratic neurons are used as nodes in introduced architecture. These simple elements can be quickly adjusted using high-speed learning procedures. Proposed approach allows to reduce time required for weight coefficients adjustment and to decrease training dataset size in comparison with conventional neural networks. Also on-board realization of quadratic neuron is quite simple and therefore implementation of entire cascade architecture in hardware is very easy.*

***Keywords**: artificial neural networks, constructive approach, quadratic neuron, real-time processing, online learning.*

***ACM Classification Keywords**: I.2.6 Learning – Connectionism and neural nets.*

Introduction

Nowadays artificial neural networks (ANNs) are widely applied for solving a large class of problems related with the processing of information given as time-series or numerical "object-properties" tables generated by the non-stationary, chaotic or stochastic systems. The most attractive ANNs properties are their approximating possibilities and learning capabilities.

Conventionally "learning" is considered as process of the neural network's synaptic weights adjustment accordingly to selected optimization procedure of accepted learning criterion [Cichocki, 1993; Haykin, 1999]. But during learning procedure not only weight coefficients but also network's architecture (quantity of nodes) can be adjusted for the purpose of increasing quality of received results. There are two basic approaches for the neural network architecture adjustment: 1) "constructive approach" [Platt, 1991; Nag, 1998; Yingwei, 1998] — starts with simple architecture and gradually adds new nodes during learning; 2) "destructive approach" [Cun, 1990; Hassibi, 1993; Prechelt, 1997] — starts with initially redundant network and simplifies it throughout learning process.

Obviously, constructive approach needs less computational resources and within the bounds of this technique the cascade neural networks (CNNs) [Fahlman, 1990; Schalkoff, 1997; Avedjan, 1999] can be marked out. The most efficient representative of the CNNs is the Cascade-Correlation Learning Architecture (CasCorLA) [Fahlman, 1990]. This network begins with the simplest architecture which consists of a single neuron. Throughout a learning procedure new neurons are added to the network, producing a multilayer structure. It is important that during each learning epoch only one neuron of the last cascade is adjusted. All pre-existing neurons process information with "frozen" weights. The CasCorLA authors, S. E. Fahlman and C. Lebiere, point out high speed of the learning procedure and good approximation properties of this network. But it should be observed that elementary Rosenblatt perceptrons with hyperbolic tangent activation functions are used in this architecture as nodes. Thus an output signal of each neuron is non-linearly depended from its weight coefficients. Therefore it is necessary to use gradient learning methods such as delta-rule or its modifications, and optimization an operation speed becomes impossible. In connection with the above it seems to be reasonable to

synthesize the cascade architecture based on the elementary nodes with linear or quadratic dependence of an output signal from the synaptic weights. It allows to increase a speed of synaptic weights adjustment and to reduce minimally required size of training set.

In [Bodyanskiy, 2007] ortho-neurons were proposed as such nodes. Also it was shown how simply and effectively an approximation of sufficiently non-linear function can be performed using this technique. But it should be noticed that on-board realization of the ortho-neuron is quite complex due to its functional specificities. At this paper we propose to use quadratic neurons as basic elements for the cascade architecture. They have simple structure and therefore their realization in hardware is simple too.

The Quadratic Neuron and Its Gradient Learning Procedure

The quadratic neuron is a nonlinear in inputs but linear in synaptic weights multi-input single output system shown on Fig. 1. It realizes the following mapping:

$$\hat{y}_j(k) = \theta_j(k) + \sum_{i=1}^{n} w_{ji}(k) x_i(k) + \sum_{p=1}^{n} \sum_{l=p}^{n} w_{jpl}(k) x_p(k) x_l(k) \tag{1}$$

where x_i is the *i*-th input (*i*=1,2,…,*n*); $\hat{y}$ is an output; θ_j is a bias in the *j*-th quadratic neuron; w_{ji} is a weight coefficient connected to *i*-th input in the *j*-th quadratic neuron; w_{jpl} is a weight coefficient connected to composition of *p*-th and *l*-th inputs in the *j*-th quadratic neuron.

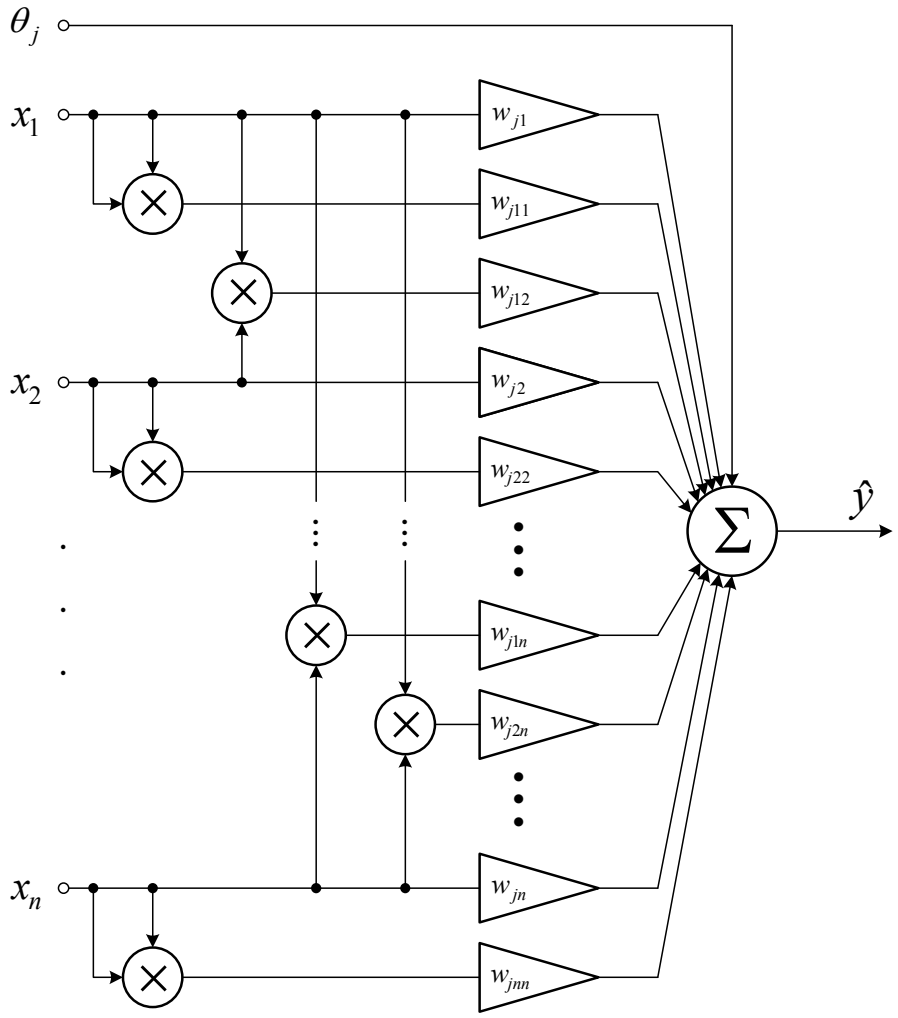


Figure 1. The Quadratic Neuron – QN

Let us define *additional* designations $w_{j0}(k) = \theta_j(k)$, $b_j(k) = (w_{j1}(k), w_{j2}(k), ..., w_{jn}(k))^T$ – $(n \times 1)$-vector, $C_j(k) = \{w_{jpl}(k)\}$ – $(n \times n)$-matrix, $x\_(k) = (x_1(k), x_2(k), ..., x_n(k))^T$ – $(n \times 1)$-vector, $x(k) = (1, x\_^T(k))^T$ – $((n+1) \times 1)$-vector. Then we can rewrite expression (1) in the form:

$$\hat{y}_j(k) = w_{j0}(k) + b_j^T(k) x\_(k) + x\_^T(k) C_j(k) x\_(k)$$

or

$$\hat{y}_j(k) = x^T(k) W_j(k) x(k)$$

where

$$W_j(k) = \left(\begin{array}{c|c} w_{j0}(k) & 0.5 b_j^T(k) \\ \hline 0.5 b_j(k) & C_j(k) \end{array} \right) \quad (2)$$

is block $((n+1) \times (n+1))$-matrix.

Weight coefficients matrix $W_j(k)$ adjustment can be performed by minimization of the quadratic learning criterion

$$E_j(k) = \frac{1}{2} e_j^2(k) = \frac{1}{2} (y(k) - x^T(k) W_j x(k))^2$$

(where $y(k)$ is an *external* learning signal) using gradient algorithm:

$$W_j(k+1) = W_j(k) + \eta(k) e_j(k) x(k) x^T(k) \quad (3)$$

where

$$e_j(k) = y(k) - x^T(k) W_j(k) x(k).$$

For the purpose of *evaluation* parameter $\eta(k)$ which provides optimal rate of convergence to algorithm (3) let us define values deviation matrix

$$\tilde{W}_j(k) = W_j - W_j(k)$$

where W_j is unknown matrix of optimal coefficients values, $W_j(k)$ (2) is its estimate on the *k*-th learning iteration.

Then solving the differential equation

$$\frac{\partial Tr(\tilde{W}_j(k) \tilde{W}_j^T(k))}{\partial \eta} = 0$$

(where $Tr(\bullet)$ is trace of matrix) optimal value of the step parameter can be obtained in the form [Bodyanskiy, 1987; Bodyanskiy, 1997]:

$$\eta(k) = \|x(k)\|^{-4}.$$

Using evaluated step parameter, expression (3) can be rewritten as

$$W_j(k+1) = W_j(k) + \frac{d_j(k) - x^T(k) W_j(k) x(k)}{\|x(k)\|^4} x(k) x^T(k). \quad (4)$$

Learning procedure (4) is Kaczmarz-Widrow-Hoff [Kaczmarz, 1937; Kaczmarz, 1993; Widrow, 1960] optimal algorithm extension for quadratic neuron.

As it can be readily seen the quadratic neuron is a generalization of the well known N-Adaline widely used in GMDH Neural Networks [Pham, 1995].

Quadratic neuron provides quite high precision of approximation and extrapolation of significantly non-stationary non-linear signals and processes but further we use it as an elementary node in the cascade architecture.

The Cascade Neural Network Based On Quadratic Neurons

The architecture of cascade neural network based on quadratic neurons is shown on Fig. 2

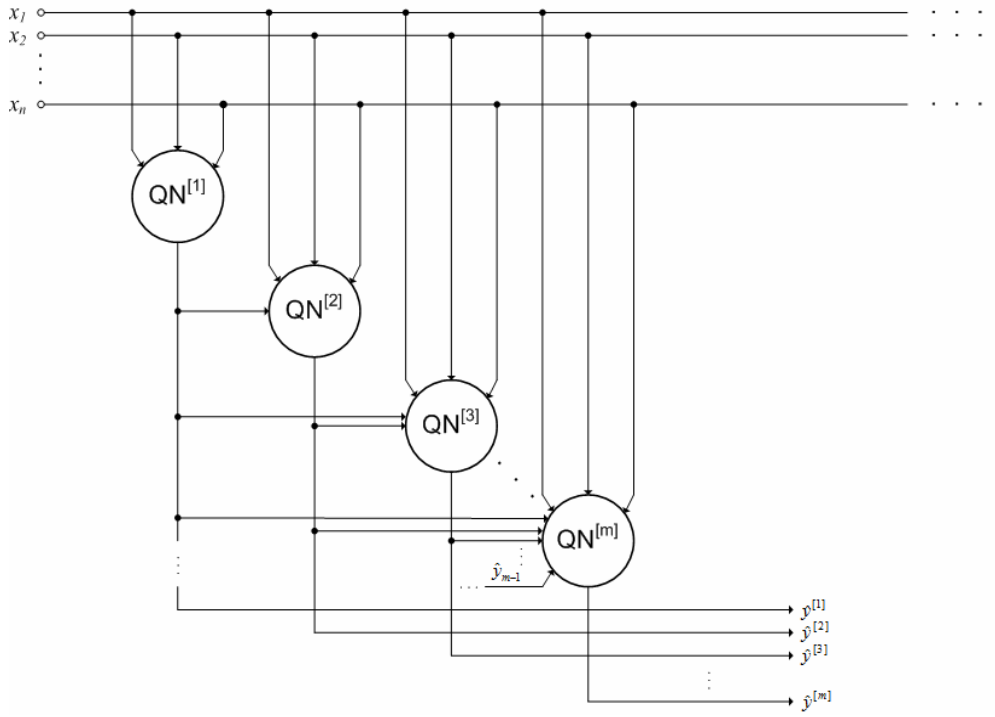


Figure 2. The Cascade Neural Network based on Quadratic Neurons

and mapping that it realizes has the following form:

– **first cascade quadratic neuron**

$$\hat{y}^{[1]}(k) = \theta_1(k) + \sum_{i=1}^{n} w_{1i}(k)x_i(k) + \sum_{p=1}^{n}\sum_{l=p}^{n} w_{1pl}(k)x_p(k)x_l(k) = x^{[1]T}w^{[1]}(k)x^{[1]}(k),\ \ x^{[1]} = (1, x^T(k))^T,$$

– **second cascade quadratic neuron**

$$\hat{y}^{[2]}(k) = \theta_2(k) + \sum_{i=1}^{n} w_{2i}(k)x_i(k) + \sum_{p=1}^{n}\sum_{l=p}^{n} w_{2pl}(k)x_p(k)x_l(k) + w_{2(n+1)}\hat{y}_1 + \sum_{p_x=1}^{n} w_{2p_x,n+1}x_{p_x}\hat{y}_1 + w_{2,n+1,n+1}\hat{y}_1\hat{y}_1 =$$

$$= x^{[2]T}w^{[2]}(k)x^{[2]}(k),\ \ x^{[2]} = (x^{[1]T}, \hat{y}^{[1]})^T,$$

– **third cascade quadratic neuron**

$$\hat{y}^{[3]}(k) = \theta_3(k) + \sum_{i=1}^{n} w_{3i}(k)x_i(k) + \sum_{p=1}^{n}\sum_{l=p}^{n} w_{3pl}(k)x_p(k)x_l(k) + \sum_{i_y=1}^{2} w_{3,n+i_y}\hat{y}_{i_y} + \sum_{l_x=1}^{2}\sum_{p_x=1}^{n} w_{3p_x,n+l_x}x_{p_x}\hat{y}_{l_x} +$$

$$+ \sum_{l_y=1}^{2}\sum_{p_y=1}^{l_y} w_{3,n+p_y,n+l_y}\hat{y}_{p_y}\hat{y}_{l_y} = x^{[3]T}w^{[3]}(k)x^{[3]}(k),\ \ x^{[3]} = (x^{[2]T}, \hat{y}^{[2]})^T,$$

– ***m*-th cascade neuron**

$$\hat{y}^{[m]}(k) = \theta_m(k) + \sum_{i=1}^{n} w_{mi}(k)x_i(k) + \sum_{p=1}^{n}\sum_{l=p}^{n} w_{mpl}(k)x_p(k)x_l(k) + \sum_{i_y=1}^{n+m-1} w_{m,n+i_y}\hat{y}_{i_y} +$$

$$\sum_{l_x=1}^{n+m-1}\sum_{p_x=1}^{n} w_{mp_x,n+l_x}x_{p_x}\hat{y}_{l_x} + \sum_{l_y=1}^{n+m-1}\sum_{p_y=1}^{l_y} w_{m,n+p_y,n+l_y}\hat{y}_{p_y}\hat{y}_{l_y} = x^{[m]T}w^{[m]}(k)x^{[m]}(k),\ \ x^{[m]} = (x^{[m-1]T}, \hat{y}^{[m-1]})^T \quad (5)$$

where *m* is quantity of cascades.

Thus the cascade neural network based on quadratic neurons contains $\sum_{q_1=1}^{n+1} q_1 + \sum_{q_2=1}^{n+2} q_2 + ... + \sum_{q_m=1}^{n+m} q_m$ adjustable parameters and it is important that all of them are linearly included in the description (5).

The Cascade Neural Network Based On Quadratic Neuron Learning Procedure

The cascade neural network learning can be performed in both the batch mode and the mode of sequential information processing using global learning criterion (6)

$$E_N^{[j]} = \frac{1}{2}\sum_{k=1}^{N} e_j(k)^2 = \frac{1}{2}\sum_{k=1}^{N} (y(k) - \hat{y}_j(k))^2. \quad (6)$$

Firstly, let us consider situation when the training data set is defined *a priory,* i.e. we have a set of points $x\_(1), y(1); x\_(2), y(2); ...; x\_(k), y(k); ...; x\_(N), y(N)$. Then for quadratic neuron of the first layer (QN[1]) weight coefficients vector is defined in form $w^{[1]} = = (w_{10}, w_{11}, ..., w_{1n}, w_{111}, w_{122}, ..., w_{1nn}, w_{112}, w_{113}, ..., w_{11n}, w_{123}, w_{124}, ..., w_{12n}, w_{134}, ..., w_{1,n-1,n})^T$ and also corresponding vector of internal quadratic neuron signals is defined as well $s^{[1]}(k) = (\theta_1, x_1(k), x_2(k), ..., x_n(k), x_1(k)x_1(k), x_2(k)x_2(k), ..., x_n(k)x_n(k), x_1(k)x_2(k), x_1(k)x_3(k), ..., x_1(k)x_n(k), x_2(k)x_3(k), x_2(k)x_4(k), ..., x_2(k)x_n(k), x_3(k)x_4(k), ...x_{n-1}(k)x_n(k))^T$.

Then using direct minimization of the learning criterion (6) vector of synaptic weights can be evaluated in the form

$$w^{[1]}(N) = \left(\sum_{k=1}^{N} s^{[1]}(k) s^{[1]T}(k)\right)^{+} \sum_{k=1}^{N} s^{[1]}(k) y(k) = P^{[1]}(N) \sum_{k=1}^{N} s^{[1]}(k) y(k) \quad (7)$$

where $(\bullet)^+$ denotes the Moore-Penrose pseudoinversion.

In the case of sequential data processing recurrent form of the least squares method can be used instead of procedure (7):

$$\begin{cases} w^{[1]}(k+1) = w^{[1]}(k) + \dfrac{P^{[1]}(k)(y(k+1) - w^{[1]T}(k) s^{[1]}(k+1))}{1 + s^{[1]T}(k+1) P^{[1]}(k) s^{[1]}(k+1)} s^{[1]}(k+1), \\ P^{[1]}(k+1) = P^{[1]}(k) - \dfrac{P^{[1]}(k) s^{[1]}(k+1) s^{[1]T}(k+1) P^{[1]}(k)}{1 + s^{[1]T}(k+1) P^{[1]}(k) s^{[1]}(k+1)}, \quad P^{[1]}(0) = \beta I \end{cases} \quad (8)$$

where β is a large positive number and I is a unity matrix of corresponding dimensionality.

Using of adaptive algorithms (3) or (4) is also possible and leads to reducing of computational complexity of learning process. But utilization of learning procedure (7) or (8) essentially reduces a learning time in comparison with gradient algorithms underlying delta-rule and backpropagation.

After the first cascade learning completion, the synaptic weights of the quadratic neuron QN[1] become “frozen”, all values $(\hat{y}^{[1]}(1), \hat{y}^{[1]}(2), ..., \hat{y}^{[1]}(k), ..., \hat{y}^{[1]}(N))$ are evaluated and the second cascade of the network which consists of a single quadratic neuron QN[2] is generated. It has one additional input for the output signal of the first cascade. Then the procedure (7) or (8) is again applied for adjusting a vector of weight coefficients $w^{[2]}$, which has dimensionality is $\sum_{q_2=1}^{n+2} q_2$.

The neural network growing process (increasing quantity of cascades) continues until we obtain required precision of the solved problem solution, and for the adjusting weight coefficients of the last *m*-th cascade following expressions are used:

$$w^{m1]}(N) = \left(\sum_{k=1}^{N} s^{[m]}(k) s^{[m]T}(k) \right)^{+} \sum_{k=1}^{N} s^{[m]}(k) y(k) = P^{[m]}(N) \sum_{k=1}^{N} s^{[m]}(k) y(k)$$

or

$$\begin{cases} w^{[m]}(k+1) = w^{[m]}(k) + \dfrac{P^{[m]}(k)(y(k+1) - w^{[m]T}(k) s^{[m]}(k+1))}{1 + s^{[m]T}(k+1) P^{[m]}(k) s^{[m]}(k+1)} s^{[m]}(k+1), \\ P^{[m]}(k+1) = P^{[m]}(k) - \dfrac{P^{[m]}(k) s^{[m]}(k+1) s^{[m]T}(k+1) P^{[m]}(k)}{1 + s^{[m]T}(k+1) P^{[m]}(k) s^{[m]}(k+1)}, \quad P^{[m]}(0) = \beta I \end{cases}$$

where vectors $w^{[m]}$ and $s^{[m]}$ have dimensionalities $\sum_{q_m=1}^{n+m} q_m$.

Simulation Results

In order to confirm efficiency of introduced architecture we have solved a dynamic plant identification problem. Proposed dynamic plant [Patra, 2002; Narendra, 1990] can be defined by the equation:

$$y(k+1) = 0.3 y(k) + 0.6 y(k-1) + f(u(k))$$

where

$$f(u) = 0.6 \sin u + 0.3 \sin 3u + 0.1 \sin 5u.$$

There was generated a sequence which contained 1500 values of signal for *k*=1,2,...,1500. On training set signal $u(k) = \sin 2k / 250$ (*k*=1,...,500) have been used and on the testing set $u(k) = \sin 2k / 250$ (*k*=501,...,1000), $u(k) = 0.5 \sin 2k / 250 + 0.5 \sin 2k / 25$ (*k*=1001,...1500). It means that on testing set sinusoidal component of the dynamic object changes and therefore output signal changes its form too. Obtained set was normalized on interval [-1 1].

For estimation of received results we have used normalized mean square error:

$$NRMSE(k, N) = \frac{\sum_{q=1}^{N} e^2(k+q)}{N\sigma}$$

where σ is a mean square deviation of the predicted process on the training set.

During simulation modeling we have used least squares method as well as adaptive algorithm (4) for the purpose of adjusting synaptic weight coefficients inside quadratic neurons. Also, the same problem had been solved using conventional multilayer perceptron. Obtained results are given in table 1 and on figure 3.

Table 1. Results of the dynamic object identification.

Artificial Neural Network	NRMSE
Multilayered perceptron (50 epochs using Levenberg-Marquardt procedure)	0.0011
Cascade architecture – 3 cascades (batch mode using LSM)	0.0009
Cascade architecture (mode of sequential real-time data processing using adaptive algorithm (4) – 1 epoch)	0.0015

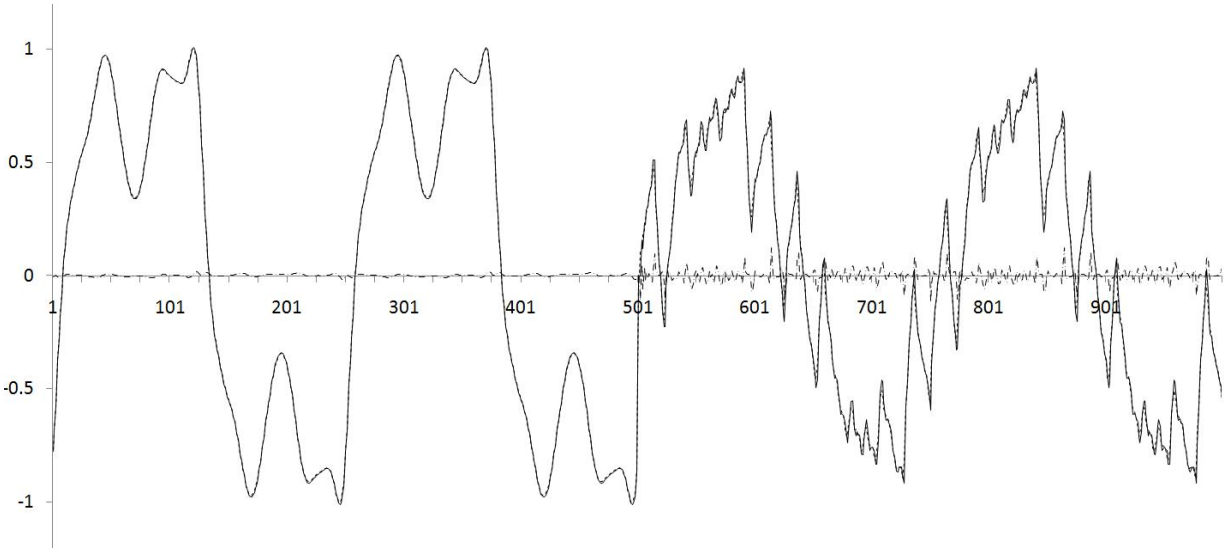


Figure 3. Dynamic object identification using cascade architecture trained with LSM: object output – solid line; network output – dashed line; identification error – chain line.

Conclusion

The Cascade Neural Network based on Quadratic Neurons is proposed. It differs from the known cascade networks in increased speed of operation, real-time processing possibility and simplicity of its on-board realization. Theoretical justification and experiment results confirm the efficiency of developed approach to cascade systems synthesis.

Bibliography

[Cichocki, 1993]. Cichocki A., Unbehauen R. Neural Networks for Optimization and Signal Processing. Stuttgart, Teubner, 1993.

[Haykin, 1999] Haykin S. Neural Networks. A Comprehensive Foundation. Upper Saddle River, N.J, Prentice Hall, 1999.

[Platt, 1991] Platt J. A resource allocating network for function interpolation. In: Neural Computation, 3. 1991. P. 213-225.

[Nag, 1998]. Nag A., Ghosh J. Flexible resource allocating network for noisy data. In: Proc. SPIE Conf. on Applications and Science of Computational Intelligence. 1998. P. 551-559.

[Yingwei, 1998] Yingwei L., Sundararajan N., Saratchandran P. Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm. IEEE Trans. on Neural Networks, 9. 1998. P. 308-318.

[Cun, 1990] Cun Y. L., Denker J. S., Solla S. A. Optimal brain damage. In: Advances in Neural Information Processing Systems, 2. 1990. P. 598-605.

[Hassibi, 1993] Hassibi B. Stork D. G. Second-order derivatives for network pruning: Optimal brain surgeon. In: Advances in Neural Information Processing Systems. Eds. Hanson et al. San Mateo, CA: Morgan Kaufman, 1993. P. 164-171.

[Prechelt, 1997]. Prechelt L. Connection pruning with static and adaptive pruning schedules. Neurocomputing, 16. 1997. P. 49-61.

[Fahlman, 1990] Fahlman S.E., Lebiere C. The cascade-correlation learning architecture. In: Advances in Neural Information Processing Systems. Ed. D.S. Touretzky. San Mateo, CA: Morgan Kaufman, 1990. P. 524-532.

[Schalkoff, 1997] Schalkoff R.J. Artificial Neural Networks. N.Y.: The McGraw-Hill Comp., 1997.

[Avedjan, 1999] Avedjan E.D., Barkan G.V., Levin I.K. Cascade neural networks. Avtomatika i Telemekhanika, 3. 1999. P. 38-55.

[Bodyanskiy, 2007] Bodyanskiy Ye., Viktorov Ye., Slipchenko O. Orthosynapse, ortho-neurons, and neuropredictor based on them. Systemi obrobki informacii, Issue 4(62). 2007. P. 139-143.

[Bodyanskiy, 1987] Ye. V. Bodyanskiy, Yu. A. Bobrovskiy. Adaptivnaya identifikaciya, prognozirovanie i upravlenie na osnove kvadratichnyx modelei. Kharkiv, 1987, 33 p. Ruk. dep. v UkrNIINTI 25.05.2987, №1515 – Uk 87.

[Bodyanskiy, 1997] Ye. V. Bodyanskiy, S. G. Udovenko, A. Ye. Achkasov, G. K. Voronovskii. Suboptimal'noe upravlenie stohasticheskimi processami. Kharkiv, Osnova, 1997. 140 p.

[Pham, 1995] D. T. Pham, X. Liu. Neural Networks for Identification Prediction and Control. London, Springer-Verlag, 1995.

[Kaczmarz, 1937] S. Kaczmarz. Angenaeherte Ausloesung von Systemen Linearer Gleichungen. Bull. Int. Acad. Polon. Sci, Let. A. 1937. S. 355-357.

[Kaczmarz, 1993] S. Kaczmarz. Approximate solution of systems of linear equations. Int. J. Control, 53. 1993. P. 1269-1271.

[Widrow, 1960] B. Widrow, M. E. Hoff Jr. Adaptive switching circuits. 1960 IRE WESCON Convention Record, Part 4. N.Y.: IRE, 1960. P. 96-104.

[Patra, 2002] J.C. Patra, A.C. Kot. Nonlinear dynamic system identification using Chebyshev functional link artificial neural network // IEEE Trans. on System, Man and Cybernetics, Part B. 2002, 32, №4. P. 505-511.

[Narendra, 1990] K.S. Narendra, K. Parthasarathy. Identification and control of dynamic systems using neural networks // IEEE Trans. on Neural Networks. 1990, 1, № 1. P. 4-26.

Authors' Information

Yevgeniy Bodyanskiy *– Professor, Dr.-Ing. habil., Professor of Artificial Intelligence Department and Scientific Head of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Lenina av. 14, Kharkiv, 61166, Ukraine, Tel +380577021890, e-mail: bodya@kture.kharkov.ua*

Yevgen Viktorov *- Ph.D. Student, Kharkiv National University of Radio Electronics, Lenin Av., 14, Kharkiv, 61166, Ukraine, Tel +380681613429, e-mail: yevgen.viktorov@gmail.com*

Iryna Pliss *– Dr.-Ing., Senior Researcher, Leading Researcher of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronic, Lenina av. 14, Kharkiv, 61166, Ukraine, Tel +380577021890, e-mail: pliss@kture.kharkov.ua*

THE APPLICATION OF ARTIFICIAL NEURAL NETWORKS AND EVOLUTIONARY ALGORITHM FOR THE DESIGNING OF GAS NITRIDING PROCESS

Mateusz Kosikowski, Zbigniew Suszyński, Roman Olik, Jerzy Ratajski, Tomasz Suszko

***Abstract**: The authors have undertaken a research task with a view to apply of evolutionary algorithms and artificial neural network to design of the course of a gas nitriding process, which guarantees to obtain the expected hardness profile in the nitrided layer. The gas nitriding process is widely used in industry in order to improve the functional properties of machine and tool parts. First of all, the artificial neural network was trained dependences between physical properties of steel and process parameters in relation to hardness profile, formed in gas nitriding process. Those data was collected experimentally. Such trained neural network was used as a mathematical model for the design approach of gas nitriding process. The aim of such designing was to predict the parameters of nitrinding industrial process, in which the required hardness profile of nitriding layer will be obtained. Prediction of manufactory conditions was realized with the help of evolutionary algorithms. In the approach developed, each chromosome includes encoded parameters of a single steel nitriding process. For each chromosome, a steel hardness profile related to it is determined by the model which is represented by neural network. On the further step, the chromosomes undergo a selection, a modification with the aid of crossing mutation operations and promotion to the next population. In this manner, by approach of a directed evolution of steel nitriding parameters, one is selected for which the hardness profile matches the best to profile sought.*

***Keywords**: evolutionary algorithm, neural network, designing of nitriding process, nitriding layer, hardness profile, model, manufactory conditions prediction.*

***ACM Classification Keywords**: D.2.2 Design Tools and Techniques - Evolutionary prototyping*

Introduction

Computer aided designing of formation technologies of surface layers with the required and repeatable functional properties constitute one of the key issues of the surface layer engineering. The gas nitriding process is known for a long time as great support for improving of durability of industrial tools such as cog-wheels or cutting tools. The layer developed during steel gas nitriding is composed of the surface zone of iron (carbon) nitrides and the diffusive zone located directly below [Somers 1995, Rozendaal 1985, Lengenhan 1992, Zyśk 1979]. The zone of (carbon) nitrides for pure iron, created at a relatively high value of nitric potential is composed, in compliance with Lehrer's diagram [Somers 1995, Rozentaal 1985]. It was assumed that his growth kinetics of the diffusive layer depends of the process temperature only. In the case of alloy and carbon steels, the sequence of phases in the zone of (carbon) nitrides changes in the duration of the process [Lenagenhan 1992, Zyśk 1979, Schewerdtfeger 1969, Lehrer 1930, Somers 1990, 1997, Mittemeijer 1980, 1983]. On the basis of the research in papers [Ratajski 2003, 2009, Malinov 2003], it was demonstrated that owing to the construction and phase composition of the iron (carbon) nitride zone being different than in the case of iron, as well as the structural changes occurring in that zone during the process, the quasi-equilibrium of nitrogen concentration is upset on the interfacial boundary of the diffusive zone/iron (carbon) nitride zone. Moreover, it was demonstrated that the phase structure of the iron (carbon) nitride zone has a significant contribution, regardless of the nitrogen potential and the temperature, to the creation of the diffusive zone, and its effective thicknesses g400, g500 and g600 in particular (Figure 1).

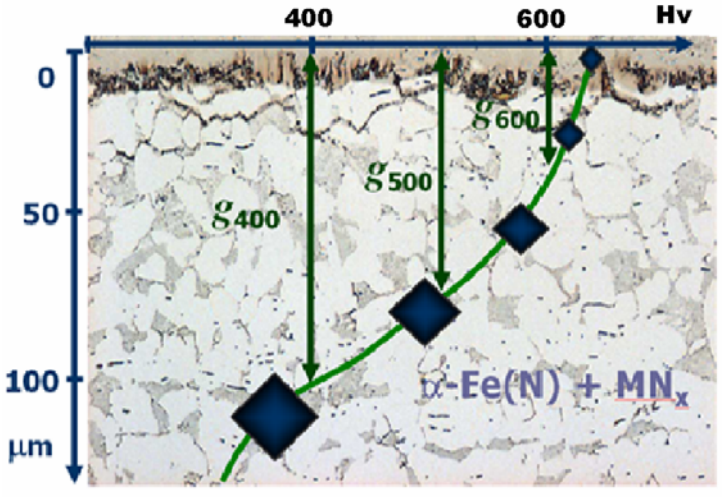


Figure 1. The effective thicknesses zone on the micro-hardness profile of nitride layer

To produce of the nitride layer is a very complex assignment therefore there is no derived accurate formula of mathematical model of such process so far. Difficulty of modelling came down from many very important parameters such as: chemical composition of steel (physical properties), process temperature which may not be constant during the process, nitrogen potential (K_N) [Ratajski, 2009] as well as method's recurrence. Designing of software for the system control of the gas nitiding process that could guarantee obtaining a layer with the required and repeatable functional properties is made difficult in connection with the phenomena and mechanisms which have an impact on the growth of the nitrided layer. As a result, there is an increasing interest in computer aided designing of surface layer formation technologies. The authors undertake attempt to design a software for prediction and control the gas nitride process. Such software is based on a artificial neural network which is used as a mathematical model for the evolutionary algorithm to solve of the inverse problem that mean to predict the determined process parameters based upon a required hardness profile. This is wary important for the industry because very often companies want to increase the endurance of tool to diamond level. In many cases the working environmental' process parameters are trying to get experimentally.

ANN as a model of a gas nitriding process

The mathematical model for determination of goal function for evolutionary algorithm is replaced by artificial neural network. Such ANN is trained with the help of data set which were collected experimentally. For the specified type of steel the sequence of gas nitridig process with different value of process parameters ware realized and the hardness profiles ware measured. Those data was used as a trainee set for the ANN as it is shown on figure 2. The implementation of such a concept for a specific surface treatment process enables to replace a very costly trial-and-error method, which is used at present in practice. However, it requires that a number of complex scientific and application problems be solved concerning modeling of confounded functional dependences between the surface layer properties and the characteristics of the process environment and the base material. The examples of process parameters and related hardness profile measured on chosen depth are shown in table 1. For same cases the process was repeated with different temperatures, duration of process or nitrogen potential and just after that the hardness profile was measured. For cases the nitriding process is repeated the hardness of surface layer is significantly greater than in the one stage approach.

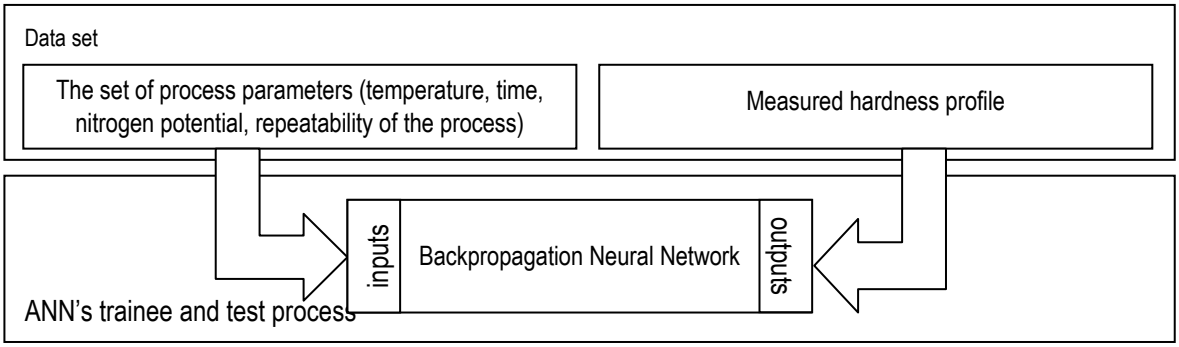


Figure 2. The schema of hardness profile determination with the help of ANN

For the experiment authors used a standard babckpropagation neural network with modified Chain's and Felside's trainee algorithm [Nałęcz 2000, Suszyński 2006, Kosikowski 2007]. The decision to choose such type of ANN was determined by its properties which are easy implementation of trainee algorithm, convergence and low duration of trainee process. Backpropagation ANN's require data to be coded in (-1;1) or (0 ;1) range which is imperfection of such networks, but they have very good approximation properties.

Table 1. Measured hardness profile received during gas nitriding process with related environmental parameters

Nr	Gas nitriding process parameters						Hardness profile [Hv] measured in chosen depth [um]							
	first stage			second stage										
	T_1[°C]	K_N	t_1 [min]	T_2[°C]	K_N	t_2 [min]	50	100	150	200	250	300	350	600
1	530	6	120	-	-	-	660	600	360	310	290	280	260	260
2	530	6	720	-	-	-	695	680	610	530	460	385	335	280
3	570	3.2	240	-	-	-	655	600	505	415	335	295	270	250
4	570	3	480	-	-	-	715	630	550	475	420	375	340	270
5	570	3	960	-	-	-	740	680	625	575	525	470	415	280
6	550	3.5	720	-	-	-	620	605	525	425	365	320	300	270
7	550	30	720	-	-	-	690	655	600	530	465	395	365	280
8	570	5.2	1200	-	-	-	700	640	585	545	520	490	455	315
9	580	2	180	-	-	-	675	585	495	410	360	325	295	270
10	580	10.2	120	580	0.4	240	690	610	535	460	385	325	395	260
11	530	8.5	240	530	0.6	120	780	710	570	420	360	320	305	280
12	530	9	240	530	0.6	240	770	730	640	600	515	465	435	270
13	530	10	240	530	0.6	480	780	725	620	525	430	380	345	280
14	490	12	120	560	2.4	30	540	325	260	260	260	260	260	260
15	480	17	225	560	3.7	25	720	440	330	295	290	270	270	270
...														
85	490	10	180	570	2	180	715	640	480	400	355	315	225	250
...														

Evolutionary algorithm combined with ANN

To predict the nitriding process parameters based on exact hardness profile, that mean to solve the inverse problem, the evolutionary algorithm was used. In such approach each chromosome was defined by encoded parameters of a single steel nitriding process. For the experiment authors used tools made from the alloyed carbon steel known as 18HGT. Chromosomes sets form the so-called population. For each chromosome, a steel hardness profile corresponding to it is determined (Figure 3.). For this purpose, the earlier trained neural network is used, which performs the role of a model. If the end-point condition of algorithm is reached the demand parameters became predicted. In the opposite, each chromosome (a set of parameters) evaluates. The chromosomes undergo a selection (a selection of chromosomes with the best matching), and a modification with the aid of crossing over and mutation operations. The chromosome selection algorithm was a nonlinear ranking with stress (pressure) coefficient equal to 0.3. The mutation treatment of individuals in population was assigned by standard single one-point crossing with randomly calculated probability which was less than 0.8. At the end of population evolution step, the homogeneous individual, with constant probability equal to 0.2, mutation was conduct. In such approach, by way of a directed evolution of steel nitriding parameters, chromosome for which the hardness profile matches best the profile sought is appointing the industrial process parameters that should be customary to achieve the requested hardness of a nitriding object.

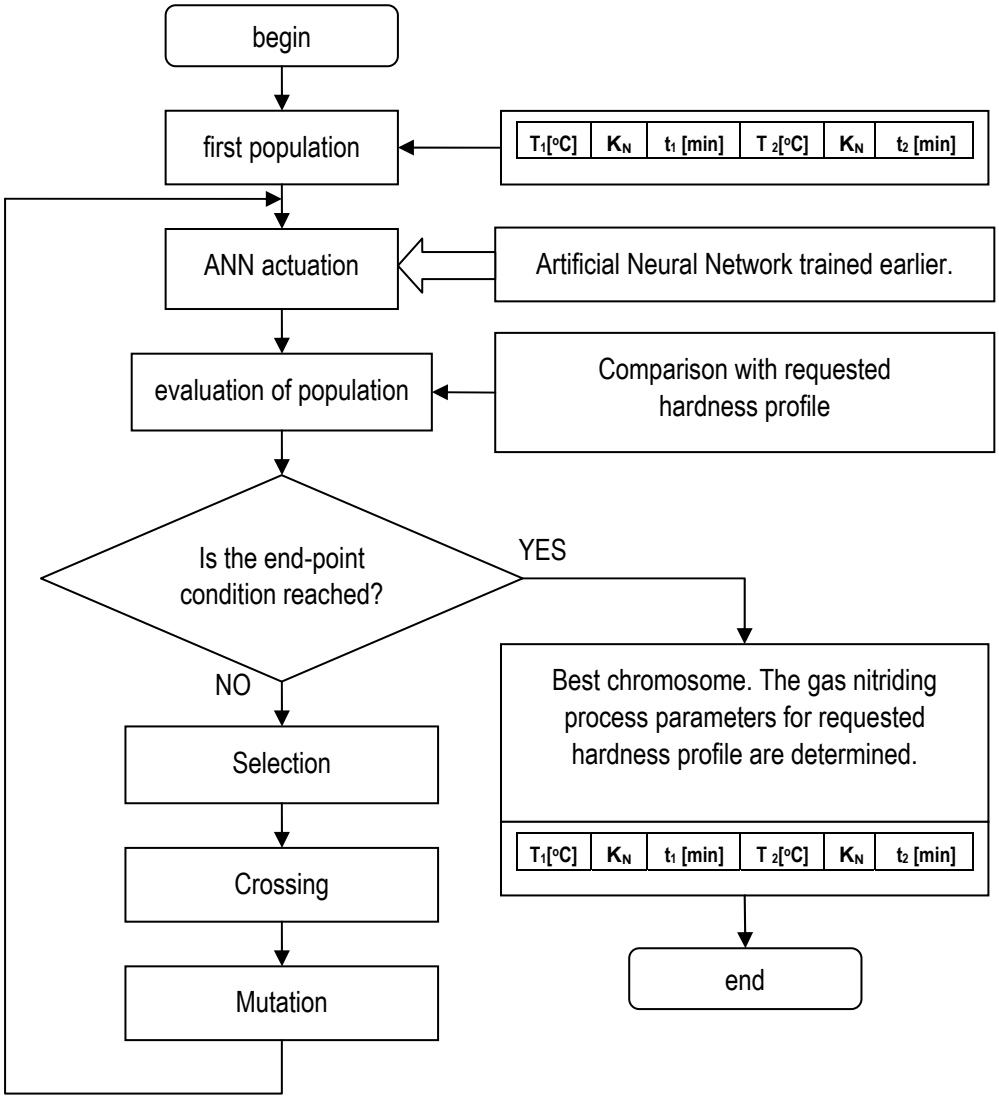


Figure 3. Flowchart of evolutionary algorithm used in experiment

Experiment – the utilization of designed algorithm for prediction of parameters of industrial gas nitriding working processes

The experiment was realized with the measurement hardness profile collected from tool parts made form 18HGT alloyed carbon steel which is European most popular nitriding grade. This steel is used with success for nitriding as well as for welding, hammersmithing, rolling and surface- of flame-hardening. At the first stage the required hardness profile is set to software, as it is shown on figure 4., and after that the evolutionary parameters, described in previous section is activated. The figure 5., shows the software during the evolutionary calculation procedure which was repeated twice. The algorithm stops when the fitting error, which is the absolute difference between required profile and the determined one by ANN, is non-changeable during 25 populations.

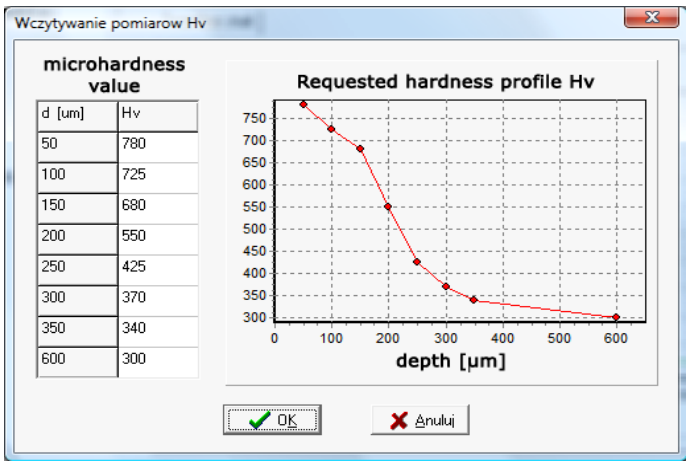


Figure 4. The software dialogue form to put the demand hardness profile in

The result of each actuation of algorithm is different – in presented case the dispersion of fitting error is between ~31÷62 (Figure 5.). Because the difference between calculated errors is so huge each time the algorithm is actuated, there is a high probability that the global minimum of error wasn't reached by the evolutionary algorithm and the calculations should continuously as long as the error is not changeable by 50 populations, for example.

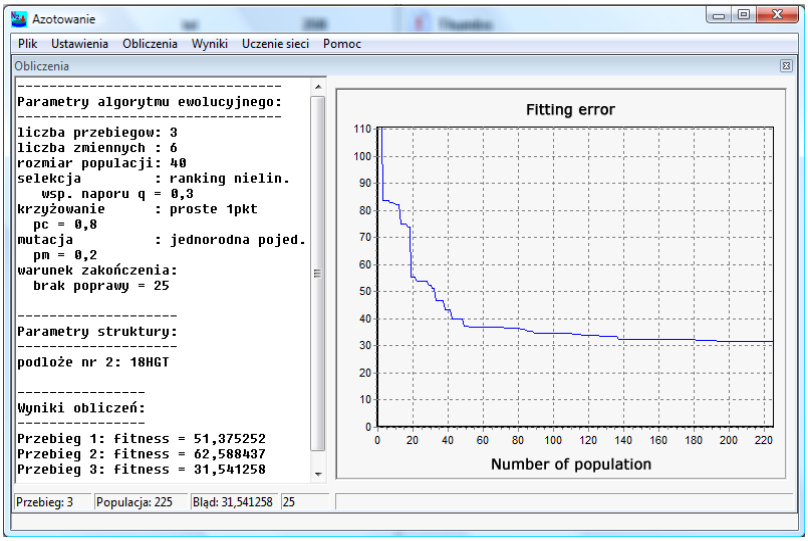


Figure 5. The screenshot of designed software window that shows the dependences of the fitting error of the best chromosome in particular population

Although the predicted harness profile for case the error is equal to 31.541258 is very similar to requested one (figure 6.). In principle the predicted profile is a very good approximation of requested one, which is conditioned by a nature of backpropagation neural network used by authors. The nitriding parameters predicted by designed

software are presented on the left side of software window shown on figure 6. The significant to explain is fact that some of calculated parameters, for example time and especially temperature of process, don't have to be determined exactly. The expected accuracy is rank of 5°C, because there is hard to hold the constant temperature up during the nitriding process. Therefore the value of 571.1°C can be understood as 570°C. The duration of gas nitriding process should be round up because, as it can be read from table 1., longer time secure better hardness. In fact those variables should be rounded by the algorithm himself and they could be represented by integer's data. It is easy to see that the goal function developed in evolutionary algorithm can be defined different. In approach authors presents the aim is to minimize the absolute difference between predicted hardness profile and the sought one. In other approach it could minimize not only the fitting error, but the production cost calculated from appointed process parameters also [Słowik 2004, 2008].

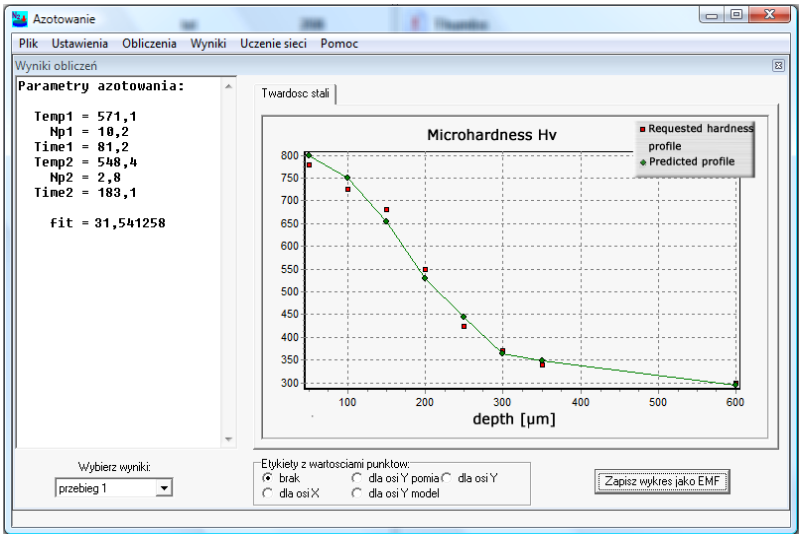


Figure 6. The result window with drawn chart which compare the requested hardness profile with predicted one

Conclusion

The work presented covers the concept of unconventional designing methods, which is use the combined artificial intelligence methods in application to predict the gas nitriding process being very popular in industry. The reason to choose such approach was numerical limitations which occur during calculations of nitriding model with the help of analytical models. Developed system of evolutionary algorithm combined with artificial neural network, as a mathematical model for determination of goal function, makes possible to determine the changes of the parameters of the gas nitriding process on the basis of the required hardness profile in the layer nitrided. The problem solved on the example of gas nitriding, i.e. determination of the process parameters which guarantee the required final result, constitutes one of the key issues of surface layer engineering. It is possible to calculate and minimize the total cost of described manufactory process, by modification of the goal evolutionary algorithm's goal function. The total calculation time ends in maximum 3 minutes, so a person interested may be informed of order cost without more ado. Hardness profile for predicted parameters is nearly identical with requested one, but in further step authors want to verify the method developed, which wasn't done yet.

Bibliography

[Somers, 1995] M. A. J. Sommers, E. J. Mittemeijer. In: Metall. Trans. A 26A, 1995, pp. 57-74

[Rozendaal, 1985] H. C. F. Rozendaal, P. F. Colijn, E. J. Mittemeijer. In: Surface Eng., 1985, pp. 30-43

[Langenhan, 1992] B. Langenhan, H. J. Spies. In: Härt-Tech. Mitt. 47, 1992, pp. 337-343.

[Zyśk, 1979] J. Zyśk, J. Tacikowski, E. Kasprzycka. In: Härt.–Tech. Mitt. 34, 1979, pp. 263-271.

[Prenosil, 1965] B. Prenosil. In: Härt-Tech. Mitt. 20, 1965, pp. 41-49.

[Schewerdtfeger, 1969] K. Schwerdtfeger. In: Trans. Aime 245, 1969, pp. 2461-2466.

[Lehrer, 1930] E. Lehrer. In: Elektochem. 26, 1930, pp. 383-392.

[Somers, 1990] M. A. J. Somers, P. F. Colijn, W. G. Sloof, E. J. Mittemeijer. In: Metallkde. 81, 1990, pp. 33-43.

[Somers, 1997] A. J. Somers, et al. In: Acta Mater. 45, 1997, pp. 2013-2025.

[Mittemeijer, 1980] E. J. Mittemeijer, et al. In: Scripta Metall. 14, 1980, pp. 1189-1192.

[Mittemeijer, 1983] E. J. Mittemeijer, et al. In: Proc. Of the Conf. On Heat Treatment, The Metal Society, London, 1983, pp. 107-115.

[Ratajski, 2003] J. Ratajski, J. Tacikowski, M. A. J. Somers. In: Surface Eng. 19, 2003, pp. 87-93.

[Ratajski, 2009] J. Ratajski. In: Surf. and Coat. Tech. 30, 2009, pp. 2300-2306.

[Malinov, 2003] S. Malinov, W. Sha. In: Computational Materials Science 28, 2003, pp.179-198.

[Nałęcz, 2000] M. Nałęcz. In: Biocybernetyka i inżynieria biomedyczna Tom 6 – Sieci Neuronowe (in Polish), EXIT, Warszawa, 2000, pp. 92-97

[Suszyński, 2006] Z. Suszyński, M. Kosikowski, R. Duer. The application of artificial neural Network of the assessment of thermal properties of multilayered semiconductor structure. In: Proceedings of the 12th International Workshop on Thermal Investigations of ICs and Systems THERMINIC, 2006, Nice France.

[Kosikowski, 2007] M. Kosikowski. In: Sieć neuronowa do wyznaczania parametrów cieplnych struktur warstwowych na podstawie obrazów termo-falowych, Politechnika Koszalińska, 2007

[Słowik, 2004] A. Słowik, M. Białko. In: Modified version of Roulette Selection for evolution algorithm – The Fan Selection. Lecture Notes in Artificial Intelligence, Springer, V. 3070,2004, pp.474-779

[Słowik, 2008] A. Słowik, J. Słowik. In: Multi-objective optimization of surface grinding process with the use of evolutionary algorithm with remembered Patero set, International Journal of Advanced Manufacturing technology, Springer, London, V. 37, No 7-8, June,2008, pp. 657-669.

Authors' Information

Mateusz Kosikowski *– M.Sc., Section of Multimedia Systems and Artificial Intelligence, Department of Electronics and Computer Science, Koszalin University of Technology, Śniadeckich 2, 75-453 Koszalin, Poland, e-mail: kosik@ie.tu.koszalin.pl*

Zbigniew Suszyński *– Ph.D., D.Sc., Chair of Multimedia Systems and Artificial Intelligence, Department of Electronics and Computer Science, Koszalin University of Technology, Śniadeckich 2, 75-453 Koszalin, Poland, e-mail: zas@man.koszalin.pl*

Roman Olik *– M.Sc., Institute of Mechatronics, Nanotechnology and Vacuum Technique, Koszalin University of Technology, Racławicka 15-17, 75-620 Koszalin, Poland, e-mail: olik@tu.koszalin.pl*

Jerzy Ratajski *– Ph.D., D.Sc., Chair of Institute of Mechatronics, Nanotechnology and Vacuum Technique, Koszalin University of Technology, Racławicka 15-17, 75-620 Koszalin, Poland, e-mail: jerzy.ratajski@tu.koszalin.pl*

Tomasz Suszko *– Ph.D., Institute of Mechatronics, Nanotechnology and Vacuum Technique, Koszalin University of Technology, Racławicka 15-17, 75-620 Koszalin, Poland, e-mail: tomasz.suszko@tu.koszalin.pl*

Knowledge Discovery and Data Mining

KERNEL-BASED METHODS FOR NON-STATIONARY TIME-SERIES IDENTIFICATION AND PREDICTION

Leonid Lyubchyk, Vladyslav Kolbasin

***Abstract**: Identification and prediction problem of nonlinear time-series generated by discrete dynamic system is considered via Kernel Method approach. A unified approach to recurrent kernel identification algorithms design is proposed. In such a way a recurrent modification of initial Kernel Method with growing windows is considered. In order to prevent the model complexity increasing under on-line identification, the reduced order model kernel method is proposed and proper recurrent identification algorithms are designed along with conventional regularization technique. Such an approach leads to a new type of Recursive Least-Square Kernel Method identification algorithms. Finally, the recurrent version of Sliding Window Kernel Method is also developed along with suitable identification algorithms. The proposed algorithm has tracking properties and may be successfully used for on-line identification of nonlinear non-stationary time-series.*

***Keywords**: identification, kernel methods, machine learning, nonlinear model, prediction, recurrent least-squares, support vector machine, time-series*

***ACM Classification Keywords**: G. Mathematics of Computing: G.1 Numerical Analysis: Least squares approximation, nonlinear approximation, G.3 Probability and Statistics: Time series analysis*

Introduction

The problems of time-series identification and prediction are of a great importance in different applications, such as signal processing, automatic control, econometrics etc. Most efficient methods based on classical identification approaches [Ljung, 1999] were developed for linear time-series, described by autoregressive - moving average (ARMAX) or, in most general case, by discrete-time state-space dynamic model. With respect to the nonlinear time-series, for instance, discrete-time chaotic processes, the most difficult problem is the admissible model selection and a model structure choice [Dorffner, 1996]. At that the complex dynamic nonlinear mapping recovery using the sample data in the classical framework of parameterized model application leads to the multi-parameter estimation problem which becomes very complicated due to «curse of dimension» difficulties. It's stipulated the expedience of nonparametric methods and intelligent data analysis approaches such as artificial neural networks and machine learning, which usually needs a long sample for training. Most efficient under the short training sample are Support Vector Machine (SVM) [Vapnik, 1998] and Kernel Methods (KM) [Scholkopf, Smola, 2002] approaches, which produce non-linear and non-parametric versions of conventional identification algorithms.

Kernel identification methods, based on the idea of input data implicit nonlinear transformation into high-dimensional (theoretically infinite) feature space, ensure the possibility of complex nonlinear model high quality approximation. Using the Mercer's theorem, the feature vectors are chosen so that its scalar products in feature space are the positive definite kernel functions. At that the identified model may be represented in a nonparametric form as linear combination of kernel function, though the weighting coefficients (auxiliary variables)

may be computed without making direct reference to feature vectors ("kernel trick") [Scholkopf, Smola, 2002]. Such an idea is proved to be very effective for non-linear identification [Espinoza, Suykens, De Moor, 2005].

In the initial version of Kernel Method the dimension of auxiliary variables vector as well as estimated model complexity increases proportionally to the training sample length, which make it unsuitable for on-line application. In order the model complexity restriction and simplification of computations, it is desirable to use the auxiliary vector of fixed dimension along with recurrent version of identification algorithm [Kivinen, Smola, Williamson, 2004]. Moreover, the recurrent on-line learning successfully used for non-stationary time-series identification.

In [Suykens, Van Gestel, De Brabanter, De Moor, Vandewalle, 2002] the SVM approach to Recursive Least-Square Kernel Method (RLSKM) has been considered. Sequential sparsification procedure was proposed in [Engel, Mannor, Meir, 2004], which may be viewed as a form of regularization and ensure the restriction of the rate of model complexity increasing. In this way the resulting RLSKM algorithm reduces the order of the feature space. Another approach, known as a Sliding Window Kernel Method (SWKM) [Vaerenbergh, Javier, Santamar, 2007], used at any time instant only fixed size subset of training sample.

In this paper a unified approach to recurrent kernel identification algorithms design is proposed. At first, we consider a recurrent modification of initial KM with “growing” windows. In order to fix the auxiliary vector dimension, the reduced order model KM is proposed and proper recurrent identification algorithms are designed. Finally, the full recurrent version of SWKM regarding to auxiliary variables, is also developed along with suitable sliding kernel matrix updating algorithms.

Problem Statement

Consider the time-series $\{x_k\}_{k=1}^{n}$, generated by nonlinear discrete dynamic system

$$x_{k+1} = f(x_k) + \varepsilon_k \ , k = 0, \ldots \tag{1}$$

where $f(\cdot)$ is unknown nonlinear function and ε_k is a noise discrete random process, $\mathbf{E}\{\varepsilon_k\} = 0,\ \mathbf{E}\{\varepsilon_k^2\} = \sigma^2$.

The problem is the nonlinear time-series (1) identification and prediction using the observed sequence $\{y_k, x_k\}_{k=0}^{n}$, where $y_k = x_{k+1}$. In the KM framework the parameterized time-series identification model is:

$$y_k = \hat{f}(x_k) = \varphi^{\mathrm{T}}(x_k)w + \varepsilon_k, \quad k = \overline{0,n}, \tag{2}$$

where $\hat{f}(x)$ - model approximation, $\varphi : \mathbf{R}^1 \to \mathbf{R}^{\mathrm{M}}$ is a nonlinear feature map, which transform the original inputs into high-dimensional feature vector $\varphi(x) \in \mathbf{R}^{\mathrm{M}}$, and $\mathbf{w} \in \mathbf{R}^{\mathrm{M}}$ is an unknown coefficient vector.

Equation (2) may be represented in matrix form as $\mathbf{y}_n = \mathbf{\Phi}_{n-1}^{\mathrm{T}}\mathbf{w} + \boldsymbol{\varepsilon}_n$, where $\mathbf{y}_n = (y_0\ y_1 \ldots y_n)^{\mathrm{T}}$ - observation vector, $\boldsymbol{\varepsilon}_n = (\varepsilon_0\ \varepsilon_1 \ldots \varepsilon_n)^{\mathrm{T}}$ - noise vector, and feature matrix $\mathbf{\Phi}_{n-1} = \big(\varphi(x_0) \quad \varphi(x_1) \quad \ldots \quad \varphi(x_{n-1})\big)$.

In according with Mercer’s theorem, feature vectors are taken hereby that its scalar products in feature space will be positive definite kernel functions $\varphi^{\mathrm{T}}(x_i)\varphi(x_j) = \kappa(x_i, x_j),\ i, j = \overline{1,n}.$ Commonly used polynomial kernels $\kappa(x, x') = (\mu + x \cdot x')^p$ of degree p or Gaussian kernels $\kappa(x, x') = \exp\{-\mu(x - x')^2\}$, where p, μ – tuning parameters of kernel model.

In accordance with KM technique introduce further the kernel matrix $\mathbf{K}_n = \mathbf{\Phi}_n^{\mathrm{T}}\mathbf{\Phi}_n$, which may be computed directly without reference to the feature vectors, because $\mathbf{K}_n = \left\| k_{i,j} \right\|, k_{i,j} = \kappa(x_i, x_j), i, j = \overline{1,n},$ and also taking into consideration the auxiliary (dual) variables vector $\boldsymbol{\lambda}_n \in \mathbf{R}^n$, such as $\mathbf{w}_n = \mathbf{\Phi}_{n-1}\boldsymbol{\lambda}_n$. At that the nonlinear time-series (1) model estimation and one step prediction may be represented as

$$\hat{x}_{k+1} = \hat{f}(x_n) = \varphi^{\mathrm{T}}(x_n)\boldsymbol{\Phi}_{n-1}\mathbf{w}_n = \mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\boldsymbol{\lambda}_n, \quad (3)$$

where $\hat{x}_{k+1}$ - one step ahead time-series prediction, $\mathbf{k}_{n-1}^{\mathrm{T}}(x_n) = (\kappa(x_n, x_0)\ \kappa(x_n, x_1)\ \ldots\ \kappa(x_n, x_{n-1}))$ - kernel vector and $\boldsymbol{\lambda}_n$ is an auxiliary variables vector estimate at instant k, which should be obtained by the training sample $\{y_k, x_k\}_{k=0}^{n-1}$, and in accordance with "kernel trick" express in terms of only kernel matrix $\mathbf{K}_{n-1}$.

The purpose of this paper is the recurrent KM identification algorithms design, which ensures on-line $\boldsymbol{\lambda}_n$ estimates. We will consider the following alternatives of recurrent KM identification:

– *Recurrent KM identification with growing window* $\boldsymbol{\lambda}_{n+1} = F(\boldsymbol{\lambda}_{n+1}, y_{n+1}, \mathbf{K}_n)$, which use the complete training sample $\{y_k, x_k\}_{k=0}^{n}$ for dual variable $\boldsymbol{\lambda}_n$ estimation.

– *Recurrent KM reduced order model identification* $\widetilde{\boldsymbol{\lambda}}_{n+1} = F(\widetilde{\boldsymbol{\lambda}}_{n+1}, y_{n+1}, \widetilde{\mathbf{K}}_{n,r})$, which also use the complete training sample but with respect to fix dimension dual variables $\widetilde{\boldsymbol{\lambda}}_n \in \mathbf{R}^r$ and kernel matrix $\overline{\mathbf{K}}_{n,r}$.

– *Recurrent KM identification with sliding widow* $\overline{\boldsymbol{\lambda}}_{n+1} = F(\overline{\boldsymbol{\lambda}}_{n+1}, y_{n+1}, \overline{\mathbf{K}}_{n,s}), \overline{\boldsymbol{\lambda}}_n \in \mathbf{R}^{\mathrm{s}}$, which uses the sliding window training sample $\{y_k, x_k\}_{k=n-s}^{n}$ and kernel matrix $\overline{\mathbf{K}}_{n,s}$, built up on the respective observations.

Recurrent Kernel Identification with Growing Window

In accordance with general SVM approach [Vapnik, 1998] nonlinear time-series (1) identification problem using the complete training sample (growing window) may be reduced to the following constrained optimization problem with regularized estimation cost function with regularization parameter $\gamma > 0$:

$$J_n(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}\gamma \cdot \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} \to \min_{\mathbf{w},\boldsymbol{\varepsilon}}, \quad \mathbf{y}_n = \boldsymbol{\Phi}_{n-1}^{\mathrm{T}}\mathbf{w} + \boldsymbol{\varepsilon}_n. \quad (4)$$

The optimization problem (4) is solved using Lagrange function with dual multipliers $\boldsymbol{\lambda} \in \mathbf{R}^n$:

$$L_n(\mathbf{w}, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}\gamma \cdot \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{y}_n - \boldsymbol{\Phi}_{n-1}^{\mathrm{T}}\mathbf{w} - \boldsymbol{\varepsilon}_n). \quad (5)$$

In such a way, using well-known condition of optimality, the solution may be obtained in the explicit form $\mathbf{w}_n = \boldsymbol{\Phi}_{n-1}\boldsymbol{\lambda}_n$, $\boldsymbol{\varepsilon}_n = \gamma^{-1}\boldsymbol{\lambda}_n$, therefore the dual variables estimate $\boldsymbol{\lambda}_n$ takes the form of ridge regression:

$$\boldsymbol{\lambda}_n = \left(\gamma^{-1}\mathbf{I}_n + \mathbf{K}_{n-1}\right)^{-1}\mathbf{y}_n = \mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{y}_n, \quad (6)$$

where $\mathbf{I}_n$ is an identity $n \times n$ matrix, and $\mathbf{K}_{n-1}(\gamma)$ is a regularized kernel matrix..

The recurrent estimation for dual variables $\boldsymbol{\lambda}_{n+1} = \mathbf{K}_n^{-1}(\gamma)\mathbf{y}_{n+1}$ at instant $n+1$ may be easily obtained using Sherman-Morrison-Woodbury formula [Golub, Van Loan, 1998] for the regularized kernel matrix $\mathbf{K}_n(\gamma)$:

$$\mathbf{K}_n(\gamma) = \left(\begin{array}{c|c} \mathbf{K}_{n-1}(\gamma) & \mathbf{k}_{n-1}(x_n) \\ \hline \mathbf{k}_{n-1}^{\mathrm{T}}(x_n) & \gamma^{-1} + k_{n,n} \end{array}\right), \quad (7)$$

$$\mathbf{K}_n^{-1}(\gamma) = \left(\begin{array}{c|c} \mathbf{K}_{n-1}^{-1}(\gamma) + \delta_n^{-1}\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n)\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1}^{-1}(\gamma) & -\delta_n^{-1}\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n) \\ \hline -\delta_n^{-1}\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1}^{-1}(\gamma) & \delta_n^{-1} \end{array}\right), \quad (8)$$

where $\delta_n = \gamma^{-1} + k_{n,n} - \mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n)$.

Consequently the recurrent algorithm for dual variables estimates may be represented as

$$\boldsymbol{\lambda}_{n+1} = \begin{pmatrix} \boldsymbol{\lambda}_n - \delta_n^{-1}[y_{n+1} - \omega_n(\boldsymbol{\lambda}_n)]\mathbf{K}_{n-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n) \\ \delta_n^{-1}[y_{n+1} - \omega_n(\boldsymbol{\lambda}_n)] \end{pmatrix}, \tag{9}$$

where $\omega_n(\boldsymbol{\lambda}_n) = \mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\boldsymbol{\lambda}_n$, along with regularized kernel matrix $\mathbf{K}_n(\gamma)$ updating procedure (8).

As a result the time-series model estimate and one-step ahead prediction may be obtained as (3). It is obvious that the dimension of dual variables estimates $\boldsymbol{\lambda}_n$ and thereafter the estimated model $\hat{f}(x)$ complexity grows proportionally to the training sample length, which make the considered scheme unsuitable for on-line application.

Reduced Kernel Model Recurrent Identification

From the computational point of view for on-line application it is desirable to limit the number of data vectors from which the kernel matrix is calculated. It allows both to reduce the order of the feature space (which prevents overfitting) and to keep the complexity of model bounded. In order to limit the size of the kernel matrix a sparsification process was proposed [Engel, Mannor, Meir, 2004], in which an input sample is only admitted into the kernel matrix if its image in feature space cannot be sufficiently well approximated by combining the previously admitted samples. We consider another approach in which the reduced order model is formed from the pre-established linear independent feature vectors, corresponding to the fixed input vectors. In such a way the size of kernel matrix is fixed in advance so the model complexity doesn't growing under identification process.

Consider the reduced order feature matrix $\widetilde{\boldsymbol{\Phi}}_r$ consist of r constant linear independent basic (support) feature vectors $\widetilde{\boldsymbol{\Phi}}_r = \left(\widetilde{\varphi}(x_1)\ \widetilde{\varphi}(x_2)\ldots\widetilde{\varphi}(x_r)\right)^{\mathrm{T}}$, whish has been initially constructed from pre-established input vectors $\widetilde{x}_i, i = \overline{1,r}$, selected in such a way, that $\mathrm{rang}(\widetilde{\boldsymbol{\Phi}}_r) = r$. In practice, such a "feature" condition may be easily verified using equivalent "kernel" condition $\mathrm{rang}(\widetilde{\mathbf{K}}_r) = r$ using appropriate kernel matrix $\widetilde{\mathbf{K}}_r = \widetilde{\boldsymbol{\Phi}}_r^{\mathrm{T}}\widetilde{\boldsymbol{\Phi}}_r$.

Any feature vector from training sample may be represent as linear combination of basic (support) feature vectors $\varphi(x_i) = \sum_{j=1}^{r} a_{ij}\varphi(\widetilde{x}_j),\ i = \overline{1,n}$, or, in matrix form, as $\boldsymbol{\Phi}_n = \widetilde{\boldsymbol{\Phi}}_r\mathbf{A}_n^{\mathrm{T}}$, where $\mathbf{A}_n = \|a_{ij}\|$ is a matrix of corresponding expansion coefficients, which may be obtained by the minimum least-squares approximation :

$$\Delta_{\mathbf{A}} = \left\|\boldsymbol{\Phi}_n - \widetilde{\boldsymbol{\Phi}}_r\mathbf{A}_n^{\mathrm{T}}\right\|^2 \to \min_{\mathbf{A}_n}. \tag{10}$$

The solution of (10) may be easily obtained in the explicit form as $\mathbf{A}_n = \widetilde{\mathbf{K}}_{n,r}\widetilde{\mathbf{K}}_r^{-1}$, where $\widetilde{\mathbf{K}}_{n,r} = \boldsymbol{\Phi}_n^{\mathrm{T}}\widetilde{\boldsymbol{\Phi}}_r$. At that the attainable approximation accuracy are determined by $\min\Delta_{\mathrm{A}} = \left\|\mathbf{K}_n - \widetilde{\mathbf{K}}_{n,r}\mathbf{K}_r^{-1}\widetilde{\mathbf{K}}_{n,r}^{\mathrm{T}}\right\|^2$.

Furthermore, the model parameters vector may be express from reduced order dual variables $\widetilde{\boldsymbol{\lambda}}_n \in \mathbf{R}^r$

$$\mathbf{w}_n = \boldsymbol{\Phi}_{n-1}\boldsymbol{\lambda}_n = \widetilde{\boldsymbol{\Phi}}_r\mathbf{A}_{n-1}^{\mathrm{T}}\boldsymbol{\lambda}_n = \boldsymbol{\Phi}_r\widetilde{\boldsymbol{\lambda}}_n, \tag{11}$$

where $\widetilde{\boldsymbol{\lambda}}_n = \mathbf{A}_{n-1}^{\mathrm{T}}\boldsymbol{\lambda}_n$. Corresponding reduced order identified model and prediction function are

$$\hat{x}_{n+1} = \hat{f}_n(x_n) = \varphi^{\mathrm{T}}(x_n)\boldsymbol{\Phi}_r\widetilde{\boldsymbol{\lambda}}_n = \mathbf{k}_r^{\mathrm{T}}(x_n)\widetilde{\boldsymbol{\lambda}}_n,\ \mathbf{k}_r(x_n) = \left(\kappa(x_n,\widetilde{x}_1)\ldots\kappa(x_n,\widetilde{x}_r)\right)^{\mathrm{T}}. \tag{12}$$

In such a way reduced order vector of dual variables estimate $\widetilde{\boldsymbol{\lambda}}_n$ may be obtained via SVM approach as a solution of suitable optimization problem.

1. *Non-regularized case* $\gamma^{-1}=0$. Using representation (11), the appropriate cost function takes the form

$$J_n=\frac{1}{2}\left\|\mathbf{y}_n-\mathbf{\Phi}_{n-1}^{\mathrm{T}}\mathbf{w}\right\|^2=\frac{1}{2}\left\|\mathbf{y}_n-\mathbf{A}_{n-1}\widetilde{\mathbf{\Phi}}_r^{\mathrm{T}}\widetilde{\mathbf{\Phi}}_{n-1}\widetilde{\boldsymbol{\lambda}}\right\|^2\to\min_{\widetilde{\boldsymbol{\lambda}}}. \tag{13}$$

The solution of (12) is $\widetilde{\boldsymbol{\lambda}}_n=\left(\mathbf{A}_{n-1}\widetilde{\mathbf{K}}_r\right)^{+}\mathbf{y}_n$, where "+" denotes Moor-Penrose generalized inversion. Taking into account that $\mathbf{A}_{n-1}=\widetilde{\mathbf{K}}_{n-1,r}\widetilde{\mathbf{K}}_r^{-1}$, due to the generalized inversion properties, the reduced order dual variables estimate may be obtained as $\widetilde{\boldsymbol{\lambda}}_n=\widetilde{\mathbf{K}}_r^{-1}\left(\mathbf{A}_{n-1}\right)^{+}\mathbf{y}_n=\left(\widetilde{\mathbf{K}}_{n-1,r}\right)^{+}\mathbf{y}_n$.

In the following way a recurrent algorithm for $\tilde{\boldsymbol{\lambda}}_{n+1}=\left(\tilde{\mathbf{K}}_{n,r}\right)^{+}\mathbf{y}_{n+1}$ updating may be easily derived.

As far as $\widetilde{\mathbf{K}}_{n,r}^{\mathrm{T}}=\left(\widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}}\vdots\mathbf{k}_n(x_n)\right)$, one can use the known Greville formula [Ben-Israel, Greville, 2003] for reduced order Moor-Penrose inverse kernel matrix updating:

$$\widetilde{\mathbf{K}}_{n,r}^{+}=\left((\mathbf{I}_r-\mathbf{q}_n\mathbf{k}_n^{\mathrm{T}}(x_n))\widetilde{\mathbf{K}}_{n-1,r}^{+}\vdots\mathbf{q}_n\right) \tag{14}$$

where $\mathbf{q}_n=\left(\alpha_n+\mathbf{K}_r^{\mathrm{T}}(x_n)\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})\mathbf{k}_r(x_n)\right)^{-1}\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})\mathbf{k}_r(x_n)$, $\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})=\mathbf{I}_r-\widetilde{\mathbf{K}}_{n-1,r}^{+}\widetilde{\mathbf{K}}_{n-1,r}$,

$\alpha_n=1-\mathrm{sgn}\left(\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{Z}(\widetilde{\mathbf{K}}_{n-1,r})\mathbf{k}_r(x_n)\right)$

Thereafter recurrent identification algorithm for dual variables estimates is

$$\widetilde{\boldsymbol{\lambda}}_{n+1}=\widetilde{\boldsymbol{\lambda}}_n+\mathbf{q}_n(\mathbf{y}_{n+1}-\mathbf{k}_r^{\mathrm{T}}(x_n)\widetilde{\boldsymbol{\lambda}}_n). \tag{15}$$

The proper initial conditions at instant $n=r$ are $\widetilde{\mathbf{K}}_{r,r}^{+}=\widetilde{\mathbf{K}}_{r,r}^{-1}$, $\widetilde{\mathbf{K}}_{r,r}=\widetilde{\mathbf{\Phi}}_r^{\mathrm{T}}\widetilde{\mathbf{\Phi}}_r$.

2. *Regularized case* $\gamma>0$. Using the introduced representation for unknown model parameters vector $\mathbf{w}$, the regularized estimation cost function for reduced order model will be taken in the form:

$$J_n=\frac{1}{2}\left\|\mathbf{y}_n-\mathbf{\Phi}_{n-1}^{\mathrm{T}}\mathbf{w}\right\|^2+\gamma^{-1}\mathbf{w}^{\mathrm{T}}\mathbf{w}=\frac{1}{2}\left\|\mathbf{y}_n-\widetilde{\mathbf{K}}_{n-1,r}\widetilde{\boldsymbol{\lambda}}\right\|^2+\gamma^{-1}\widetilde{\boldsymbol{\lambda}}^{\mathrm{T}}\widetilde{\mathbf{K}}_r\widetilde{\boldsymbol{\lambda}}\to\min_{\widetilde{\boldsymbol{\lambda}}}. \tag{16}$$

The explicit solution is $\widetilde{\boldsymbol{\lambda}}_n=\mathbf{P}_{n-q,r}^{-1}\mathbf{K}_{n-1,r}^{\mathrm{T}}\mathbf{y}_n$, where $\mathbf{P}_{n-q,r}=\gamma^{-1}\widetilde{\mathbf{K}}_r+\widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}}\widetilde{\mathbf{K}}_{n-1,r}$.

As far as $\widetilde{\mathbf{K}}_{n,r}^{\mathrm{T}}\widetilde{\mathbf{K}}_{n,r}=\widetilde{\mathbf{K}}_{n-1,r}^{\mathrm{T}}\widetilde{\mathbf{K}}_{n-1,r}+\mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n)$, the recurrent form for reduced order dual vector estimate $\widetilde{\boldsymbol{\lambda}}_{n+1}=\mathbf{P}_{n,r}^{-1}\widetilde{\mathbf{K}}_{n,r}\mathbf{y}_{n+1}$ may be represented as

$$\widetilde{\boldsymbol{\lambda}}_{n+1}=\widetilde{\boldsymbol{\lambda}}_n+\gamma_n\mathbf{P}_{n-1,r}^{-1}\mathbf{k}_r(x_n)\left(\mathbf{y}_{\mathrm{n}+1}-\mathbf{k}_r^{\mathrm{T}}(x_n)\widetilde{\boldsymbol{\lambda}}_n\right),\quad\gamma_n=\left(1+\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{P}_{n-1,r}^{-1}\mathbf{k}_r(x_n)\right)^{-1}. \tag{17}$$

Using the matrix inversion lemma [Haykin, 1996], the recurrent procedure for inverse matrix $\mathbf{P}_{n,r}=\mathbf{P}_{n-1,r}+\mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n)$ updating takes the following form:

$$\mathbf{P}_{n,r}^{-\backprime1}=\mathbf{P}_{n-1,r}^{-\backprime1}-\gamma_n^{-1}\mathbf{P}_{n-1,r}^{-\backprime1}\mathbf{k}_r(x_n)\mathbf{k}_r^{\mathrm{T}}(x_n)\mathbf{P}_{n-1,r}^{-\backprime1}. \tag{18}$$

Thereby equations (15), (16) may be treated as RLSKM version for nonlinear time-series identification based on reduced order model.

Recurrent Kernel Identification with Sliding Window

Sliding window KM approach consider for estimation at instant n only last $n-s$ observations, so observation vector is $\mathbf{y}_{n,s}=(y_{n-s+1}\ldots y_n)^{\mathrm{T}}$. Consequently, the kernel matrix $\overline{\mathbf{K}}_{n,s}=\mathbf{\Phi}_{n,s}^{\mathrm{T}}\mathbf{\Phi}_{n,s}$ has a fixed dimension $(s\times s)$ and observation equation takes the form $\mathbf{y}_{n,s}=\mathbf{\Phi}_{n-1,s}^{\mathrm{T}}\mathbf{w}+\boldsymbol{\varepsilon}_{n,s}=\mathbf{K}_{n-1,s}\overline{\boldsymbol{\lambda}}_n+\boldsymbol{\varepsilon}_{n,s}$, where $\overline{\boldsymbol{\lambda}}_n\in\mathbf{R}^{\mathrm{s}}$.

Consider the "sliding" estimation cost function includes at any instant $n+1$ *a priori* information term determined by previously estimate at instant n. Using the representation $\mathbf{w}=\mathbf{\Phi}_{n,s}\overline{\boldsymbol{\lambda}}$, the optimization problem is defined as:

$$J_{n,s}=\left\|\mathbf{y}_{n+1,s}-\mathbf{K}_{n,s}\overline{\boldsymbol{\lambda}}\right\|^2+\gamma^{-1}(\overline{\boldsymbol{\lambda}}-\overline{\boldsymbol{\lambda}}_n)^{\mathrm{T}}\mathbf{K}_{n,s}(\overline{\boldsymbol{\lambda}}-\overline{\boldsymbol{\lambda}}_n)\to\min_{\overline{\boldsymbol{\lambda}}}. \tag{19}$$

Condition of optimality leads to the following normal equations:

$$\left(\overline{\mathbf{K}}_{n,s}^{\mathrm{T}}\overline{\mathbf{K}}_{n,s}+\gamma^{-1}\overline{\mathbf{K}}_{n,s}\right)\overline{\boldsymbol{\lambda}}_n=\overline{\mathbf{K}}_{n,s}^{\mathrm{T}}\mathbf{y}_{n+1,s}+\gamma^{-1}\overline{\mathbf{K}}_{n,s}\overline{\boldsymbol{\lambda}}_n. \tag{20}$$

Using the identity $\mathbf{A}^{-1}\left(\gamma^{-1}\mathbf{I}_s+\mathbf{A}\right)^{-1}\mathbf{A}=\left(\gamma^{-1}\mathbf{I}_s+\mathbf{A}\right)^{-1}$, the recurrent dual vector estimate takes the form:

$$\overline{\boldsymbol{\lambda}}_{n+1}=\left(\gamma^{-1}\mathbf{I}_s+\overline{\mathbf{K}}_{n,s}\right)^{-1}\left(\gamma^{-1}\overline{\boldsymbol{\lambda}}_n+\mathbf{y}_{n+1}\right) \tag{21}$$

At last, it is necessary to put forward the updating algorithm for inverse regularized sliding kernel matrix $\overline{\mathbf{K}}_{n,s}^{-1}(\gamma)$. Using the approach, proposed by [Vaerenbergh, Javier, Santamar, 2007], consider two step inverse regularized kernel matrix updating algorithm $\overline{\mathbf{K}}_{n-1,s}^{-1}(\gamma)\to\overline{\mathbf{K}}_{n-1,s-1}^{-1}(\gamma)\to\overline{\mathbf{K}}_{n,s}^{-1}(\gamma)$, which use auxiliary "downsizing" matrix $\overline{\mathbf{K}}_{n-1,s-1}(\gamma)$, determined from the sliding kernel matrix representation:

$$\overline{\mathbf{K}}_{n-1,s}(\gamma)=\left(\begin{array}{c|c}\gamma^{-1}+k_{n-s,n-s} & \mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_{n-s})\\ \hline \mathbf{k}_{n-1,s-1}(x_{n-s}) & \overline{\mathbf{K}}_{n-1,s-1}(\gamma)\end{array}\right), \tag{22}$$

where kernel vector $\mathbf{k}_{n-1,s-1}(x_{n-s})=\left(\kappa_{n-1}(x_{n-s})\ \ldots\ \kappa_{n-s+1}(x_{n-s})\right)^{\mathrm{T}}$.

Then at 1-st step of the algorithm including the "downsizing" matrix inverse is:

$$\overline{\mathbf{K}}_{n-1,s-1}^{-1}=\mathbf{R}_s\overline{\mathrm{K}}_{n-1,s}^{-1}\mathbf{R}_s^{\mathrm{T}}-(\mathrm{e}_1^{\mathrm{T}}\overline{\mathbf{K}}_{n-1,s}^{-1}\mathrm{e}_1)^{-1}\mathbf{R}_s\overline{\mathbf{K}}_{n-1,s}^{-1}\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}\overline{\mathbf{K}}_{n-1,s}^{-1}\mathbf{R}_s^{\mathrm{T}}, \tag{23}$$

where $\mathbf{R}_s=\left(0_s\vdots\mathbf{I}_{s-1}\right)$, $\mathbf{e}_1=(1\ldots0)^{\mathrm{T}}$.

Using the regularized sliding kernel matrix $\mathbf{K}_{n,s}^{-1}(\gamma)$ representation

$$\overline{\mathbf{K}}_{n,s}(\gamma)=\left(\begin{array}{c|c}\overline{\mathbf{K}}_{n-1,s-1}(\gamma) & \mathbf{k}_{n-1,s-1}(x_n)\\ \hline \mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n) & \gamma^{-1}+k_{n,n}\end{array}\right), \tag{24}$$

where $\mathbf{k}_{n-1,s-1}(x_n)=\left(\kappa_{n-1}(x_n)\ \ldots\ \kappa_{n-s+1}(x_n)\right)^{\mathrm{T}}$, the 2-nd step of $\overline{\mathbf{K}}_{n,s}^{-1}(\gamma)$ updating is the following:

$$\mathbf{K}_{n,s}^{-1}=\left(\begin{array}{c|c}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)+\delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1}(x_n)\mathbf{k}_{n-1}^{\mathrm{T}}(x_n)\mathbf{K}_{n-1,s-1}^{-1}(\gamma) & -\delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n)\\ \hline -\delta_n^{-1}\mathbf{K}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n) & \delta_n^{-1}\end{array}\right), \tag{25}$$

where $\delta_n=\gamma^{-1}+k_{n,n}-\mathbf{k}_{n-1,s-1}^{\mathrm{T}}(x_n)\overline{\mathbf{K}}_{n-1,s-1}^{-1}(\gamma)\mathbf{k}_{n-1,s-1}(x_n)$.

Finally, expressions (21), (23), (25) produce a recurrent form of SWKM for nonlinear time-series identification.

Conclusion

Recurrent KM approach for nonlinear time-series identification and prediction combining with model reduction technique leads to identification algorithms efficiency improvement. The advantage of such an approach consists not only in computing difficulties reducing and amount of calculation restriction but also in the possibility of on-line operating in non-stationary environments. The key feature of proposed algorithms is that the identified model complexity does not increase as the number of samples increases and time-varying model may be on-line estimated, so recurrent KM algorithms may be successfully used for non-stationary time series identification. Another preference connected with the sufficiently simple possibility of robust modification recurrent KM algorithms design using suitable nonlinear estimation function.

Further inquiry and improvement of proposed approach should be connected with identified model optimization methods development. The most important problem is the model parameters optimization, namely, regularization parameter and kernel tuning parameter. The cross-validation technique is seemed to be the most suitable approach to model parameters optimization. Such issue is closely connected with the general problem of model optimization in compliance with available information via structural risk minimization approach. Dynamic approach to model optimization may be considered as one of the most important directions of further investigation.

Bibliography

[Ljung, 1999] L. Ljung. System Identification: Theory for the User. Prentice Hall, New Jersey, Second edition, 1999.

[Dorffner,1996] G. Dorffner. Neural networks for time series processing // Neural Network World, 1996, 6(4), Pp.447–468.

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

[Scholkopf, Smola, 2002] B. Scholkopf, A. Smola. Learning with kernels. Cambridge, MA: MIT Press, 2002.

[Espinoza, Suykens, De Moor, 2005] M. Espinoza, J.A.K. Suykens, and B. De Moor. Kernel based partially linear models and nonlinear identification // IEEE Transactions on Automatic Control, 2005, 50 (10), Pp. 1602–1606.

[Kivinen, Smola, Williamson, 2004] J. Kivinen, A.J. Smola, R.C. Williamson. Online learning with kernels // IEEE Transactions on Signal Processing, 2004, Vol. 52, Pp. 2165 – 2176.

[Suykens, Van Gestel, De Brabanter, De Moor, Vandewalle,2002] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. Least Squares Support Vector Machines. World Scientific, Singapore, 2002.

[Engel, Mannor, Meir, 2004] Y. Engel, S. Mannor, R. Meir. The kernel recursive least-squares algorithm // IEEE Transactions on Signal Processing, 2004, Vol. 52, Pp. 2275 - 2285.

[Vaerenbergh, Javier, Santamar, 2007] S. Van Vaerenbergh, V. Javier, I. Santamar. Nonlinear System Identification using a New Sliding Window Kernel RLS Algorithm // Journal of Communications, 2007, Vol. 2, No. 3.

[Golub, Van Loan, 1989] G.H. Golub, C.F. Van Loan. Matrix Computations. Johns Hopkins Univ. Press. Baltimore MD, 1989.

[Ben-Israel, Greville, 2003] A. Ben-Israel, T. Greville. Generalized Inverses. Theory and Applications. Springer, 2003.

[Haykin, 1996] S. Haykin. Adaptive Filter Theory. Prentice Hall, 3rd edition, 1996.

Authors' Information

Leonid Lyubchyk *– Head of Department, Department of Computer Mathematics and Mathematical Modeling, National Technical University "Kharkov Polytechnic Institute", Frunze str. 21, Kharkov, 61002, Ukraine; e-mail: lyubchik.leonid@gmail.com*

Vladyslav Kolbasin *– PhD student, Department of Computer Mathematics and Mathematical Modeling, National Technical University "Kharkov Polytechnic Institute", Frunze str. 21, Kharkov, 61002, Ukraine; e-mail: ka_vlad@list.ru*

ARES SYSTEM - INTEGRATION OF ANALYSIS METHODOLOGIES

Roman Podraza, Mariusz Kalinowski

***Abstract**: ARES System is an application dedicated to data analysis, data exploration and knowledge discovery. This versatile tool offers rough set based methodology as well as emerging patterns and Support Vector Machine algorithms to be applied to input data presented as an information system. Many analytical capabilities provided with an user-friendly, intuitive graphical interface and XML based support for input/output operations make ARES System a challenging and promising software for application in scientific research and didactic work. An unique feature of ARES System is its ability of identifying improper objects within information system by employing credibility coefficients. The credibility coefficients attempt to assess a degree of typicality of each object in respect to the whole information system.*

***Keywords**: information system, classification, rough sets, emerging patterns, SVM, credibility coefficients.*

***ACM Classification Keywords**: H.1.1 Systems and Information Theory*

Introduction

ARES System has been persistently developed to extend its potential in data analysis, data exploration and knowledge discovery [Podraza, 2005]. Its initial functionalities were based on rough set theory [Pawlak, 1991] and contained all phases leading to discovering rules from an information system. Results of each step of data analysis can be presented by an intuitive graphical user interface. This feature should enable user to observe the whole process of data analysis and/or exploration to learn how to tune the whole experiment to achieve the most sound results. A number of algorithms of data discretization, finding frequent sets and reducts or extracting rules were implemented with the first release of the system.

From the very beginning ARES System offered credibility coefficients [Podraza, 2007], [Podraza 2006] , which were heuristic measures of objects' typicality in respect to other ones. The main idea of introducing the credibility coefficients could be summarized that rules involve exceptions to them and the exceptions very often can be more interesting that the rules themselves. The evaluation of credibility coefficients is based on observations that typical objects are considered credible, because they appear more frequently and we used to generalize their appearance as something predictable, and describable by rules.

In the development of ARES System there have been introduced algorithms for discovering discriminants of information systems – LEM1, LEM2 and AQ along with required checking of consistency of information systems [Grzymala, 2005]. A next new feature of ARES System enables discovering Emerging Patterns (EP) in objects of decision tables. This is an alternative approach to mine rules from set of data. Support Vector Machine is yet another methodology used to classify data, which has been integrated within the system.

ARES System becomes a common platform for a number of different classification approaches. The same data can be processed in several ways and the results can be immediately compared and examined. Each approach can be parameterized in many aspects giving a really flexible and powerful classification tool. Formats of input/output files have been enriched by flexible XML based enabling universal communication with other different systems. In particular quite large system output like hundreds of rules can be presented in web browser or more elaborated presentation tools.

In the next section a general description of ARES System functionalities can be found. Then follows a section covering capabilities of three methodologies of data analysis and knowledge acquisition, which have been introduced to ARES System. A brief presentation of system development goes after a short record of the original

system potential. Subsequently credibility coefficients, which are unique feature o ARES System, are concerned. And finally conclusions and further perspectives of ARES System are submitted.

System Overview

The system has been designed to give a full interactive access to process of data analyses involving different approaches. A multi-window graphical user interface enables tracking different steps of the processing and/or comparing results of applying different algorithms for the whole procedure of data analysis or for its particular phase. A sample screenshot from ARES System is shown in Fig. 1. The main application window consists of two views. *Directory browser* presents a list of all elements currently available for user (e.g. workspaces, decision tables, analysis results) and *Workspace view* provides a space for windows opened according to user selection. Views of elements from the directory browser can be presented for investigations and comparisons with the others (e.g. window with a decision table, the corresponding frequent sets, a list of all found reducts).

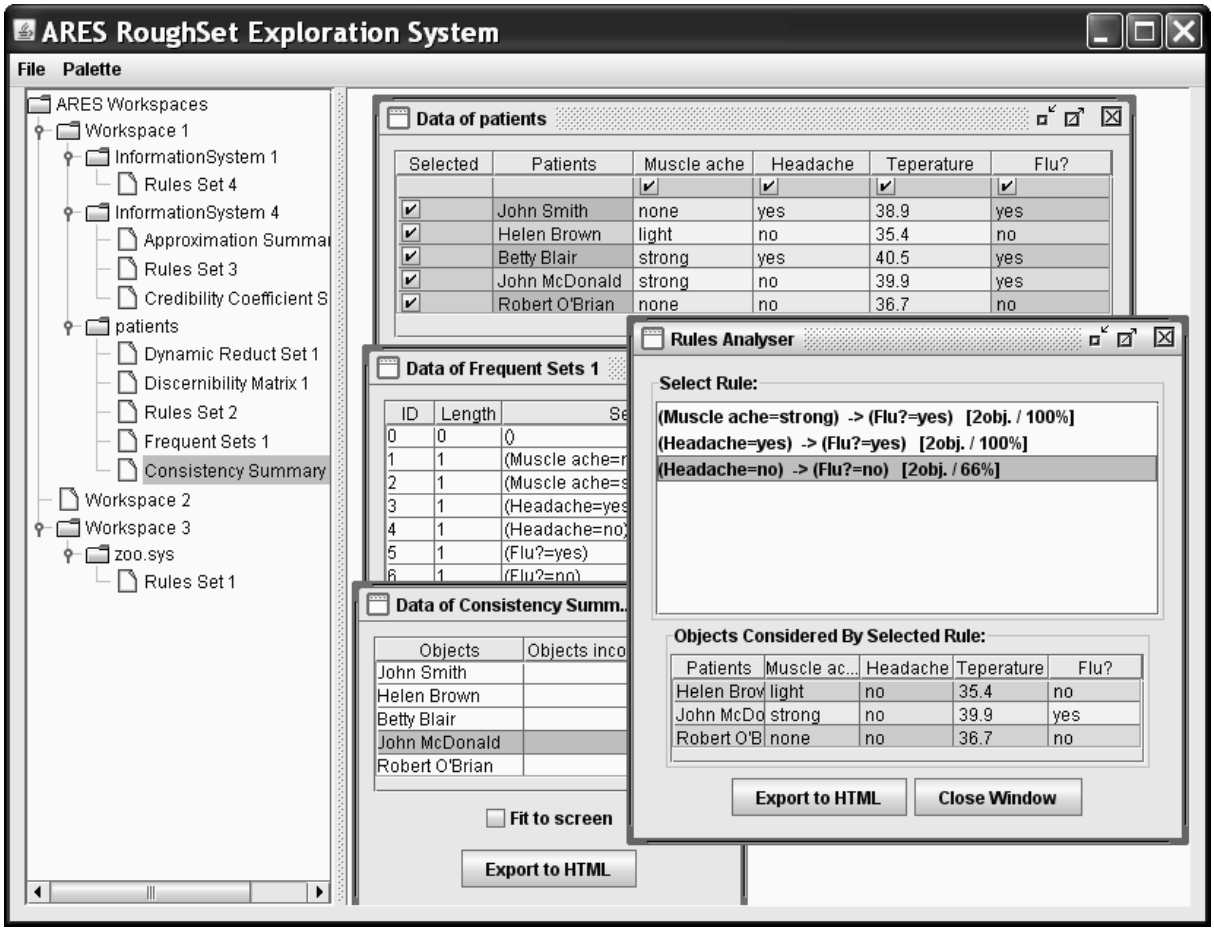


Figure 1. Multi-window GUI of ARES System

A relationship between elements of processed data is reflected by their hierarchical placements within the system directory, however it can be accentuated by naming conventions. Each directory item stores information on its type, its data and description. Item type determines operations, which can be applied to it. Context menu associated with the item (pulled down by the right button of mouse) contains positions for starting the operations. The data and the description can be presented in appropriate windows in the workspace view. The description contains some statistic and explanatory information as execution time used to produce the item, algorithm applied and some other specific data (e.g. number of generated rules for rule set item).

An information system or more precisely a decision table is input to ARES System. The input data can be stored in three formats: numerical one (specific for ARES System only), CSV and XML. Two latter ones give sufficient

flexibility to cooperate with other systems. In the decision table rows represent data objects and columns represent attributes of objects. In ARES System only one attribute can be chosen as a decision. There are checkboxes associated with each row and each columns. They enable cutting a desired decision table from the existing one by removing unchecked rows and/or columns (Fig. 2). It is possible to view the information system in its internal numerical representation as well.

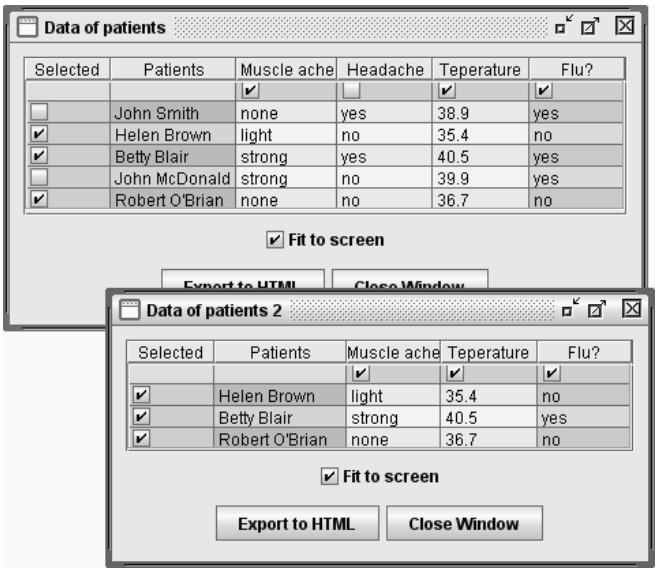


Figure 2. Modifications of a Decision Table

All data displayed in a tabular form in any window of the workspace view can be exported to HTML file. The information format is similar to presented in the window. This option allows saving the results of ARES System, present them and print in internet browsers and/or publish in formats accepting HTML.

Integration of Platforms

The system has been developed to comprise three methodologies of knowledge acquisition. To original contents based on Rough Set theory [Pawlak, 1991] capabilities of Emerging Patterns [Dong, 1999] approach and Support Vector Machines [Schmilovici, 2005] methodology have been appended to the system. Rough Set approach has been extended as well by adding algorithms for discovering discriminant of information system. SVM algorithms have already been implemented and user interface for them is just under construction, and currently only credibility coefficients based on SVM are fully operational. The original ARES System [Podraza 2005] has comprised modules for performing the following tasks form rough set sphere such as

- Discretizing continuous domains of objects' attributes
- Discovering approximations of decision classes
- Determining discernibility matrices
- Finding relative reducts by applying set of algorithms to calculate
    - all reducts
    - minimal reducts
- Discovering frequents sets
- Mining decision rules

All these tasks can be performed by a number of algorithms. The operations are selected for particular items from the directory browser. Each such item has a number of operations applicable to it. They are all grouped in a context menu pulled down by the right button of mouse while selecting the item. A collection of operations applicable to directory item representing a set of rules is presented in Tab. 1. Usually results of each such

operation is presented in a new window in the workspace area and sometimes a new item representing the result is inserted to the directory. There is a possibility to export the result into HTML file. Particular operation can be performed by a number of algorithms. The choice and required parameters are determined in a interactive way.

Table 1. List of commands (from context menu) for item representing set of rules

Command	Description
Close Rules Miner	Deletes the rules set
Show Data	Displays a window with rules and information about them (support and confidence of given rules)
Show Properties	Displays a window with information about given rules set (name and parameters of the rule mining algorithm, count of rules, time of mining)
Show Disjunctive Rules Set (AQ)	Displays a window with disjunctive rules (generated by the AQ algorithm)
Objects Coverage	Displays a window presenting objects coverage
Rules Coverage	Displays a window presenting rules coverage
Rules Analyser	Displays a window of a rules analyser
Show All Rules	Displays a window with all rules
Show Certain Rules	Displays a window with certain rules
Show Possible Rules	Displays a window with possible rules
Close Menu	Closes the menu

In implementation of ARES System many known procedures of rough set area have been instantiated and some updates have been formulated. There are some analyzing tools to present relationships discovered between results and objects of original decision table. In Fig. 3 there is a window showing objects from decision table supporting antecedent of a selected rule.

There have been a number of algorithms for calculating credibility coefficients for objects of decision table. Module of credibility coefficients is very special for ARES System because it allows for systematic treatment of exceptional cases and the next section is devoted to this subject in a systematic way.

A domain of rough set theory in ARES System has been supplemented by module for discovering discriminant of information system. The module comprises three algorithms LEM1, LEM2 and AQ [Grzymala, 2005]. The last algorithm generates rules with disjunctive representation. There is a possibility to check consistency of information system, which is a necessary condition to calculate a discriminant of information system. If the information system is inconsistent (there are at least two objects, which have the same values for all conditional attributes and have different values of the decision attribute) then the operation results in a report highlighting objects causing this inconsistency.

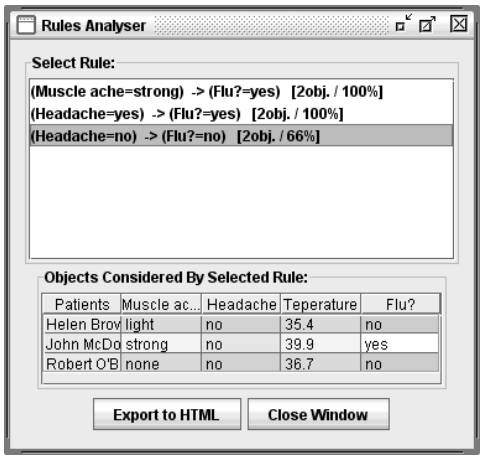


Figure 3. Window presenting relationship between antecedent of selected rule and objects from decision table.

The next platform in ARES System is the KTDA system [Podraza 2007], [Podraza, 2006a] based on Emerging Patterns (EP) approach. The KTDA system was designed and implemented as an independent platform and then was integrated with ARES System. To distinguish two decision classes it is desirable to find out such patterns which are frequent itemsets in one class and are infrequent in the other one. These patterns are just called emerging patterns. The ratio of the pattern support in its target class to the pattern support in the rest of the dataset is a *growth rate* for this pattern. Larger values of the growth rate denote more characteristic EPs for its target class.

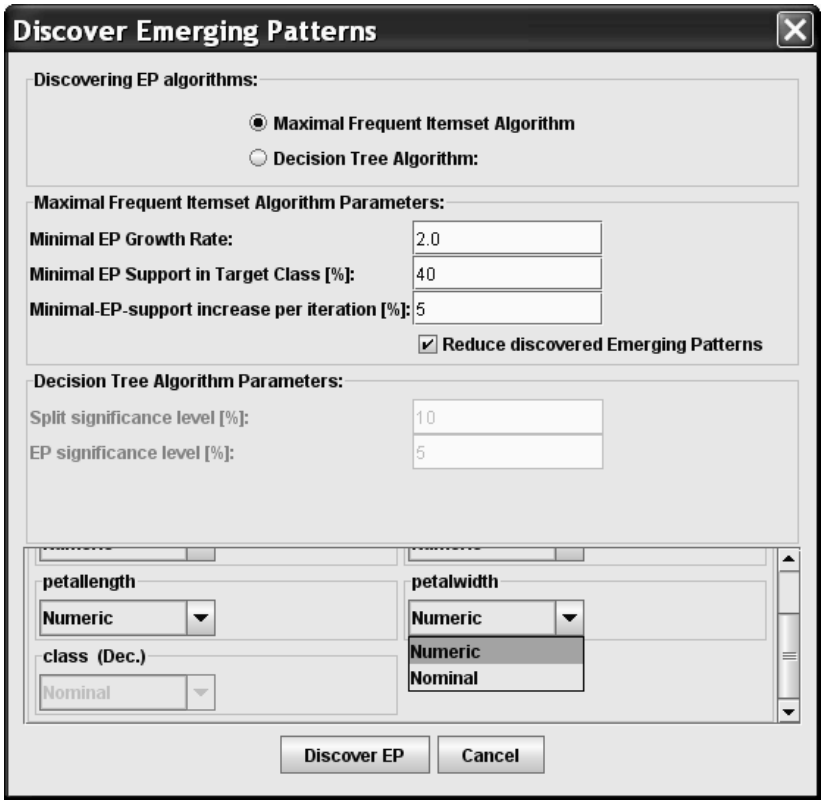
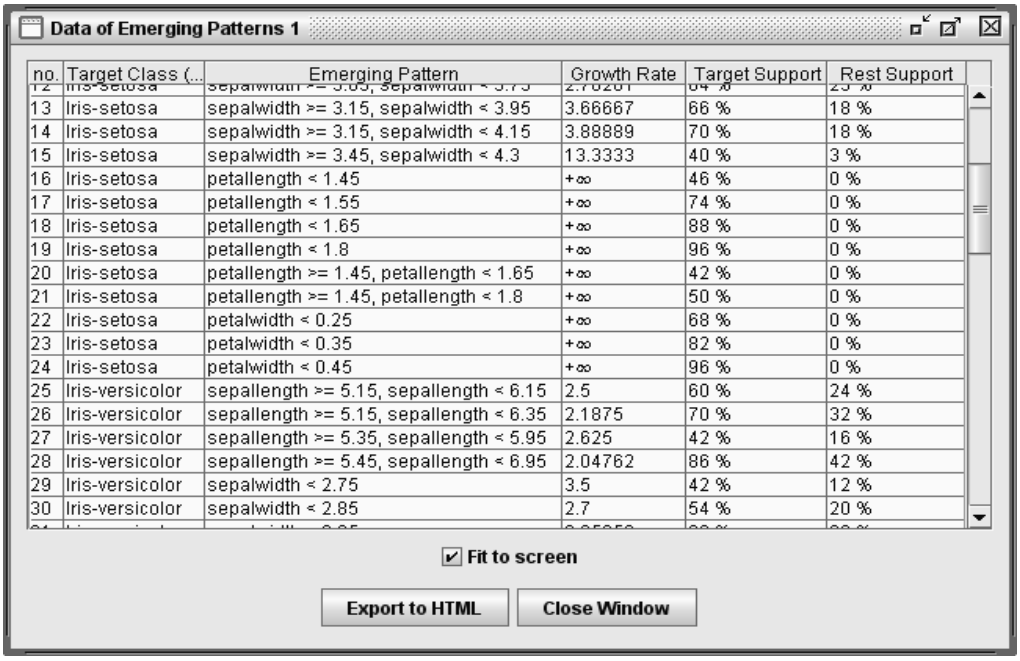


Figure 4. A choice of algorithms for discovering emerging patterns.

The KTDA system implements two different algorithms of discovering EPs – using maximal frequent itemsets proposed in [Dong, 1999] and using decision tree [Boulesteix, 2003] , but with some extensions and improvements. The former one reflects the classical approach and requires stating minimal growth rate and minimal support in the target class, while the latter one uses Fisher's Exact Test used to discover only such EPs which are statistically significant. Algorithm using decision tree is quicker one, produces smaller set of EPs, however all of them are statistically significant. There are default parameters provided for both algorithms and some preliminary research shows, that except minimal growth rate for maximal frequent itemsets algorithms the other parameters have limited impact on set of discovered EPs. In Fig. 4 there is a window for choosing algorithm discovering EPs and setting the appropriate parameters and in Fig 5 there is presented set of EPs revealed by algorithm using maximal frequent itemsets.

EPs enable data classification for which CAEP (Classification by Aggregating Emerging Patterns) algorithm [Dong, 1999a] is applied. For this classifier set of EPs discovered by maximal frequent itemsets algorithms gives usually slightly better classification. On the other hand algorithm using decision tree produces more compact and more significant knowledge, which probably can be more interesting for expert trying to update his/her knowledge on the analyzed problem.

Support Vector Machines is yet another methodology being integrated to ARES System. Currently there are attempts to expose results of classification of data done with this approach. Only credibility coefficients calculated for each object from information system are available from ARES System now, but their calculations involve the classification itself. The credibility coefficients are presented in the following section.



no.	Target Class (...	Emerging Pattern	Growth Rate	Target Support	Rest Support
13	Iris-setosa	sepalwidth >= 3.15, sepalwidth < 3.95	3.66667	66 %	18 %
14	Iris-setosa	sepalwidth >= 3.15, sepalwidth < 4.15	3.88889	70 %	18 %
15	Iris-setosa	sepalwidth >= 3.45, sepalwidth < 4.3	13.3333	40 %	3 %
16	Iris-setosa	petallength < 1.45	+∞	46 %	0 %
17	Iris-setosa	petallength < 1.55	+∞	74 %	0 %
18	Iris-setosa	petallength < 1.65	+∞	88 %	0 %
19	Iris-setosa	petallength < 1.8	+∞	96 %	0 %
20	Iris-setosa	petallength >= 1.45, petallength < 1.65	+∞	42 %	0 %
21	Iris-setosa	petallength >= 1.45, petallength < 1.8	+∞	50 %	0 %
22	Iris-setosa	petalwidth < 0.25	+∞	68 %	0 %
23	Iris-setosa	petalwidth < 0.35	+∞	82 %	0 %
24	Iris-setosa	petalwidth < 0.45	+∞	96 %	0 %
25	Iris-versicolor	sepallength >= 5.15, sepallength < 6.15	2.5	60 %	24 %
26	Iris-versicolor	sepallength >= 5.15, sepallength < 6.35	2.1875	70 %	32 %
27	Iris-versicolor	sepallength >= 5.35, sepallength < 5.95	2.625	42 %	16 %
28	Iris-versicolor	sepallength >= 5.45, sepallength < 6.95	2.04762	86 %	42 %
29	Iris-versicolor	sepalwidth < 2.75	3.5	42 %	12 %
30	Iris-versicolor	sepalwidth < 2.85	2.7	54 %	20 %

Figure 5. Emerging patterns produced by algorithm using maximal frequent itemsets

Credibility Coefficients

Calculations of credibility coefficients is a unique feature of ARES System. A credibility coefficient is a heuristic measure, which assesses typicality of a given object in respect to other objects of information system. Value of credibility coefficient ranges from 0 to 1 and lower values denote worse credibility. The concept of credibility coefficients is based on assumption that majority of data is correct. Minority of data can be incorrect or corrupted. The goal of credibility coefficients is to identify this minority by applying different approaches.

Currently in ARES System there is a number of algorithms for calculations of credibility coefficients based on the following concepts:

- Approximation of rough set classes
- Statistics of attribute values
- Hybrid of previous two
- Frequent set
- Extracted Rules
- Voting Classifier (CAEP)
- Support Vector Machines
- Multi Credibility Coefficient

The first five algorithms for calculations of credibility coefficients belong to the original version of ARES System and mostly exploit concepts of rough set theory. They were described elsewhere in details [Podraza 2007] [Podraza, 2006b] . Credibility coefficient based on voting classifier was incorporated as a part of KTDA system. It

takes into account the real classification of the object and vector of weights of votes for classification determined by the classifier. Any voting classifier outcome can be utilized - CAEP and SVM classifiers are used in ARES System.

Values of different credibility coefficients are incomparable, although all belong to interval <0; 1>. There was proposed a new kind of credibility coefficient, namely ordinal credibility coefficient. Ordinal credibility coefficient is associated with any arbitrary chosen "normal" credibility coefficient, whose values are data for the former one. Ordinal credibility coefficient expresses the relative amount of records with credibility coefficients less or equal to the credibility coefficient for this record. Important feature of ordinal credibility coefficient is fact that its and its input counterpart's values introduce the same ordering of data set objects. In other words, ordinal credibility coefficient of a given record defines its relative position within the data set according to ranking given by the value of the basic credibility coefficient. Multi Credibility Coefficient method [Podraza 2007] combines a number of ordinal credibility coefficients to obtain an aggregate outcome. The resulting value is (weighted) average of aggregated ordinal credibility coefficients.

Credibility coefficients are supposed to reveal "improper" data. Quite often it may be extremely important to focus user attention on such exceptional cases. For instance in medical application the case supporting known rules and procedures can be proceed routinely, while exceptions may require extra check-up and treatment. And in practical data analysis exceptions may appear more interesting then the rules themselves.

Conclusion

The paper presents ARES System functional capabilities. All typical stages of data exploration based on the rough set theory can be performed and presented with support of ARES System. The system has been extended by Emerging Patterns approach and Support Vector Machines methodology.

ARES System has its unique characteristics for discovering non-typical objects in information system. Credibility coefficients are used to evaluate object's measure of typicality in respect to the rest of the information systems. Many algorithms for evaluating credibility coefficients are offered.

ARES System was designed to be used in medical application. Medicine and natural sciences appear to be often interested in exceptions more than in rules – a patient, who reacts exceptionally to a routine treatment causes the highest concern of physicians. ARES System's unique feature to recognize exceptional cases by employing credibility coefficients seems valuable to medicine and other natural sciences but, generally, it can be used on any kind of data, e.g. for engineering purposes.

A multi-document architecture of the ARES System allows for detailed analysis of the data exploration process, what makes the system a perfect and easy to use learning tool.

The ARES System has been implemented in Java and is portable. Its architecture enables permanent development by adding new items with appropriate algorithms to the documents presented and processed by the system. The module structure of ARES System makes its development quite obvious – new functional items inherit structure features of the system.

Although ARES System has been designed to allow its permanent development. It is planned to implement ARES System as Service Oriented Architecture (SOA). The server part will contain all services, very often mutually independent, responsible for steps of data analysis. A number of client programs, tailored to user needs, will call the services. This approach should support developing the system by adding new methodologies, updating existing algorithms and testing parts of the system without interfering with its working part. For applications it should be very attractive to compare results of data analysis performed by different approaches.

Bibliography

[Podraza 2005] Podraza R., Walkiewicz M., A. Dominik A.: Credibility Coefficients in ARES Rough Set Exploration System. Proc. 10th Int. Conf. on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, RSFDGrC 2005, Regina, Canada, Lecture Notes in Artificial Intelligence, LNAI 3642, Part II. Springer-Verlag, Berlin Heidelberg New York (2005) 29-38.

[Pawlak, 1991] Pawlak Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer, (1991).

[Podraza, 2007] Podraza R., Tomaszewski K.: Ordinal credibility coefficient - a new approach in the data credibility analysis. In A. An, J. Stefanowski, S. Ramanna, C. J. Butz, W. Pedrycz, G. Wang (Eds.), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, Proc. 11th Int. Conf., RSFDGrC 2007, Toronto, Canada, Lecture Notes in Artificial Intelligence LNAI 4482, Springer-Verlag Berlin Heidelberg (2007) 190-198.

[Podraza, 2006] Podraza R., Dominik A.: Problem of Data Reliability in Decision Tables. Int. J. of Information Technology and Intelligent Computing (IT&IC), Vol. 1 No. 1, (2006) 103-112.

[Grzymala, 2005] Grzymała-Busse J.W.: Rule Induction. In Data Mining and Knowledge Discovery Handbook. Maimon O., Rokach L. (Eds.), Springer Science + Business Media Inc. (2005) 277-295.

[Dong, 1999] Dong G., Li J.: Efficient Mining of Emerging Patterns: Discovering Trends and Differences. Proc. of the SIGKDD (5th ACM Int. Conf. on Knowledge Discovery and Data Mining), San Diego, USA, (1999) 43-52.

[Schmilovici, 2005] Schmilovici A.: Support Vector Machines. In Data Mining and Knowledge Discovery Handbook. Maimon O., Rokach L. (Eds.), Springer Science + Business Media Inc. (2005) 257-274.

[Podraza, 2006a] Podraza R., Tomaszewski K.: KTDA: Emerging Patterns Based Data Analysis System. Annales UMCS, Informatica, AI, Vol.4, Lublin, Poland, (2006) 279-290.

[Boulesteix, 2003] Boulesteix A.-L., Tutz G.: A Framework to Discover Emerging Patterns for Application in Microarray Data. Institute for Statistics, Ludwig-Maximilian-Universitat, Munich (2003).

[Dong, 1999a] Dong G., Zhang X., Wong L., Li J.: CAEP: Classification by Aggregating Emerging Patterns. Proc. of 2nd Int. Conf. on Discovery Science, Tokyo, Japan, (1999) 30-42.

[Podraza, 2006b] Podraza R., Dominik A.: Credibility coefficients for objects of rough sets. Studia Informatica, No. 1/2 (7) (2006) 93-104.

Authors' Information

Roman Podraza *– Institute of Computer Science, Warsaw University of Technology; Nowowiejska 15/19, 00-665 Warsaw, Poland ; e-mail: R.Podraza@ii.pw.edu.pl*

Mariusz Kalinowski *– Institute of Computer Science, Warsaw University of Technology; Nowowiejska 15/19, 00-665 Warsaw, Poland ; e-mail: M.M.Kalinowski@elka.pw.edu.pl*

THE APPLICATION OF DATA MINING METHODS TO THE CLASSIFICATION OF FINISHED PRODUCTS

Monika Piróg-Mazur, Galina Setlak, Wioletta Szajnar, Tomasz Kożak

Abstract: *In this paper the review of applications of data mining is presented. The algorithm of the induction across decision trees that was proposed for classifying large set of data is described in the part 2 of the paper. A short characteristic of the proposed algorithm is included in the part 3. In the part 4 we present characteristics of R-environment and the implementation of the algorithm. The method of conducting the computational experiments and the obtained results are presented in the part 5. The last part of the paper contains conclusions and further perspectives.*

Keywords*: data mining, classification, critical defects, induction of decision trees formatting rules*

ACM Classification Keywords*: I. Computing Methodologies, I.2 Artificial Intelligence, J. Computer Applications J.6 Computer-aided engineering,*

Introduction

Data mining is the analysis of observational data sets (often huge sets) in order to find unexpected relations and to sum up data in an original way to make it both comprehensible and useful for its owner. Relationships and summaries which are the result of data mining are called models or prototypes. The examples are linear equations, rules, concentrations, graphs, tree structures and re-entry prototypes in time series.

Data mining covers a wide range of topics in the field of computer science and statistics. They have a lot in common but at the same time each of them preserves its distinct character and deals with specific problems and the ways of solving them [Hand, Mannila, Smith, 2005].

The essence of classification

One of the oldest and the most important methods of data mining, which is also of great practical importance, is the classification method. The process of classification in data mining was presented in papers [Kennedy, Lee, Van Roy, Reed, Lipman, 1995], [Mehta, Agrawal, Rissanen, Slig, 1996].

This method was introduced by Breiman and others [Breiman, Friedman, Olsen, Stone, 1984] and it combines classification with application in which it works. The most popular in this group are: CART [Breiman, Friedman, Olsen, Stone, 1984], ID3 [Quinlan , 1986], C4.5 [Quinlan 1993].

Classification means finding data mapping in a set of predefined classes. The model (for example: a decision tree, logical rules) is being built on the basis of database contents. The built model is designed for classifying new objects in the database or for deeper understanding of the existing division of objects into predefined classes.

The main aim of classification is to build a formal model called a classifier. Input data in the process of classification is a set of tuples (a training set of examples, observations, samples), which is a list of values of description attributes and a selected decision attribute. The result of the classification process is a certain model (a classifier) which assigns each tipple (an example) the value of a decision attribute on the basis of values of other attributes [Wazniak, 2008].

A decision tree construction

The task of classifying objects on the basis of putting together their features (attributes) is a key task for technologies connected with data mining. It can be generally said that the main task of classification is to find the function which assigns *n*-set of well-known attributes of a certain object marked as *(x1….xn)* to *y* class which this object belongs to. The tasks of this type appear in many domains. The popular ones include: picture recognition, text analysis, economic trends analysis and customers classification [Kasprowski, 2005].

A method based on decision trees enjoys great popularity. The classifier is represented by a binary tree, there are some enquiries about values of a given feature in its nodes and class assessments in its leaves.

The basic algorithm of a decision tree structure used while constructing it belongs to the method family called "divide-and-conquer". There are many variants of the basic algorithm. ID3 and C4.5. are the algorithms that are the most frequently applied. The main difference between these algorithms is the established criterion of division, which is the way of creating new inner nodes in a decision tree, used while constructing the tree. The method of division should maximize the accuracy of constructing the decision tree that is it should minimize the wrong classification of data records [Kasprowski, 2005].

Let's begin with finding such a division of a training set so as to minimize its entropy that is an unordered set of data, which can be counted by using this formula:

$$H(d) = -\sum_{y} p(y)\log p(y)$$

Where *d* is a set of data and *p(y)* is probability that a sample from this set belongs to class *y*.

The probability is counted as the ratio of the number of samples in a set belonging to class y to the total number of samples.

The tree generated in the stage of learning is used to classify new samples with the unknown class membership.

The database is divided into two sets, where the first one is a training set (it serves to build the model) and the second one is a testing set (it serves to test the model).

Implementation in the R environment

The R package is a tool for data analysis (financial, industrial, biological, medical and other data). It allows carrying out a reliable analysis, to visualize the results by creating legible graphs and to generate the reports automatically. Statistical methods and data analysis and visualization have been implemented in the R environment.

The R package can be used for many applications, both educational and business applications.

R users have the possibility to use CLI (Command Line Interface – writing commands directly into a command line), creating script languages or GUI (Graphical User Interface – ready-made user interfaces).

The method of creating decision trees is available in many different functions in the R package. The functions which are commonly used are: tree(tree), rpart(rpart) oraz cpart (party). All mentioned above functions are used in the same way while constructing trees [Biecek, 2008].

To make classification trees visualized the following functions can be used (depending on the function we used to create the tree):

```
plot.BinaryTree(party)
plot.tree(tree)
text.tree(tree)
draw.tree(maptree)
plot.rpart(rpart)
text.rpart(rpart)
```

While building a classification tree the criterion of a division should be determined, that is which value is to be minimized while creating the next branches (most often it is the error of classification) and the stopping criterion (how long the tree should be divided). Different variants of trees allow to control different criteria.

```
> # okreslamy kryteria budowy drzewa klasyfikayjnego
> ustawienia <- ctree_control(mincriterion = 0.1, testtype = "Testt")
> # uczymy drzewo
> drzewo <- ctree(gr_1~wada_krytyczna, data=dane, subset = zbior.uczacy, controls = ustawier
> # rysujemy drzewo
> plot(drzewo)
> # proces klasyfikacji
> oceny = predict(drzewo, dane[-zbior.uczacy,])
> table(predykcja = oceny, prawdziwe = dane[-zbior.uczacy,3])
prawdziwe
predykcja neg pos
      neg 111 26
       pos 25 34
```

Figure 1. The hypothetical session of constructing a decision tree

Computational data analysis

The characteristic of a data set.

The data set of a real production company, which manufactures ready-made products (glass products – bottles), was used in research. The production company does three-shift work. It is able to produce 200 000 bottles of one type during one shift.

There is some data prepared to classify the bottles. The data is prepared in the form of text files including the value set of the selected parameters (classification of critical defects with the division into groups).

In the first place preparing data was based on bringing it to the flat structure. Next, all names applied in a source system were removed. During this stage of preparing data incoherent and empty data were rejected. The research was carried out by means of the tool – the R package.

Table 1 Hypothetical data from the studied data set

Name tag	fault	group
1	Birdcage	Group 0
2	Spikes	Group 0
3	Internal Fused glass	Group 0
4	Flanged finish	Group 0
5	Overpress finish	Group 0
6	Internal dirt	Group 0
7	Broken stones, Open blisters – internal	Group 1
8	Chipped finish	Group 1
9	Open blister on sealing surface, burrs	Group 1
10	Finish dimensions	Group 2
11	Checks on finish	Group 2
12	Unfilled, bulged finish	Group 2
13	Sharp finish seams	Group 2
14	Chocked finish	Group 2
15	Flanged bottom	Group 2
16	Thermal shock	Group 2
17	Any checks, crizzels, chips	Group 3
18	Microchecks on finish	Group 3
19	Offset finish	Group 3
20	Sharp, thick seams	Group 3
21	Breaking and open blisters	Group 3
22	Dimensional defects	Group 3
23	Sharp crizzels, sharp and open cords	Group 3
24	Bulged body, cold mould	Group 3

Name tag	fault	group
25	Any microchecks	Group 4
26	Crizzels, washboard, threads, cords, ripples, tears	Group 4
27	Wavy, rough, hammer, dirty mould	Group 4
28	Shear mark, dirt outside	Group 4
29	Unffilled thread, thin finish	Group 4
30	Fused bottom, baffle mark, wedged bottom	Group 4
31	Engraving, lettering	Group 4
32	Solid stones	Group 4
33	Closed blisters	Group 5
34	Drag marks, double seams, offset bottom	Group 5
35	Offset mould	Group 5
36	Unstable, bulged, push up, thick bottom	Group 5
37	Faint, no stippling	Group 5

For hypothetical data, included in table 1, sets describing this data are presented in table 2

Table 2. Set

Group 0	Birdcage, Spikes, Internal Fused glass, Flanged finish, Overpress finish, Internal dirt
Group 1	Broken stones, Open blisters – internal, Chipped finish, Open blister on sealing surface, burrs
Group 2	Finish dimensions, Checks on finish, Unfilled, bulged finish, Sharp finish seams, Chocked finish, Flanged bottom, Thermal shock
Group 3	Any checks, crizzels, chips, Microchecks on finish, Offset finish, Sharp, thick seams, Breaking and open blisters, Dimensional defects, Sharp crizzels, sharp and open cords, Bulged body, cold mould
Group 4	Any microchecks, Crizzels, washboard, threads, cords, ripples, tears, Wavy, rough, hammer, dirty mould, Shear mark, dirt outside, Unffilled thread, thin finish, Fused bottom, baffle mark, wedged bottom, Engraving, lettering, Solid stones
Group 5	Closed blisters, Drag marks, double seams, offset bottom, Offset mould, Unstable, bulged, push up, thick bottom, Faint, no stippling

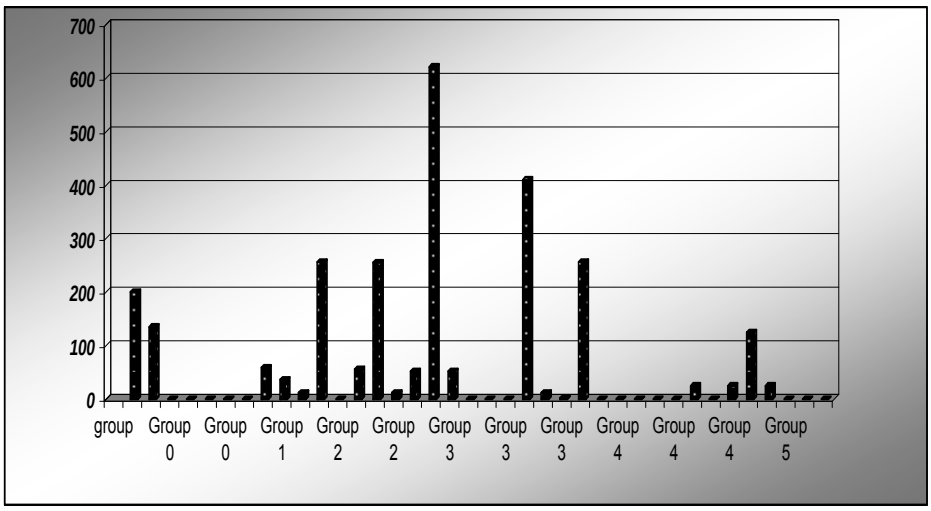


Figure 2. The hypothetical histogram with the division into groups

Decision trees allow data set mining to be done by analyzing values which decide about creating nodes and can be designed for classification of customers or products [Bauer, 2006] [Blundon, 2003] [Shearer, 2004].

The advantage of classification by means of decision trees is the ease of interpretation.

The drawn classifier can be evaluated and corrected by experts in different domains.

Conclusion

Having a training set including samples with well-known classification at disposal makes it possible to create a classification model (called in other words - a classifier).

It is possible to create different classifiers for the same training set taking into consideration:

- which attributes and their combinations are to be considered,
- which classification method will be used.

Thus the same input data (a training set) can have very different solutions. The selection of the optimum solution is usually made on the basis of tests carried out on the set of samples with the known classification [Kasprowski, 2005].

It is necessary to remember that classification algorithms relying on the same data can easily lead to the phenomenon of overloading. The overloaded classifier provides very good results for the data which was used to create it, but it is much weaker for the new data which has not been used yet [Kasprowski, 2005].

The paper presents the preliminary results of the research concerning the classification of ready-made products of the production company. The achieved results prove the usefulness of the applied method of data mining – decision trees. The results will be verified in collaboration with experts from the control department of the studied company. The aim of further research will be to make a series of detailed analyses within the confines of individual ready-made products.

This exemplar is meant to be a model for manuscript format. Please make your manuscript look as much like this exemplar as possible.

In the case of serious deviations from the format, the paper will be returned for reformatting.

Bibliography

[Bauer, 2006] Bauer K.: The power of metrics: Predictive analytics – addressing the business vicissitudes, DM Review Magazine, 2006

[Biecek, 2008] Biecek P Przewodnik po pakiecie R, Wrocław, 2008

[Biecek, 2008] Biecek P Na przełaj przez Data Mining, Wrocław, 2008

[Breiman, Friedman, Olsen, Stone, 1984] Breiman L., Friedman J.H., Olsen R.A., Stone C.J., Classification and Regression Press, Wadsworth, Belmont CA, 1984.

[Blundon, 2003] Blundon W.: Predicting Success: using analytics to drive more profitable customer interactions DM, Direct Newsletter, 2003

[Hand, Mannila, Smith, 2005] Hand D, Mannila H., Smith P Eksploracja danych, WNT, Warszawa, 2005

[Kasprowski, 2005] Kasprowski P, Zastosowanie głosowania klasyfikatorów w przykładowym zadaniu klasyfikacyjnym, Rozdział w monografii Bazy danych: Modele, Technologie, Narzędzia, Kozielski S., Małysiak B., Kasprowski P., Mrozek D., WKŁ, 2005

[Kennedy, Lee, Van Roy, Reed, Lipman, 1995] Kennedy Runy L., Lee Y., Van Roy Benjamin, Reed Ch., Lipman R.P., Solving Data Mining Problems through Pattern recognition, Pearson Education, Upper Saddle River, Nj, 1995.

[Mehta, Agrawal, Rissanen, Slig, 1996] Mehta M., Agrawal R., Rissanen J., Slig: A Fast Scalable Classifier for Data Mining, in: Proc. of the Fifth International Conference on Extending Database technology, Avignon, France 1996

[Shearer , 2004] Shearer C.: Anticipating customer behavior with analytics, CRM Today 2004.

[Quinlan , 1986] Quinlan J.R., Induction of Decision Trees, Machine Learning, Vol.1, No.1, 1986.

[Quinlan 1993] Quinlan J.R., C4.5: Programs for Machine Learning, Morgan Kauffman, 1993.

[Wazniak, 2008] http://wazniak.mimuw.edu.pl/index.php

Authors' Information

Monika Piróg-Mazur *– PhD Student, Editor in chief, Institute of Technical Engineering State School Of Higher Vocational Education, Czarniecki Street 16, 37-500 Jaroslaw, Poland; e-mail: m_pirog@pwszjar.edu.pl*

Galina Setlak, *D.Sc, Ph.D., Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science , W. Pola 2 Rzeszow 35-959, Poland, and The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland, e-mail:* gsetlak@prz.edu.pl

Wioletta Szajnar *– Editor, Institute of Technical Engineering State School Of Higher Vocational Education, Czarniecki Street 16, 37-500 Jaroslaw, Poland; e-mail wiola@pwszjar.edu.pl*

Tomasz Kożak *– Editor, Institute of Technical Engineering State School Of Higher Vocational Education, Czarniecki Street 16, 37-500 Jaroslaw, Poland; e-mail: tomasz.kozak@pwszjar.edu.pl*

A NEURAL NETWORKS APPLICATION IN DECISION SUPPORT SYSTEM

Galina Setlak, Wioletta Szajnar, Leszek Kobyliński

Abstract: *This work presents the using of the chosen neural networks for the data classification. The Statistic Neural Networks has been used for the rating. The neural networks have been successfully applied for building one decision support systems for solving managerial problem.*

Keywords*: artificial intelligence, neural networks, classification, decision support*

ACM Classification Keywords*: I. Computing Methodologies, I.2 Artificial Intelligence*

Introduction

The problems of object recognizing and classification are currently the most important, of which the modern science is interested in at present.

Classification is the most frequently solved problems both in technology and economy. Generally one can define that the classification algorithm lies in finding the data mapping into the set of predefined classes:

$$f_c : R^p \supset X \to C, \tag{1}$$

where $C = \{C_1, C_2, ..., C_n\}$ is the finite set of classes, whereas the set $X \subset R^p$ is the attribute space, and the decision about the classification result is based on them. Classification mapping f_c devides space X into n decisive areas, grouping the attribute formulas belonging to one category [Zieliński, 2000]. Input data is the set of examples, observations, samples which are the value list of the descriptive features constituting the model (classifier). On the basis of the data content the model has been built, which is then used to classify new objects. Building a new classifier is the main aim of the classification of data. The classification process consists of few stages: building a model, testing unknown values . Whereas, the classification means assigning an object on the basis of the chosen distinctive features, to one of the model classes. Classification is mentioned when the classes to which we would like to divide the input set, will be defined before the division process. [Hand, Mannila, Smith, 2005], [Adamczak, 2001].

Statistical methods are traditionally used to solve the classification problems [Witkowska, 2002], [StatSoft, 2005]. Moreover there are also used the following methods: Bayes classifiers, decision trees, neural networks, genetic algorithms, rough sets, fuzzy logic and neuro-fuzzy classifiers [Stąpor, 2005], [Rutkowska, Piliński, Rutkowski, 1997].

The aim of this thesis is investigating the possibilities of applying the artificial neural networks in the object classification and evaluation and comparative analysis of these results. As an analysis result there is the choice of the net done, the net with the best classification results with solving the sample problem of the market research analysis.

Methodology and results of the experimental research

There were the following neural networks used in these object classification research, they are different in structure and teaching method:

- Multilayer Perceptron (MLP) ,
- Radial Basis Function (RBF).

These experimental researches have been done on the practical example of the market research analysis of household appliances. The task of this classification is choosing the most prospective market for the produced goods in a chosen factory. We will use the information about hoovers for description of the classified observations, which should be divided into four classes, that is markets. Initial parameters characterizing the particular goods (hoovers) are presented in table 1.

The first column of table 1 includes the verbal description of the feature, a type is presented in the second column. The third column includes the shortened name responding to the parameter, under which it is in the programme package - Statistic Neural Networks (STNN), which will be used to solve the made tasks [Setlak, 2004].

Table1. Characteristic parameters of the products (hoovers)

Description	Type	Shortened name
Engine Power/capacity	number	ENGINE_W
price	number	PRICE
Presence of the air filtration system	{Yes, No}	FILTR_SYS
Presence of automation	{Yes,No}	AUTOFUNC
Automatic cord rolling	{Yes, No}	AUTOCORD
Power regulation and speed switch	{Yes, No}	SPD_CTRL
Noise reducing system	{Yes, No}	NOISSYS
Device moistness function	{Yes, No}	WASH
Aesthetic of the external appearance	{Yes, No}	VIEW
Additional possibilities and improvements (merits)	{Yes, No}	FEATURE
Well known producer	{Yes, No}	BRAND
Service level	{low, average, high}	SERVICE

The shortening CLASS corresponds to initial classification parameters – one of the four classes of the presented market in the STNN package [9]. The market description with the division into classes is presented in the following way:

- Class 1 – market where the highest quality goods are of the most interest, of huge power with additional features and a very good service,
- Class 2 is the average class of purchasers, who are less interested in expensive products, but they still demand good products,
- Class 3 – purchasers paying attention to low price, but not interested in improvements,
- Class 4 – market where the low price is the priority, but other parameters are less important.

All data for this classification were taken from the market analysis of one of the production factories. These data were prepared in advance. This processing phase in the STNN program is called „pre-processing". In this problem the learning set describes 116 models of Hoovers with different parameters.

In order to search for the best architecture and optima teaching method the Intelligent Problem Solver (IPS) has been used in this thesis in the advanced version, included in the STNN package [Statistic Neural Networks,

1999]. IPS is exceptionally useful device that is helpful for the user in the most difficult, laborious and intensive phase of neuron network construction –testing and choosing different models.

IPS formulates and initially evaluates the architecture of neural networks with different initial variable sets. It allows to asses which parameter is the most important one. In classification problems, one can also control the way of classification for particular values. With the IPS use in the analysis of the initial parameters importance, there have been the optima neuron net structures described and only the most important parameters have been used in the research then. In complex problems with the numerous amounts of initial parameters in STNN, one can use the following device to choose the net structure – The Genetic algorithm of the initial variables selection.

IPS for the classification task, creates and tests many nets which are compared and then, with the chosen option from many examples, chooses and keeps some that are different due to the quality, construction and number of hidden layers and conducts the additional tests.

After the research done the following results were presented in table 2 on this initial data set (object „Hoover") and solving the classification problem with the use of different net types MLP and RBF

Table 2. Classification results obtained after using the IPS for the object „hoover"

Net type	Error	Number of the initial parameters	Number of the hidden layers	The most important parameter	Working time in net [s]	Chosen market
MLP	0,099	10	12	10	2	1
MLP	0,111	9	12	1	4	1
MLP	0,175	9	11	1	6	1
MLP	0,126	8	9	10	8	2
MLP	0,088	10	13	1	10	1
MLP	0,102	9	14	9	15	2
MLP	0,185	11	10	1	20	1
MLP	0,211	8	9	1	25	2
MLP	0,266	9	8	2	30	4
MLP	0,299	12	6	1	35	2
RBF	0,098	9	7	3	2	1
RBF	0,095	9	8	1	4	1
RBF	0,127	10	6	3	6	3
RBF	0,131	11	5	3	10	1
RBF	0,175	9	2	3	15	2
RBF	0,155	9	3	3	20	3
RBF	0,191	12	3	1	25	2
RBF	0,130	10	7	1	50	2
RBF	0,093	9	8	3	55	2
RBF	0,116	10	6	3	60	2

Source: own study

Some of these relations achieved during the analysis are presented in the following graphs. Some of the points in the graphs are the average values responding to the error value for the same number of initial parameters or hidden layers.

Conclusions

The research done resulted in working out two types of neuron net s of different architecture, taught with different methods that were used for classification the object „hoover".

The analysis of net work is directed to the division of numerous hoovers model s into four markets. The precision of assignment to the particular classes have been evaluated.

During the classification 116 types of hoovers, more difficult was RBF to teach than MLP, which influenced the final result. The proper variable choice played a very important role in the classification process, and it influenced the classification result. The best result (marked in table 3) was received with the use of MLP net with ten inputs and thirteen hidden layers. The optimal working time of the networks was then 10 seconds.

The most rejected parameter was 5 and 8. These were the least important parameters then. For these parameters the correlation ratio had lower than 1 value. The most chosen parameters for this analysis were 1 and 10. They were chosen for each analysis. They were also the most important parameters of the best value quotient ratio. For RBF networks, the best result was for 9 inputs and 8 hidden layers. Parameter 1 was also the most important and the most chosen parameter too.

Analyzing all results for the particular networks it appears that MLP networks gave worse results in short time of networks working and the working time of the networks was longer of over 15 seconds. The best result was for the working time of 10 seconds. The result was dependent on the initial parameters amount and the number of hidden layers. The best results were for 9-11 parameters at the initial point. The more hidden layers, the best the result was.

RBF networks were behaving similarly, if we take initial parameters and hidden layers into consideration. The more hidden layers, the best the result was.

Depending on which parameter was the most important, the networks were doing the classification for the particular classes (markets). In case of the results with low error, the choice of the market was the accurate one in case of the both networks. If the most important parameter was a huge power and additional improvements, the most chosen class would have been the class 1 and 2. If the important role in the classification was played by the parameter 2, that is the choice price, the market 3 or 4 was chosen.

Comparing two best results for the particular networks, one can state that in such a case, the better classifier was MLP networks.

Summing up, we can state that neural networks, due to their possibilities can be successfully used in solving the classification problems. It is easy to change their architecture, which results in bigger versatility. Looking at the results, one can state that the neural networks are very good classifiers, even if they are not 100% compliant with the model.

The reason that some of the data were incorrectly classified can be the fact that the classes overlap. To gain better results, one should search for additional solutions, among others test networks of different architecture, and as for the teaching, to use bigger teaching set. It can have better results that the one achieved in this study.

Bibliography

[Hand, Mannila, Smith, 2005] D. Hand, H. Mannila, P. Smith. Eksploracja danych, WNT, Warszawa, 2005.

[StatSoft, 2005] StatSoft. The statistic homepage is a public service provided by StatSoft, Inc. 1984 – 2005.

[Stąpor, 2005] K. Stąpor. Automatyczna klasyfikacja obiektów, Wyd. Exit, Warszawa, 2005.

[Adamczak, 2001] R. Adamczak. Zastosowanie sieci neuronowych do klasyfikacji danych doświadczalnych, Praca doktorska, Uniwersytet M. Kopernika, Toruń, 2001.

[Witkowska, 2002] D. Witkowska. Sztuczne sieci neuronowe i metody statystyczne, C.H.BECK, Warszawa, 2002.

[Rutkowska, Piliński, Rutkowski, 1997] Rutkowski L., Rutkowska D., Piliński M.: Sieci neuronowe, algorytmy genetyczne i systemy rozmyte, PWN, Warszawa – Łódź 1997.

[Statistica Neural Networks, 1999] Statistica Neural Networks. Addendum for Version 4.0, Statsoft, Inc., 1999.

[Zieliński, 2000] J. Zieliński. Inteligentne systemy w zarządzaniu – teoria i praktyka, PWN, Warszawa, 2000.

[Setlak, 2004] G. Setlak. Inteligentne systemy wspomagania decyzji, Wyd. LOGOS, Kijów (w jęz. rosyjskim), 2004.

Authors' Information

Galina Setlak - *D.Sc, Ph.D., Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science , W. Pola 2 Rzeszow 35-959, Poland, and The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland, e-mail: gsetlak@prz.edu.pl*

Wioletta Szajnar *- M.A. The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland, e-mail: wiola@pwszjar.edu.pl*

Leszek Kobyliński *- M.A. The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland, e-mail: leszek@pwszjar.edu.pl*

Knowledge Representation and Management

ORGANISATION OF DISABLED PERSONS' TELEWORK AS PART OF PROCESS OF KNOWLEDGE MANAGEMENT

Tatjana Bilevičienė, Eglė Bilevičiūtė

***Abstract**: Challenges of knowledge society, knowledge economics change models of management. Work trends, organisational structures are changing. Most of knowledge management processes one way or another are connected with main elements of information management: information technologies and information systems. Knowledge management is new branch of management, the main purpose of that is bigger effectiveness of business applying synergy of humans, processes and technologies. Knowledge management helps for employees single-mindedly to create, share, collect, keep and practice the knowledge. Fundamentals of knowledge management could be used for organisation of telework of disabled persons, solving economical problems of state. Knowledge management could help to increase productiveness of employees, expanding sources of reachable for them knowledge. Improving shapes of disabled persons' education the very important is to pay attention at how educational knowledge is implemented practically. Knowledge management should encourage employees to integrate such educational processes together with their work methods and to apply results of educational processes in their daily activities. For purposeful and beneficial education there should be forecasted what knowledge is the most important. Authors of article analyses the application of knowledge management methods in structure of professional rehabilitation of disabled persons, organising telework of disabled persons. During such process new knowledge should be used not only by disabled persons, but both represents of governmental institutions and consultants.*

***Keywords:** knowledge society, disabled persons, telework.*

***ACM Classification Keywords:** K.4.3. Computers and society - Computer-supported collaborative work*

Introduction

Transformation of modern society to knowledge society originates the absolutely new global social and economical contexts that require different management principles, skills, abilities and competences. Experts, who prognosticate changes of European economy, declare that the main factor of development of European economical space business organizations and economy would be the knowledge, generation of innovative products, perfection of production and management's methods.

Most of knowledge management processes one way or another are connected with the main elements of information management: information technologies and information systems. Information technologies are the main connection of information management and knowledge management too, and it is the main accent. Inquiring objects – information and knowledge, its expression's levels, determine the connections of information and knowledge management.

Information and communication technologies assume huge importance in all spheres of modern society life – in professional life, in education, in daily life, but not for all members of society it is available. Application of information and communication technologies could challenge new forms of social disjuncture. New technologies not only require special skills and competences, but both it becomes dangers of social disjuncture for special groups of employees, for example, for disabled.

Telework and new form of work organization became important aspects of creation of new work's places in European Union. 30 percents of all staff in Europe regularly practise telework. Together with additional telework it would be 36 percents. Big enterprises mostly practise telework. There are more than 3 millions teleworking employees at home in European Union. The alternative telework composes quite big part; it means that such work is performing both at home and at work. 1,4 million of employees we can classify as private teleworkers owning home office. Also 2,3 million of employees are mobile teleworkers [Bergum, 2007].

Problem of employment for disabled persons is partly connected with permanent changes and technical achievements on labour market. One of the tasks of Lisbon strategy is creation of information society, helping to implement knowledge economy and create new work places in states that have the biggest development potential. Spread of new information technologies changes character of work – its quantity, quality, particularity of performing work. Telework at home could supply better opportunity for disabled persons to receive work places.

Practical value of knowledge management

Environment of enterprise is changing continually because of globalisation. Expansion of competition, intensification of international economical cooperation, spread of technologies, changing of social priorities, attitudes and values – all are such transformations that cause new phenomenon – knowledge based economy. The main attention in knowledge economy is played to person, his abilities, knowledge and opportunities of its consumption. Information technologies precede revolution in business world: expand computer networks, globalise market, create business environment without borders. Internet and communications become equipment of successful business and bottomless source of information that properly using we can reach competitive advantage in any activity. Types of social economic activity are transferring to electronic environment, e-commerce replaces traditional commerce.

Challenges of knowledge society, knowledge economy are changing models of management, enabling in scientific view basically to analyse the advantages and disadvantages of state or organisation, to establish fields of strategic excellence creation. Elements of information and knowledge management are in every modern filed of management: in processes management, in time and space management, in changes' management, in crisis and conflicts management, in organisations' management, in education management, in quality management and etc. Information and knowledge management in system of information and communication sciences composes nucleus of going together manage mental disciplines of information sciences that perfectly fits the modern conception of management.

Knowledge management's essence becomes the management of individuals with particular skills and experience, with purpose to encourage particular behavioural models in organisation and interaction of individual employees – socialisation. The main attention is paid to knowledge based activities and processes that educate abilities of enterprise to work effectively. Effective knowledge management in organisation directly depends on technologies, methods and interaction of persons [Bhatt, 2001].

Practical space of knowledge management realisation is making knowledge economy. Knowledge management becomes strategic discipline, mostly influence development steps of knowledge economy and information society. If prerequisites of knowledge economy would be evaluated successfully, we could scientifically basically to analyse advantages and disadvantages of state, to establish fields of creation of critical strategic excellence and to analyse good practice of successful work.

Modern management as society generally is one-to-many and heterogeneous. Methodological armamentarium of management and variety of methods are changing and increasing. By information attitude, management connects

most scientific elements to holistic space. Management generally is concurrent part of information phenomenon – management always is firstly informational, not formal solution. In information space very different ways and methods of social and economic life are developing and existing [Augustinaitis, 2005].

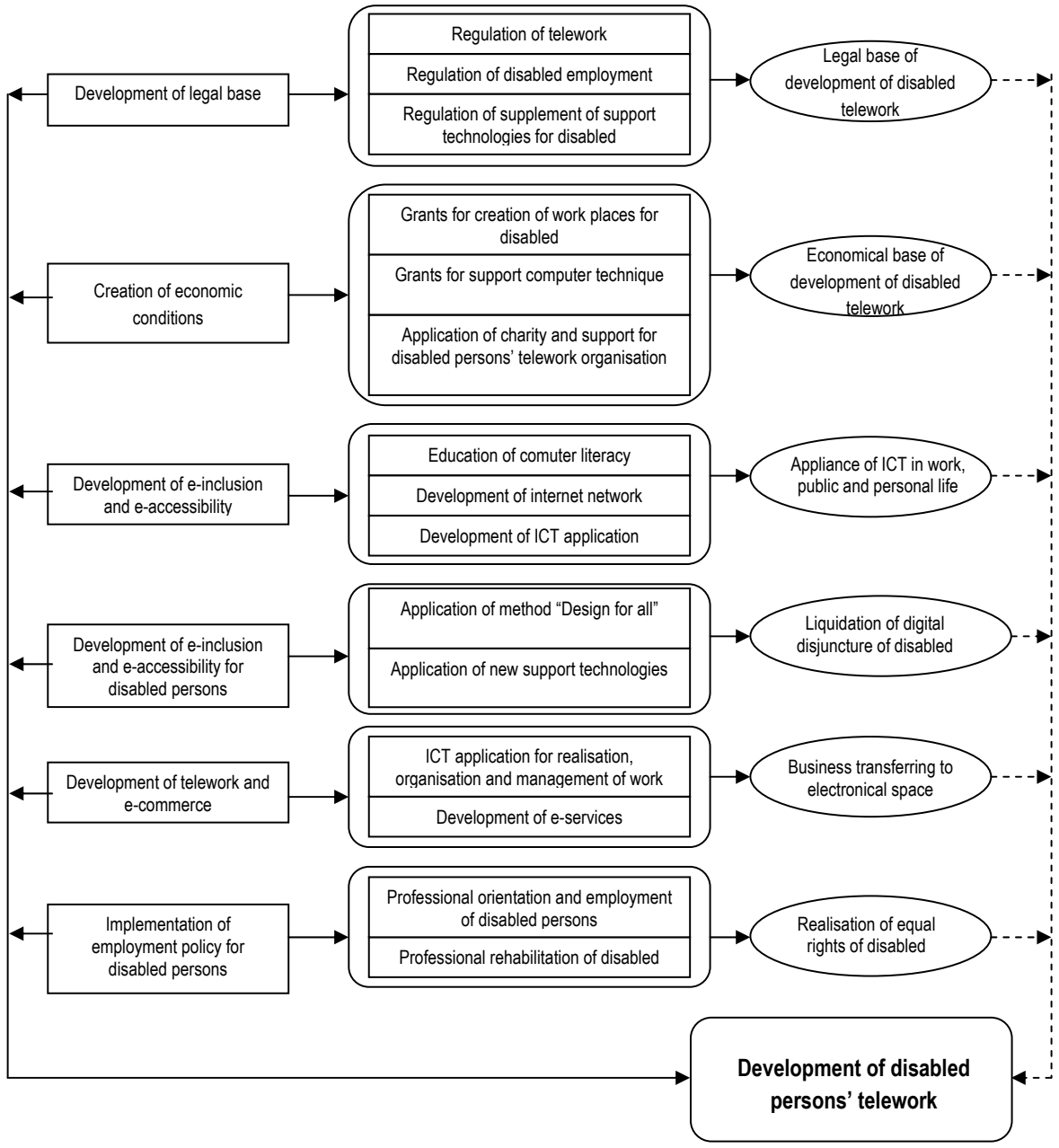


Figure 1. Scheme of development conditions of disabled persons' telework

Knowledge management previously was as attachment for information and technology management, developed in independent branch, essential purpose of which is management of organization sources as intellectual capital, employees'' knowledge, organisation's image and others. Development of new information technologies, that relieved types of most global business operations and education, transferred business to new level, supplying the increasing importance for knowledge. Conjunction of different knowledge parts to management of strategic intellectual capital brings people to new practice of management in information age.

Knowledge management is optimal application of theoretical and practical knowledge in business processes – with purpose to reach durable advantage against rivals and bigger benefit of all shareholders of enterprise – investors, employees, managers, so common state benefit would be implemented [McGinn, 2001].

Model of professional rehabilitation of disabled persons in knowledge management context

Development of employment's levels is effective measure causing development of economy and stimulating economy of social inclusion, both securing protection for disabled persons to work. Disabled persons as others members of society have demand to perform significant activity, beneficial for society and themselves.

Together with rapidly developing information technology the economical and social environmental is changing, the new opportunities are starting, so character of work is changing, new work forms are starting. Today often to remove work at employees' place is easier than opposite. Although, properly organised, flexible, mobile and independent of place work could essentially reduce costs, raise effectiveness, it is profitable for employers, customers and employees. Telework commonly is understandable as opportunity to accomplish work or its part at home maintaining connection with work place using technical equipment (internet, fax, and phone). It lets to reduce level of unemployment, to keep specialist with high qualification. Telework and e-commerce illustrate transferring of labour market to information society [Benchmarking, 2000].

According authors, the theoretical united system of development of disabled persons' telework could be presented as sustainable operative structure with related inter-connections (see Figure 1). Creation and sustainable development of such system depend on legal basis development, economical conditions creation, development of e-inclusion and e-accessibility for disabled, development of telework and e-commerce, implementation of policy of disabled persons' employment. Analysing presented scheme, we could predicate that successful organisation of disabled persons' telework is closely connected with the main development processes of information society and highly depends on disabled persons' e-inclusion and e-accessibility.

Seeking to increase disabled person's efficiency, professional competence and potential to participate in labour market professional rehabilitation services are providing for disabled persons. Professional rehabilitation services are the part of professional integration system in European Union, closely cooperating with systems of social services, labour market, medical and educational systems. Professional rehabilitation is coherent process, consisting of several stages that differ in every country but essentially professional rehabilitation composes such stages: establishment of person's working ability; evaluation of work experience and arrangement to work; professional education, training, and cultivation; training of working skills, organization of working test; employment or support in creation of new enterprise or in practise the autonomous activity. Structure of professional rehabilitation processes is presented on Figure 2. Institutions of professional rehabilitation services could play the main role in process of telework development.

Modern organization should comprehend knowledge management and implement it inside. Knowledge management is manage mental instrument supporting by different measures to create working environment in that seeking the best result they optimally create, spread and use their and others knowledge. The main result of knowledge management is environment stimulating employees to create, spread, keep and apply knowledge and consisting of all processes, roles, measures and structures that let to implement it.

Principles of knowledge management rarely are applying in development of disabled persons' professional rehabilitation. Analysing system of disabled persons' professional rehabilitation in context of knowledge management we will use triad that describes the knowledge management as process and as result: knowledge creation, knowledge spread, knowledge appliance.

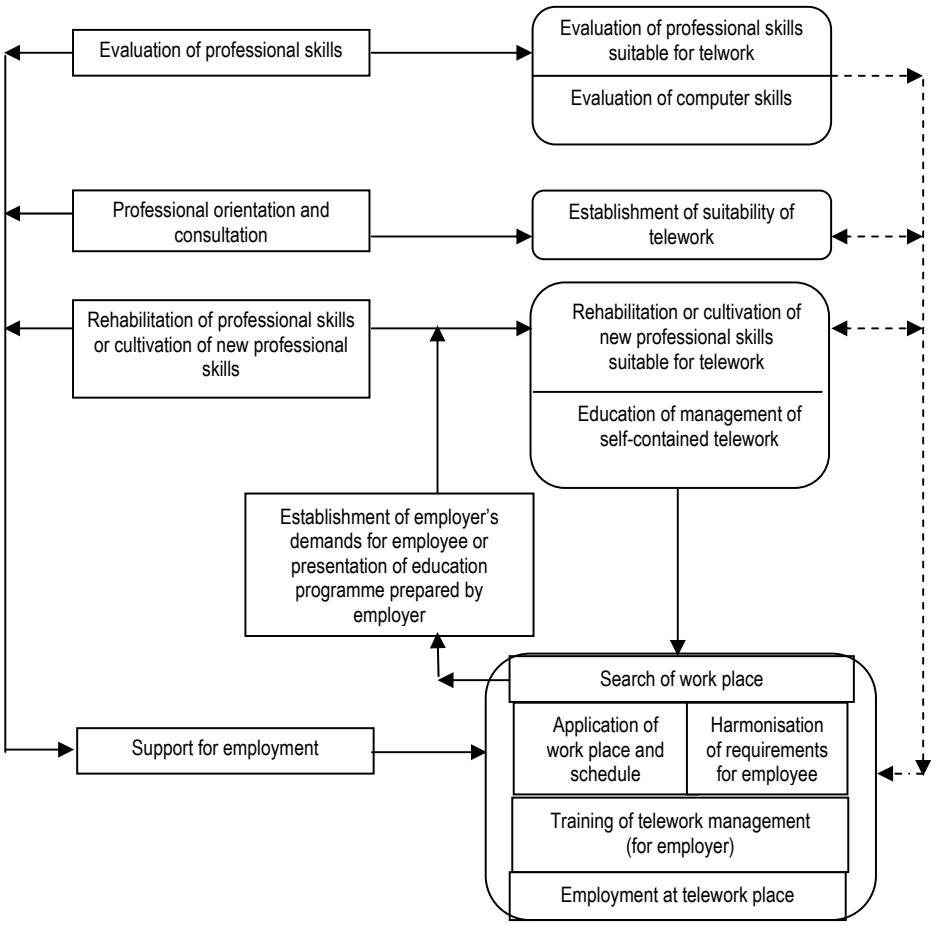


Figure 2. Structure of professional rehabilitation services

Creation of knowledge. In our opinion, model of disabled persons' professional rehabilitation applicable in Lithuania, hasn't been discussed in context of its suitability for disabled persons' telework. In February'2009 Tatjana Bilevičienė accomplished interview of experts. Only few specialists in Lithuania could qualitative evaluate the condition of disabled persons' social and professional rehabilitation, quality of disabled persons' employment. So, as experts' representatives of disabled persons' organizations, persons, responsible for disabled persons' professional rehabilitation, social work specialists – scientists, businessmen, were questioned. One of the most important questions supplied for experts was what, in their opinion, changes are needful and possible on every stage of system of disabled persons' professional rehabilitation and employment, on purpose to successfully employ united disabled persons' telework system. Experts solidly determined that new management programmes of disabled persons' professional skills evaluation, establishment of suitability of telework, rehabilitation or cultivation of professional skills and telework are necessary. There is predicated that for successful organization of telework part of employees of professional rehabilitation system should acquire additional qualification.

Spread of knowledge. Research of theoretical material and experts' opinion showed that in spite of quite intensive system of professional rehabilitation in Lithuania, organization of disabled persons' telework requires forecasting new methods of professional rehabilitation: professional education should be combining with training of computer literacy. Specialists who working with disabled persons in such way should have not only skills of social work and profession but both good skills in ICT. ICT knowledge is necessary even for medics who evaluating suitability of choosing work for disabled persons, because noxiousness of computer works and necessity of appliance of support computer technique should be evaluated. Doctor should evaluate affections of

disabled persons' health and organism and discuss with ICT specialist the opportunities to compensate such disadvantaged by ICT (using support technologies). Also, ICT specialist should describe possible effect of computer work for disabled person, and doctor should decide if particular person could work particular works (and how long he could work such works). ICT specialist composes list of necessary compensative technique and software for employment of disabled persons.

Medical adviser should have comprehensive description of employee's responsibilities; that such information would be useful the character of work, person, working time, preparation for flexible work, physical requirements of work (stay, intellectual and emotional requirements), intensity factors, forecasted working results, conditions of work contract should be described.

Discussing advantages of telework, both we should describe problems of such work. Employers fear to lose control of work, demands of teleworkers' socialization is hardly satisfied, informal meetings proceed without teleworkers, some of teleworkers overwork, teleworkers could wrongly evaluate the main task of work, leaders have problems in evaluation of work only by results, electronic connection not always successfully displace direct communication, it is complicate to plan meetings, harder to accomplish urgent work, professional unions and others organizations could lose nearness with its members [Coope, Kurland, 2002].

One of the most important problems of telework is management problem, because additional requirements for leaders originate. Leaders should be able to manage in distance, evaluate work by results, not by time, what employee is at his working place. Also leader should be able properly formulate work task. Employee should be able properly plan his work, his activity, properly understand tasks, should be able actively, successfully to find solution in every day small problems. Both, he should be able properly present his activity. Telework could be accomplished only then task is clearly described. Successful telework requires particular management style and good selection of employees [Nilles, 2007].

Employment of knowledge. Improving education forms it is very important to highlight how many of training knowledge is used in practice. Knowledge management is part of common management of organization. The main purpose of it is to increase effectiveness of organization, purposely to improve only essential educational processes: knowledge management should stimulate employees to integrate these educational processes with their working methods and to apply educational processes' results in their every day activity. There should be forecasting what knowledge seeking strategic tasks are the most important for organisation [McGinn, 2001].

System of disabled persons' professional rehabilitation should forecast the adaptation of work places for disabled persons and their education preparing specialist for particular work place.

Conclusion

Knowledge management is connected with innovations, inter-connections, ideas, competences, structures. This management supports individual or groups' education, stimulates and enhances spread of experience, distribution of failures and good practises, choice of optimal solutions. Knowledge management technologies could be used for stimulation of dialogues, bargains, communication, but it is not essence of such management.

Accomplished researches show that modern professional rehabilitation system could be conforming for organization of disabled persons' telework only applying principles of knowledge management.

If disabled person has professional, psychological, functional suitability, way of telework could be recommended for him. If disabled person agrees, suitable professional rehabilitation programme would be created for him. In such way, the search of future work place should be oriented particularly at telework method.

Bibliography

[Augustinaitis, 2005] A.Augustinaitis. New Criteria of Rationality in Knowledge Society. In: Information Sciences. 2005 Vol. 35, p. 32-46. Vilnius University Publishing House, 2005.

[Benchmarking, 2000] Benchmarking progress on new ways of working and new forms of business across Europe. ECaTT Final Report IST Programme. KAII : New Methods of Work and Electronic Commerce, 2000. http://www.ecatt.com/freport/ECaTT-Final-Report.pdf [visited 2009 04 05].

[Bergum, 2007] S.Bergum. What has happened to telework? Failure, Diffusion or Modification? In: The Journal of E-working, Vol 1, February 2007, pages 13-44. - Norway: Lillehammer.

[Bhatt, 2002] D.G.Bhatt. Management strategies for individual knowledge and organizational knowledge. In: Journal of Knowledge Management, vol. 6, no. 2, p. 31–39.

[Cooper, Kurland, 2002] C.D.Cooper, N.B.Kurland. Telecommuting, professional isolation, and employee development in public and private organizations. In: Journal of Organizational Behaviour, Volume 23, Issue 4, June 2002, p. 511-532.

[Holbrook, Wolfe, 2002] J.A.Holbrook, D.A.Wolfe. Social Capital and Cluster Development in Learning Regions. In Knowledge, Clusters and Regional Innovation: Economic Development in Canada, 2002, eds, Kingston: Queen's School of Policy Studies and McGill-Queen's University Press, p. 11-38.

[McGinn, 2001] N.F.McGinn. Knowledge Management in the Corporate Sector: Implications for Education. 2001. http://www.norrag.org/db_read_article.php?id=734 [visited 2009 04 05].

[Niles, 2007] J.M.Nilles. The Future of e-Work. In: The Journal of E-working, 2007, Vol 1, January 2007, p. 1 – 12. Jala International, Inc, USA.

Authors' Information

Eglė Bilevičiūtė *– Mykolas Romeris University, Faculty of Law, Department of Administrative Law and Procedure, PhD, Assistant Professor, Ateities 20, LT-08303 Vilnius, Lithuania, e-mail: eglek@mruni.lt*

Tatjana Bilevičienė *– Mykolas Romeris University, Faculty of Economics and Finance Management, Department of Business Economics lecturer, Ateities 20, LT-2057 Vilnius, Lithuania, e-mail: tbilev@mruni.lt*

ALGEBRA LOGIC APPROACH TO PERSON'S THINKING MECHANISMS FORMALIZATION

Olga Kalinichenko

Abstract*: It is known, that person's thinking is inaccessible to studying by direct physical and psychological methods. In this case it is necessary to have indirect ones. Computers do not understand the psychological description and formalization of thinking mechanisms. Algebra logic analysis of natural language and person's thinking plays an important role for development of logic mathematics and its applications in artificial intelligence. Only axiomatic method works in this situation. On the basis of axioms' system we can propose an approach that helps to investigate the structure and properties of objects. The main problem of formal studying of a natural language is shortage of the mathematical apparatus. The axiomatic description of logic mathematics' objects requires preliminary realization of constructive logical tools, which subsequently become a subject of the axiomatic analysis. The paper is devoted the algebra of ideas to axiomatic construction. The carrier of this algebra is naturally interpreted as the set of intelligence ideas (thoughts, concepts and, in general, any subjective conditions of the person). There are devised some methods for application of proposed formal apparatus. Simultaneously with algebra of ideas formal introducing there is considered its intentional interpretation.*

Keywords*: a predicate, algebra of ideas, artificial intelligence, Cartesian set, algebra of predicates, algebra of single k-dimentional first order predicates.*

ACM Classification Keywords*: I.2 Artificial intelligence - Natural Language Processing*

Introduction

One of artificial intelligence lacks, which is much limiting sphere of its practical use, is the inability of the machine to understand human speech and, as a consequence, impossibility of semantic processing of the natural language texts. For studying of human thinking, in particular, of natural language semantics mechanisms are successfully used comparative method and logic apparatus of predicates and predicate operations. [Bondarenko, 2000] For development of logic mathematics and its appendices in artificial intelligence the special role plays algebra logic analysis of natural language. In this case we can apply axiomatic method. On the basis of axioms' system we can propose an approach that helps to investigate the structure and properties of objects. The axiomatic description of logic mathematics' objects requires preliminary realization of constructive logical tools, which subsequently become a subject of the axiomatic analysis. Yet there are no many abstract concepts for description of natural intellectual processes. [Shabanov-Kushnarenko, 2005]

In logic mathematics the central role is played by relations. The relation are formally described with the help of predicates. [Ivanilov,2007] In language of algebra of predicates it is possible to describe any information process but the algebra of predicates is constructed structurally. It is enough, if the internal structure of information process is known for us, but it is not enough, if we have only results of this process, as in natural language. What occurs inside, what algorithms are working – it is not known. That is why it is necessary to set algebra of predicates axiomatically. Then these properties we can observe in real human speech and behavior and to make conclusions about structure of these information processes.

The algebra of predicates has appeared as a result of attempts of the formal description of natural language - modeling of declinations, conjugations, words' formation. Now, there is more complex task - formalization of understanding and semantics of language. In this case it is necessary to describe concept of a predicate axiomatically because relations, but not functions, lays in the basis of thinking. The task about the formal

description of a predicate is naturally divides into two tasks. The first is a consideration of a single predicate from variable x, and the second is expansion of this integrated variable x in a set of variables, i.e. studying of a many-place predicate or structure of the Cartesian product.

The paper is devoted to construction of the methods of formal description of natural language structure with the help of algebra of ideas - mathematical apparatus, which is constructed axiomatically as algebra logic analogue of natural language. The urgency of this area is defined by perspectives of applying of the received methods for developing systems of dialogue with the computer in natural language. In this work the properties of single predicates are considered, as it is enough for modeling rather wide area of natural language. Except studying of single predicates, there are some adjacent questions, such as predicate of equality and models, i.e. circle of tasks closely connected with axiomatic of a single predicate.

A model of ideas' equality. Formal representation of ideas

We shall use algebra of single k-dimentional predicates of the first order in a role of the algebra of ideas prototype. It appears, that exactly the algebra of single k-dimentional predicates of the first order brings to the most general algebra of ideas definition that is necessary to us. Abstract analogues of the more general algebras of final predicates (many-placed and the any order) turn out simply by detailed elaboration of initial algebra of ideas.

single k-dimentional predicates of the first order are entered as follows. Let $A_k = \{a_1, a_2, ..., a_k\}$ is the set, that consists of k letters $a_1, a_2, ..., a_k$. All letters are numbered, everyone has the serial number. The variable x is set on A_k and it named alphabetic. We enter the set $\Sigma = \{0, 1\}$ that consists of logic constants 0 and 1, named accordingly zero and unity. The variable x is set on Σ and it named logic. Each function $y = P(x)$ that display set A_k in set Σ we named as single k-dimentional predicate of the first order. Let's speak, that predicate P is set on set A_k. Set of all single k- dimentional predicates of the first order we designate by a symbol M_k. Let $N_0(k)$ is a number of all predicates included in set M_k. It is equal $N_0(k) = 2^k$.

algebra of ideas Construction we shall begin with introducing of its carrier - set of all ideas. We shall designate by a symbol S_k the set consisting of 2^k various elements $s_0, s_1 ..., s_{2^k-1}$. we Accept the set S_k in a role of the algebra of ideas carrier with dimension k. Elements of set S_k we name ideas of dimension k. Single k-dimentional predicates of the first order serve for us as prototypes of elements of set S_k. The number of elements 2^k of set S_k is chosen so that it coincided with number of all single *k*-dimentional predicates of the first order. Set S_k we shall name k-dimensional space of ideas. The question on concrete value of number k is left open. While we shall consider, that in a role k any natural number *k*=1, 2, … can be chosen. Let's notice, that at any value k the set S_k is not empty. In some tasks we need not all the set S_k but only some part N of it. The number of elements in set N can be any, but it should be less, than 2^k. Set N we shall name incomplete set of ideas, and set S_k - full.

Let's enter bijection $\Phi: S_k \to M_k$, establishing univocity between all ideas of dimension k and all k-dimentional predicates that set on set A_k. It always can be made, because sets S_k and M_k contain identical number of elements. Predicate $P = \Phi(x)$ we shall name a predicate corresponding to idea x, and idea $x = \Phi^{-1}(P)$ - the idea corresponding to predicate P. There are two examples of bijection Φ'and Φ" in tables 1 and 2. Bijection Φ': $S'_k \to M_k$ is determined on three-dimensional space of ideas $S'_3 = \{s'_0, s'_1, ..., s'_7\}$, bijection Φ'': $S''_k \to M_k$ is determined on space of ideas $S''_3 = \{s''_0, s''_1, ..., s''_7\}$ with the same

dimentional. The symbol x ' designates variable that sets on set S'_3 and symbol x " - a variable that sets on set S''_3. Sets S'_3 also S''_3 can be considered as different systems of designations for the same three-dimensional ideas.

Table 1 Variables that set on set S'_3

x'	s'_0	s'_1	s'_2	s'_3	s'_4	s'_5	s'_6	s'_7
$\Phi'(x')$	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7

Table 2 Variables that set on set S''_3.

x"	s''_0	s''_1	s''_2	s''_3	s''_4	s''_5	s''_6	s''_7
$\Phi''(x'')$	P_4	P_5	P_7	P_1	P_6	P_0	P_3	P_2

Elements of set S_k we shall psychologically interpret as ideas of the examinee. Predicate P (the question is about single k-dimentional predicates of the first order) that accept for all letters $x \in A_k$ zero value $P(x) = 0$, we shall name identically false. Predicate P that accept for all letters $x \in A_k$ individual value $P(x) = 1$, we shall name identically true. We designate these predicates accordingly symbols 0 and 1. The predicate 0 has number 0, a predicate 1 - number $2^k - 1$.

The idea that corresponds to identically false predicate 0 we shall lie, and designate it by the same symbol 0. The idea that corresponds to identically true predicate 1, we shall name true and designate it by symbol 1. Thus, $\Phi^{-1}(0) = 0,\ \Phi^{-1}(1) = 1.$ Operation of the bijection Φ reference is designate by symbol $^{-1}$. In a role of function Φ^{-1} arguments they mean predicates, that are the elements of set M_k, and in a role of function Φ^{-1} values they mean ideas, that are the elements of set S_k. This circumstance, however, will not result in misunderstanding because the true sense of signs 0 and 1 is easily determined on a context. For example, we shall find ideas 0 and 1 in sets S'_3 and S''_3 with help of tables 1 and 2. In a role of the predicate 0 acts predicate P_0, in a role of the predicate 1 - predicate P_7 in both tables. We find $\Phi'^{-1}(P_0) = s'_0$, $\Phi'^{-1}(P_7) = s'_7.$ from table 1. Thus, for set S'_3 we have $0 = s'_0$, $1 = s'_7.$ We find $\Phi''^{-1}(P_0) = s''_5$, $\Phi'^{-1}(P_7) = s'_2$ from table 2. Thus, for set S''_3 we have $0 = s''_5$, $1 = s''_2$.

Statement that express lie, we shall name the contradiction. The statement that express true, we shall name a tautology.

A predicate of ideas' equality

Let's consider a predicate of equality $\mathsf{D}_k(P,\ Q)$ of predicates P and Q, which are set on the Cartesian square of set M_k of all single k-dimentional first order predicates.

It defines by equality:

$$\mathsf{D}_k(P,\ Q) = \forall x(P(x) \sim Q(x)), \tag{1}$$

that fair for anyone $P,\ Q \in M_k$. The predicate D_k puts in conformity to equal predicates P and Q a logic constant 1, unequal - 0. The equation $\mathsf{D}_k(P,\ Q) = 1$ sets the relation of equality $P = Q$ of predicates

$P,\ Q \in M_k$. The equality relation of predicates can be considered as the diagonal relation set on the Cartesian square of set M_k, i.e. as set of all pairs a kind $(P,\ P)$ where $P \in M_k$. In our example the set $\{(P_0,\ P_0),\ (P_1,\ P_1),...,\ (P_7,\ P_7)\}$ serves as equality relation. the Equation $\mathsf{D}_k(P,\ Q) = 0$ sets the inequality relation $P \neq Q$ predicates P and Q. The inequality relation of predicates can be considered as the antidiagonal relation that set on the Cartesian set M_k.

Let's introduce a predicate of equality of ideas D_k on set $S_k \times S_k$, defining it for anyone $x,\ y \in S_k$ as follows:

$$D_k(x,\ y) = \mathsf{D}_k(\Phi(x),\ \Phi(y)). \tag{2}$$

Here Φ is bijection that display set S_k on set M_k. The predicate $D_k(x,\ y)$ predicate $D_k(x,\ y)$ displays set $S_k \times S_k$ on set Σ. Being sent from definition (2) and using equality and inequality relations of predicates, we can present a predicate D_k as

$$D_k(x,\ y) = \begin{cases} 0, \text{if } \Phi(x) \neq \Phi(y), \\ 1, \text{if } \Phi(x) = \Phi(y). \end{cases} \tag{3}$$

Let's consider two models $\langle S_k,\ D_k \rangle$ and $\langle M_k,\ \mathsf{D}_k \rangle$. First of them represents set S_k together with the predicate D_k set on its Cartesian square, another - set M_k together with the predicate D_k set on its Cartesian square. Equality (1) means, that models $\langle S_k,\ D_k \rangle$ and $\langle M_k,\ \mathsf{D}_k \rangle$ are isomorphic each other. The relation of isomorphism of models is equivalence.

We shall make some specifications of the introduced terminology. Ideas we shall name, in the first place, mathematical objects - elements of set S_k, at the second place, psychological objects - any subjective conditions of the person. In the second meaning the term idea we shall use only at the expanded statement of tasks of the intelligence theory. We shall name psychological objects by ideas - all those subjective conditions of the person which can be expressed in the form of statements.

The signals showed to the examinee during carrying out of experiences, we shall name physical stimulus. We shall speak, that physical stimulus serve as prototypes of ideas, and ideas are images of physical stimulus. At narrow problem definition in a role of physical stimulus will act statements, and in a role of their images will act only ideas. At expanded problem definition stimulus can be any physical objects.

Properties of a equality predicate of ideas

Let's consider properties of a predicate D_k. It submits to laws of reflexivity, substitution, symmetry and transitivity. In formal record these laws look like the following logic equations:

$$\forall x\ D_k(x, x) = 1, \tag{4}$$

$$\forall x \forall y\ (D_k(x, y) \supset D_k(y, x)) = 1, \tag{5}$$

$$\forall x \forall y \forall z\ (D_k(x, y) \wedge D_k(y, z) \supset D_k(x, z)) = 1, \tag{6}$$

$$\forall R_k \forall x \forall y\ (R_k(x) \wedge D_k(x, y) \supset R_k(y)) = 1, \tag{7}$$

Here, variables x, y, z are set on set of all ideas S_k, the variable R_k is define on set of all predicates which are determined on set S_k. The variable predicate connected by the logic equations (4) - (7) designates by symbol D_k.

We have defined a equality predicate of ideas D_k and have deduced its four properties, being sent from a equality predicate of predicates (1) and using expression (2). However, it would be desirable to construct the approach of equality of ideas on the bases, not dependent on concept of a final predicate which in our statement carries out only auxiliary role of the prototype of concept of idea.

As it is proved in the statement resulted below, it can be made, basing on properties (4) - (7) of equality predicates of ideas as on axioms. Value of the statement will be, that it gives axiomatic definition of a predicate of equality of ideas.

Statement 1.

To present in form (1) predicate D_k, that defines on set $S_k \times S_k$, it is necessary and enough that it satisfied to conditions of reflexivity, symmetry, transitivity and substitution.

That is why any two models that isomorphic the third are isomorphic each other. We shall take models $\langle S'_k,\ D'_k \rangle$ and $\langle S''_k, D''_k \rangle$. Both of them are isomorphic to model $\langle M_k,\ \mathsf{D}_k \rangle$, so they are isomorphic to each other.

From here follows the existence of bijection $\Omega:\ S'_k \to S''_k$, for which at anyone $x,\ y \in S'_k$ takes place the equality:

$$D'_k(x,\ y) = D''_k(\Omega(x),\ \Omega(y)). \tag{8}$$

Expression (8) means, that in abstract sense predicates of ideas equality, and, consequently, relations of ideas equality, that appears in any algebras of ideas of the same dimension, are indistinguishable from each other. Insignificant distinction from the mathematical point of view consists only in a concrete way of a designation of elements of set S'_k and S''_k of carriers of these algebras. If we replace names of set S'_k elements with names of set S''_k elements by bijection Ω the predicate of ideas equality D'_k, which is set on set $S'_k \times S''_k$ will turn in a predicate of ideas equality D''_k, which is set on set $S''_k \times S''_k$.

The equality predicate $D_k(x,\ y)$ of ideas x and y is practically realized by the examinee in a series of experiences. Every experience consists of researcher suggestion to examinee of two ideas $x = a$ and $y = b$ which are showed in the certain order so that examinee always knows what is the first of them and what is the second. He needs to compare the ideas showed to him and to establish, they are equal or not. In case of full concurrence of ideas *a* and *b* the examinee reacts the answer 1 if they are differ in something the answer will be 0. Experience shows, that the examinee recognizes two ideas equal in all those and only those cases when statements that express these ideas are logically equivalent.

When we define the algebra of ideas formally, we have introduced the set of all ideas S_k and only after that have set on it an equality predicate $D_k(x,\ y)$ for any ideas $x,\ y \in S_k$. At substantial introduction of algebra of ideas (i.e. such algebra of ideas at which a role of ideas play ideas of the person) it is necessary to make on the contrary: first to introduce an equality predicate of ideas, and then the set of all ideas with help of this equality predicate.

The researcher has no direct access to ideas of the examinee. Therefore he is compelled to find set of ideas of the examinee, basing exclusively on supervision results of examinee behavior. The researcher can act as follows. He shows to the examinee various pairs physical signals which from his point of view can carry out a role of ideas names, and suggests examinee to establish, are equal or not ideas that corresponding to these signals.

Thus the researcher, first of all, should find out, is examinee capable to react on those or other pairs of signals. If it appears, that examinee always reacts by quite certain answer on some pair of entrance signals, the researcher, should establish, will be a reaction of examinee on this pair of signals unequivocal or not.

With this purpose the researcher in a random way between other pairs of signals, repeatedly shows the same pair of signals that is interesting for him. If the examinee reacts once to this pair of signals the answer 0, and other time - the answer 1 the signals of such pair should not be included in structure of set S_k as names of ideas. So, using a predicate of equality as the tool, the researcher forms the set of all ideas for the given examinee. It is necessary to specify, that actually the researcher collects in set S_k not ideas of the examinee, but names of these ideas. If for any idea have been used several different names, the researcher select only one of them. If the researcher puts before itself any private tasks he can be limited to revealing not all ideas of the examinee but only some part of then that is interesting for him, for example, ideas of mathematical character.

Conclusion

On the basis of algebra single k-dimntional predicates of the first order is offered the algebra of ideas that intended for formalization of subjective conditions of the person.

The algebra of ideas structure is developed: the carrier of algebra and its axiomatics.

The equality predicate of ideas is introduced as the tool for experimental studying ideas of the person, the axiomatics of this predicate is determined.

Bibliography

[Bondarenko, 2000] M. Bondarenko, Z. Dudar, Yu. Shabanov-Kushnarenko, N. Protsai, V. Cherkashin, V. Chikina Algebra of predicates and predicate operations // Radio electronics and informatics. – 2000. - №3. – p.15-23.

[Shabanov-Kushnarenko,2005] Shabanov-Kushnarenko Yu. / Brainlike computing // Shabanov-Kushnarenko Yu., Klimushev V., Lukashenko O., Nabatova S., Obrizan V., Protsay N. // Proceedings of the East-West Design and Test Workshop. Odessa, Ukraine, September 15 – 19, 2005, p. 274 - 279.

[Ivanilov,2007] A. Ivanilov. Relational algebras and algebras of predicates / A. Ivanilov, Yu. Shabanov-Kushnarenko // The East Europe magazine of high technologies. – 2007. - № 4/2. – p. 43-48.

Authors' Information

Olga Kalinichenko *– PhD, Associate Professor of Software Department KNURE; Ave. Lenin 14, 61166, Ukraine; e-mail: okalinichenko@mail.ru*

THE DEVELOPMENT SUPPORT SYSTEM "ONTOINTEGRATOR" FOR LINGUISTIC APPLICATIONS

Olga Nevzorova, Vladimir Nevzorov

***Abstract**: The article is described architecture and functionality of development support system "OntoIntegrator" for NLP.*

***Keywords**: Natural language processing, ontological models, program tools for NLP*

***ACM Classification Keywords**: H.3.1.Information storage and retrieval: linguistic processing*

Introduction

The intelligent development support system "OntoIntegrator" is a ontolinguistic research integrated environment for NLP using complicated structured ontological models.

The system "OntoIntegrator" is a specialized systems for the decision of next main tasks: design of ontological models with free structure (supporting of large-scale of data visualization); modelling of applied tasks used natural language texts; NLP based on ontological and linguistic models.

The system "OntoIntegrator" is focused on the application development used NLP and realizes the onto-linguistic approach integrating the next main processes: new domain adaption; design of task decision based on ontological models: deep linguistic analysis; knowledge representation in the ontological models and specialized knowledge bases; modularity and extensibility of toolkits.

Applications spectrum of developed technologies is wide: text information structuring and visualization; information extraction from texts for computer-aided database updating; systems for automatic document annotation; information retrieval systems; document classification systems; knowledge extraction systems.

Architecture and functionality of development support system "OntoIntegrator"

Next important methods of software development were used for design of intelligent development support system "OntoIntegrator" which determine the efficiency of developed technological solution:

- using of Clarion development technology which represents the projected system as the system of hierarchy levels. The levels differ the degree of abstractness (database drivers, database dictionaries, application conceptual framework, basic templates of object-oriented programming, process-oriented programming).
- method of toolkits vertical integration which provides links and compatibility under control between different software tools.
- method of software tools balance which provides "necessity and sufficiency" for each tool (table and graphic ontology editors, import/export data processors, search engine, logical inference processor and others).
- method of horizontal integration of data processing models and inference models which provide data specifications compatibility.

Main functional subsystems of system "OntoIntegrator" is shown in Fig. 1.

The system includes next main functional subsystems: the subsystem "Integrator"; the ontology development subsystem "OntoEditor"; the subsystem "Text Analyzer"; the subsystem for supporting external linguistic recourses; the subsystem of ontological models.

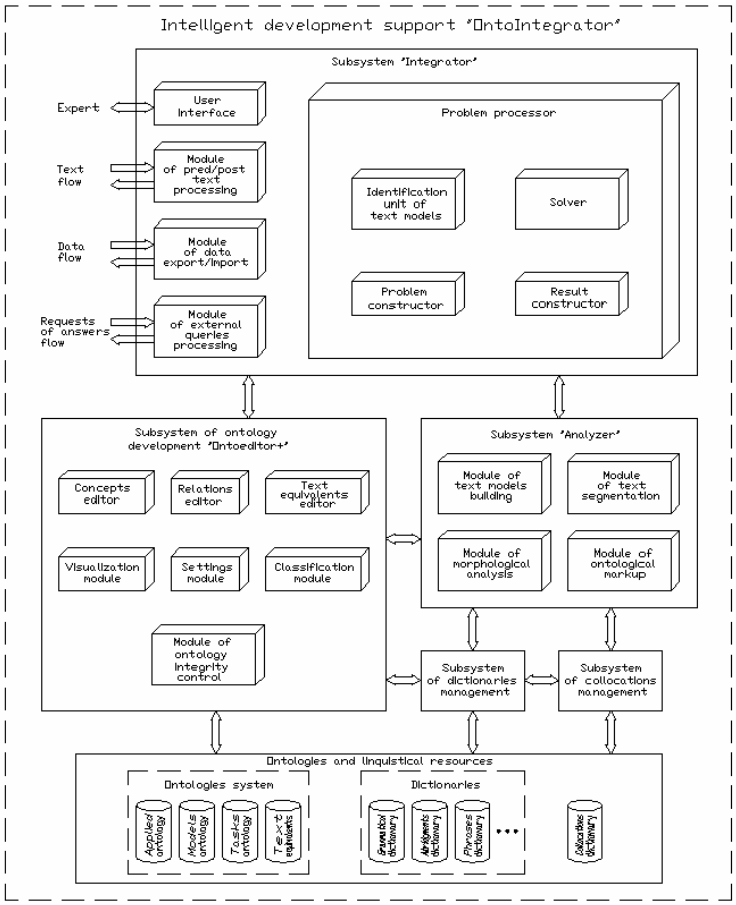


Figure 1. Architecture of system "OntoIntegrator"

A brief survey of main functional characteristics of listed systems is given below.

The ontology development subsystem "OntoEditor" [Nevzorova et. al., 2004] supports main table functions for ontology processing (editing the records; automatic record correction; supporting the using of several ontologies including mixed ontologies (for instance, the ontologies which have the common lists of relation types, classes, synonymic rows and others); import the ontologies in different data formats from external databases, gathering automatically statistics of ontology objects; ontology filtering; search the chains of concept relations with fixed properties and others).

Fig. 2 shows the graphical form with concepts hierarchy of applied ontology. Different form components show information about statistical characteristics of hierarchy including the list of ontology concepts unclassified in hierarchy; set of filtering buttons for hierarchy display; set of buttons for editing concepts hierarchy.

The subsystem "Text Analyzer" includes linguistic tools which are meant for solution different linguistic tasks: morphological analysis, ontological markup, disambiguation , segmentation, building text models for applied purposes. The solution of these tasks is based on the technologies of ontological models and external linguistic recourses interactions [Nevzorova, 2007]. In [Nevzorova et. al., 2007] the technologies of homonymy disambiguation used in the system are described.

The process of building (extraction) of text models (set of T– components) is based on the results of segmentation. The different types of segments are interpreted as the different T– components (T– components interpreted as the properties models of domain; T– components interpreted as examples models of domain; T–components interpreted as tasks models of domain).

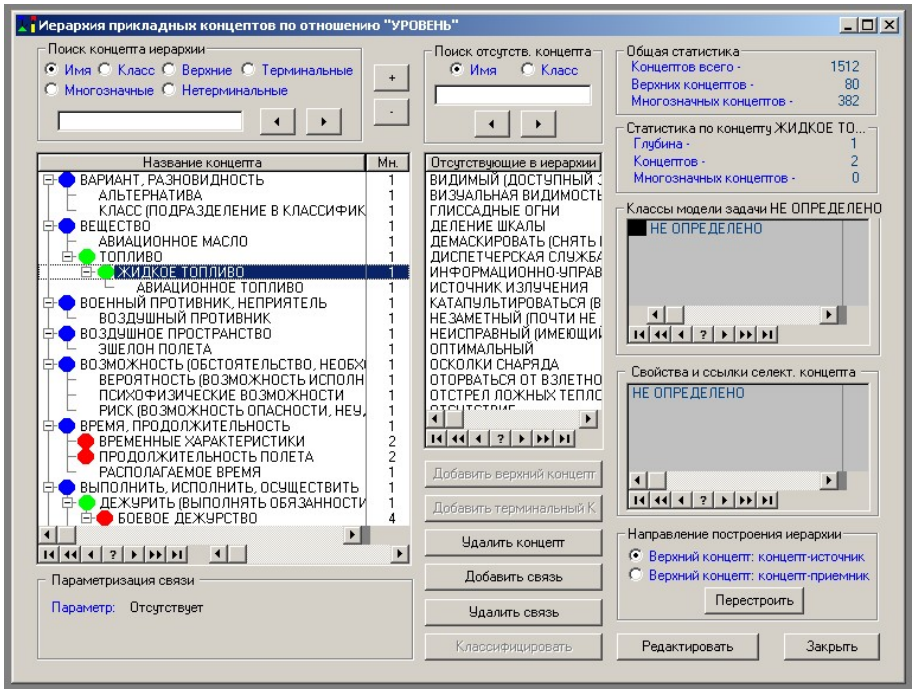


Figure 2. Graphical form of the subsystem "OntoEditor"

Fig. 3 shows the results of text analysis represented in database table structure. Information about lexical items, grammatical characteristics of word forms, references on concepts of ontological markup, results of homonymy disambiguation based on homonymy index, results of segmentation are included in table columns. The set of buttons are meant for editing table records, showing the results of segmentation and information about text statistics.

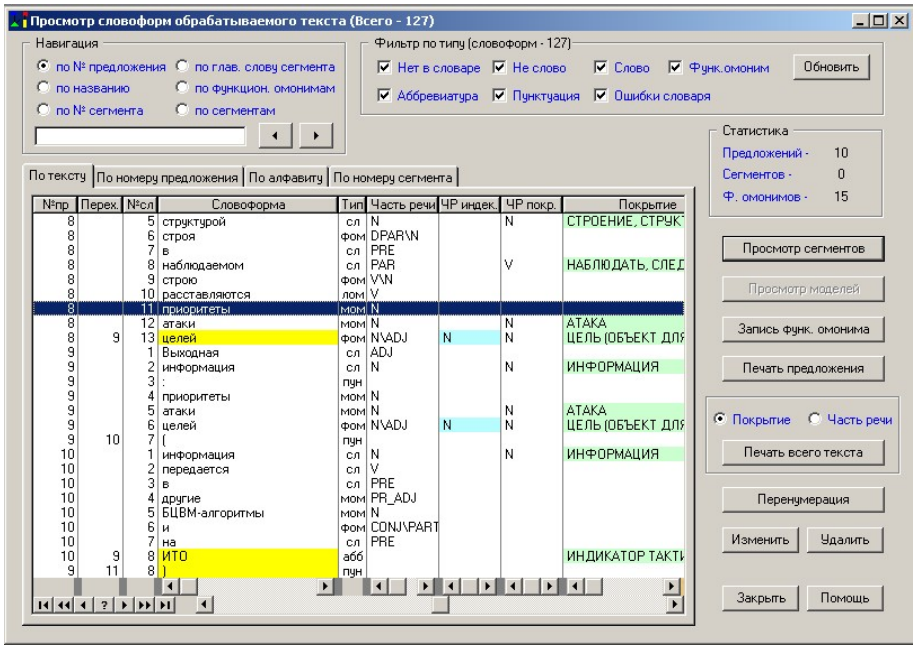


Figure 3. Graphical form of the subsystem "Analyzer"

The subsystem for supporting external linguistic recourses is meant for supporting of main linguistic recourses including marked grammatical dictionary and different specialized databases. Grammatical dictionary involves various markups included semantic markup that used in the algorithms of text analysis. The example of specialized databases is the database of collocation used for homonymy disambiguation in Russian. The

collocation context allows to disambiguate homonyms included in given collocation. The method of homonymy disambiguation based on collocations models uses the search algorithms of collocation in text and assigns pre-defined grammatical characteristics to homonyms included in collocation model.

The subsystem of ontological models includes different ontologies types (applied ontology, the models ontology, the tasks ontology). The core of ontological system is the models ontology.

The concepts of models ontology are being created and edited by user in table mode or graphical mode. The concepts have the next mandatory attributes:

- name (any string). For link to T-components is used a set of text equivalents of name.
- functional class (F-class). The next F-classes are used:
  - basic class (the models of given class are built automatically in any new ontology);
  - derived (custom) class included the models of special purpose.
- color (for graphical mode);
- contents (the description of concept-model).
- semantic class. The next semantic classes are used:
  - relation;
  - property;
  - constructed by user;
  - nonsemantic;
  - referential.

All types have specific visual markings.

The concepts of tasks ontology are being created and edited by user in table mode or graphical mode. The concepts have the next mandatory attributes:

- name (any string). For link to T-components is used a set of text equivalents of name.
- functional class (F– class). The next F– classes are used:
  - operation (operation is being imported to new tasks ontology);
  - input data(input data determine information source for operations);
  - result (result determines the output result of operation);
  - task (the concept determining the structure of task decision).
- contents (the description of concept-task).

Any concept-task may be added as the subtask to another concept-task. The link between tasks is realized as the relation "operations sequence". The relation "operations sequence" is implemented with the attribute "sequence number", the operation metric is used for parameters transmission between operations.

The concepts of applied ontology are being created and edited by user in table mode or graphical mode. The concepts have the next mandatory attributes:

- name (any string). For link to T-components is used a set of text equivalents of name.
- abstractness (the attribute is used for research purposes).

Applied concepts may be classified, i.e. applied concept is being linked to some class (some concept-model). At that additional links (properties, references) arise in relation to description model.

Current version of software system "OntoIntegrator" was used to the developing of applied ontology (avia-ontology). Avia-ontology was developed for the domain which describes behavior of both an operator (air crew)

and board equipment in various flight situations. Avia-ontology is a hierarchical concepts network (1600 concepts, 4700 terms) [Dobrov et. al., 2003].

The subsystem "Integrator" contains software tools for assignment of system interface settings including set of interface functions for settings of graphical mode of ontology (selection of object icons, object color; selection of functions being linked with mouse buttons and keyboard combinations). Another functional abilities support the functions of initial text processing (text corpora loading into databases, automatic statistics gathering for all linguistic objects; sentence segmentation; abbreviation recognition; building of linguistic shell of ontology and others), functions of data export/import, functions for processing of external text queries. The core of given system is problems processor that includes problems constructor, identification unit of text models, problem solver and result constructor. Functioning of problems processor is based on ontological models system.

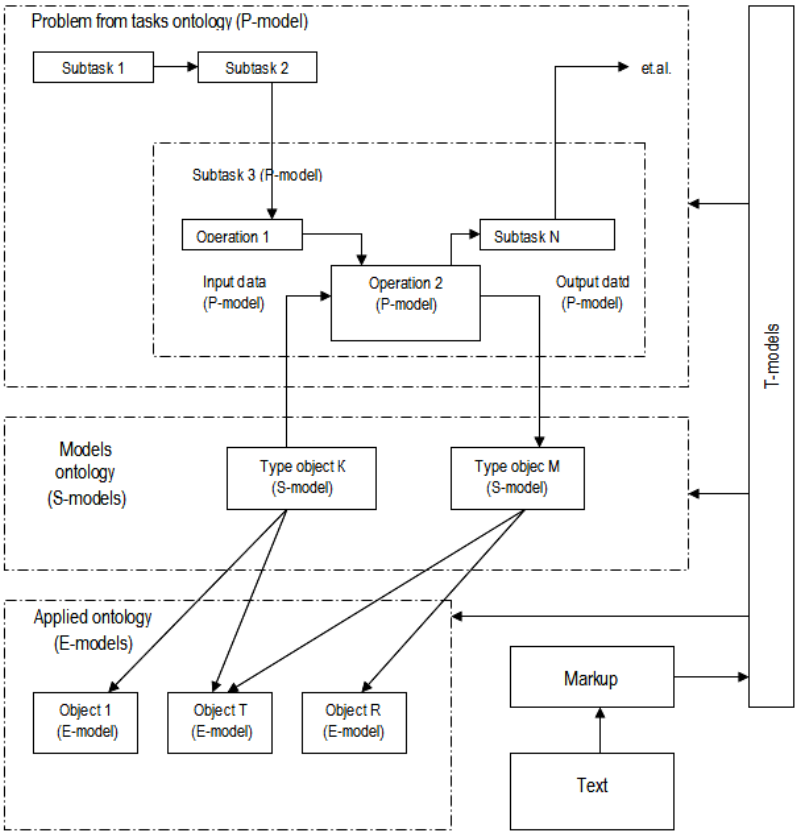


Figure 4. Task decomposition

The problems constructor performs decomposition of applied task into the structures of tasks ontology (set of P–components). The components of tasks ontology go into some set of structures of models ontology (set of S–components). The components of models ontology go again into the set of structures of applied ontology (set of E–components). Identification unit of text models establishes a one-to-one correspondence between the set of text models (set of T– components) and sets of structures of tasks ontology, models ontology and applied ontology. The problem solver forms a problem-solving process as the sequence of P– components based on built mapping.

Fig. 4 shows the method of problem decomposition into structures of ontological system. Input task (P– model) is being represented in the form of solvable subtasks sequence that are linked by consequence relation. Each subtask ((P– model) goes into the set of S– models (task operations). Each operation goes into the set of E–models that realize the operation.

Selection of P,S,E–models is based on the T-models that correlate with P,S,E–models. Recognition of T–models is implemented by the methods of linguistic analysis including text segmentation method and method of ontological markup [Nevzorova, 2007].

Reference method is used for different tasks of linguistic analysis including disambiguation, segmentation, ontological markup and others. The software tools of system "Integrator" support realization of decomposition method (development of concepts of task ontology, models ontology and applied ontology; building the mapping of various ontological levels and others).

Solution of applied linguistic problem "Analysis of functional homonymy"

The system is oriented to process linguistic tasks including as standard linguistic problems (morphology analysis, disambiguation, syntactical analysis) as well as applied problems (knowledge acquisition. annotation, classification). The solution of these problems is based on the method of problem decomposition into structures of ontological system. We will describe the using of method for applied linguistic problem "Analysis of functional homonymy" (Fig. 5).

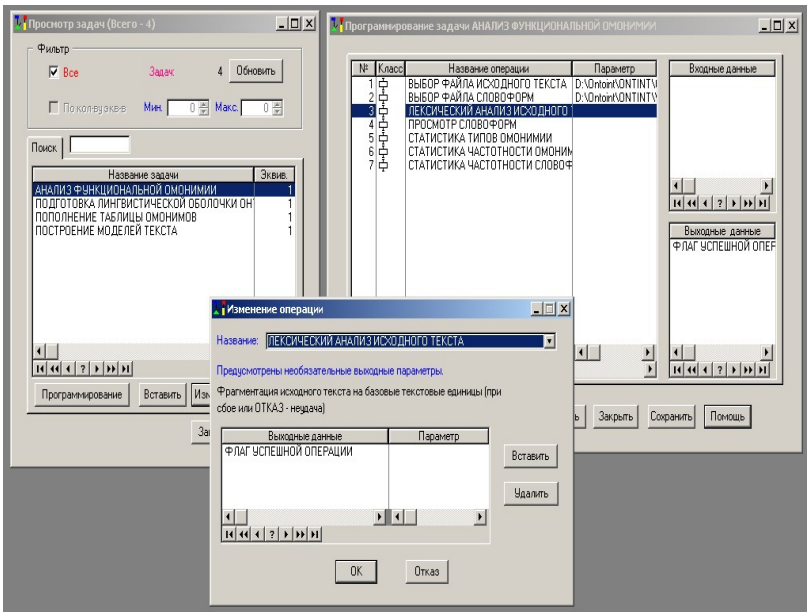


Figure 5. Development of solution for the task "Analysis of functional homonymy"

Given task is one of the basic tasks that the system "Analyzer" performs during text analysis. The objective of this task is to disambiguate functional homonymy using different applied methods. Disambiguation of functional homonymy is being implemented by group of methods [Nevzorova et. al., 2007]:

- method of disambiguation based on linguistic shell of applied ontology;
- method of disambiguation based on homonyms index of database of collocations;
- method of disambiguation based on context rules.

The solution of given problem is described as sequence of definite operations (sequence of P– components). Fig. 5 shows the sequence of given P– model: lexical analysis of text, show the word forms, flag of successful operation and others. All operations are being linked by consequence relation that is belong to S–models. Mapping of T–models into P,S,E–levels is being realized by procedures of morphological analysis. The development of solution is realized in special interactive mode in which the concepts of ontological system are being selected and the links between concepts are being set up. First two methods have the precision estimate 100 %, the third method have the precision estimate about 95% for different types of functional homonymy.

Conclusion

In the article we discussed the methodology of development of NLP–problems solution based on interactions of ontological models system. Given methodology was realized in program tools research environment "OntoIntegrator" oriented for NLP. At the present time discussed methods are successfully used for NLP including these tasks as disambiguation, text segmentation, ontological markup and others. Program tools of system "OntoIntegrator" support the realization of method of decomposition (development of ontology tasks concepts, ontology models concepts, applied ontology concepts; mapping of different ontological levels and others).

Acknowledgements

The work has been completed with partial support of Russian Foundation Basic Research (grant № 08-06-00183).

Bibliography

[Dobrov et. al., 2003] B. Dobrov, N. Loukachevitch, O. Nevzorova An Approach to New Ontologies Development: Main Ideas and Simulation Results // Int. Journal “Information Theories & Applications”. – 2003. – Vol.10. – P.98-105.

[Nevzorova et. al., 2004] Nevzorova O.A., Nevzorov V.N. Analysis of the Structures of the Linked Objects by “OntoEditor” Software Toolkit // Proc. of the 6-th Int. Workshop on Computer Science and Technologies CSIT’2004, Budapest, Hungary, 2004. – P.177-181.

[Nevzorova et. al., 2007] Nevzorova O., Nevzorov V., Zinkina J., Pjatkin N. Integral Technology of Homonymy Disambiguation in the Text Mining System "LoTA"// Proceedings of Fifth Int. Conf. Information Research and Applications, Varna, Bulgaria, 2007. – Vol. 1. – P. 129-133.

[Nevzorova, 2007] Nevzorova O.A. Ontolinguistic system: some technologies of interaction with applied ontology //Interactive systems and technologies. - Collection of scientific papers. – Ulyanovsk: UlSTU, 2007. – P. 44-51.

Authors' Information

Olga Nevzorova *– Chebotarev Research Institute of Mathematics and Mechanics, Tatar State University of Humanities and Pedagogics, Kazan, Russia; e-mail: olga.nevzorova@ksu.ru*

Vladimir Nevzorov *– Kazan State Technical University, Kazan, Russia ; e-mail: nevzorov@mi.ru*

THERMALS MAP – ASSIST FLIGHT SYSTEM

Arkadiusz Rzucidło

***Abstract**: The article presents proposition of the assist pilot system. The system helps pilot with correct decision of cross country flights. Presentation of specially calculated information stored by the system delivered to pilot could help him with his decision about correct directions of crossing.*

***Keywords**: Knowlage base, Data mining*

***ACM Classification Keywords**: E. Data, E.5 Files, E.5.6. Sorting/searching*

Introduction

Dream about flying has been following in most of people through the years. Some people make this dream came true. By those dreams, people all the time have been trying to get to the clouds. The ways of their trip were different. Human constructed a glider, plane, rocket and other flying machines as answer of this needs. But in every cases life showed them that flying is not as easy as it looks like. Birds and they behaviors, that people follow all the time, are still the question. How they do that ?

When people get to the air they tray to stay there as long as they can. Sometimes it was easy, sometimes not but all the time people ask how to get to the sky faster and stay there higher and longer. The observations and all of experiments became a science. The flying has been a topic of thesis and experiments. Lots of later theories, analysis and practices were based on bird's behaviors. At the end of all of experiments people discovered instruments which helped them with conquest their nature. But no matter how the instruments ware perfect flaying was always a big mystery. This mystery was a fuel for development of flying science.

Flying, in his pure form like soaring and thermal flying is a good place for many experiments and thesis. These drive to working out solutions which make flaying easy. These solutions help pilots make a final decision during their flights. Decisions are different. Sometimes they realize on carrying about temporary direction, height of the flight or choosing the landing place. But sometimes they are more complicated and depend on many things. That's decisions are usually critical aspects of flying, specially when the security is considered.

Flying on different kinds of gliders is what we call as a free flying. That means that the pilot doesn't need anything but pure nature forces like thermal or soaring wind. Many things which help pilot with his trip are necessary, specially when the flight is cross county type.

The main topic of this article is finding the system which helps to choose the right direction of the flight. This paper focuses on paragliders. With comparison to classic glide and delta-glide this constructions have a number of limits like: fly speed and glide ratio. Because of this limits paraglider pilot's decisions are always based on large knowledge about flying in different conditions. The way of crossing area where the fly is made is an individual tactic of flight. In many cases pilot sees the area of flying for the first time, so his knowledge about tactic is poor. His flights are usually not as long as they are supposed to be. The question is: how to get the special tactic knowledge for pilot ? How to decide which direction is correct ? These and many other questions connected with the flight have no answers.

This problem seems to have no solution but some of specially calculated information delivered to pilot could help him with his decision about correct directions of crossing. All cross country pilots after taking-off plan their trip. In many cases the plan is based on studying the log files of other pilot's flights. The files are stored in Internet site of XCC system where pilots report they flights in. It is a good source for any study in this topic. Files have the same format made in world standard named igc. Internet source of log files is public available.

Knowledge base

Mentioned XCC (http://xcc.paragliding.pl/) is not the only place of gathering data about paragliding flights. In the Internet there are many other similar web pages like XContest (http://www.xcontest.org/world/en/) or local data bases. They also use mentioned files format (igc). The greatest advantage of IGC format is its universality and that it is used by every XC pilot worldwide, which is a big help in flights analysis.

...
B1405365448657N01822350EA0000000043
...
where:

B140536 - utc Time 14h 05m 36s (T)

GPS position:
5448657N - 54-48.657 N (X)
01822350E - 018-22.350E(Y)

Height
A
00000 – height by barometer
00043 – height by GPS (Z)

Figure 1. Flies format igc [Davies, 1995]

IGC is type of text file. Thanks to quality of files gathered information are easier to analyze. There are no additional interfaces that usually cause problems in data reading. Data record igc is not complicated. It consists records, which are sequence of pilot's position in four dimension space (T,X,Y,Z) (Fig.1). Such accuracy allows fix route of flight. It also enables to check if there was a record of descending or ascending in the area where pilot was flying.

This information is essential in data analysis in topic of this article. Obtained geographic coordinates allow us to place and estimate behavior of flying object in space. Analysis of greater amount of data on set penetrated area give configuration of track points from particular flights in effect. This points can be taken as a compilation, which will be used to form areas being potential zone of ascend of zones thermal inactive. Mentioned groups of points can be presented in time and thus we can have changeable configuration of registered ascends. Time of record of geographic point makes these zones more dynamic.

Project will use files saved during flights in mountains. This area lets to cut down number of factors that influence the flight. Mountains and valleys determine wind direction and in some level limit variables like for instance ground contrast connected with character of grown plants which have influence on generating lifts as well as many other important weather conditions [Burkhard, 2007]. This thesis are concentrated on visualization of thermal active points as a "ascend map", not on analysis of arising those places and points in GPS tracks.

Analysis of data from tracklogs proposed in this article is a stage of initial recognition of prognosis possibilities and support navigation activities. In descriptions there were used only the most basic parameters. Many environmental factors that can have influence on flight were not taken into consideration. Full analysis of data is not possible due to too many mentioned variables and their impossible to predict character.

Characteristics of ascend map

Data in IGC files are often recorded in set time interval. Recording device to each point of GPS coordinates applies time value and saves it in standard UTC format. This is first value, starts next point in IGC file.

Time factor is used to display thermal active points on ascend map and record them in set moment of flight, thus to show the pilot probability of finding lifts in certain point and time. GPS coordinates enable for special orientation within first recorded point that starts the lift. Further points, where pilot stays in lift, have minor meaning due to the fact that thermal is being drifted by wind and its drift depends on wind direction. Different weather factors and pilots individual preferences of circling in thermal are also important. Therefore registration of all „lift points within thermal" can darken the map and not bringing any benefits in creating the map. GPS coordinates show position of finding a lift in horizontal setup with additional value – height. Horizontal setup places point of first contact with lift. This is position registered, what means that certain point will not be the point of triggering the thermal but only point of finding a lift by pilot. Registering height value during flight will show in three dimensional space pilot found a lift. When many tracks near to this point will be analyzed this will enable to visualize air activity in this area. The more tracks the easier to find places where it is the most probable to find a lift. Using time value helps us to show thermal activity of area while creating ascends maps. Using time also benefits in possibility of showing map not only in hours scale but also days, weeks, years scale. It makes easier to create statistics and detailed analysis of reasons for arising of "hot" – thermal active and inactive – "cold" places.

Technical details of ascend map

Creating ascend map process consists of a few stages. First stage is gathering data for marking interesting points. Choosing process based on data contained in IGC files. These data have to be initially processed by picking only lift zones from whole tracks. Chose method is classic comparison of forward recorded heights and picking first meaningful point which started the lift. Points selection is realized by differential method by comprising next track points.

During flight pilot can notice short turbulences connected with air instability and activity of thermals. These air movements are usually short and natural [Burkhard, 2007],[Suchan, 2001]. They are meaningless for gaining height, therefore they will be put aside.

Due to the fact that turbulences are making distraction and only blur the map is proposed not to take them into account. Therefore choosing of points, as previously, will be made by differential method, however point of first lift will be marked as first – beginning set time in which progressive values of lift, are being registered. If value of lift remains forward, points are showing growing value of height and points are being registered in 15 seconds interval. Fixed lift and starting point has being chosen as a start of lift. If within 15 seconds lift disappears registered starting point is treated as start of turbulence and it won't be taken into consideration when making the map.

After track analysis chosen thermal active points are compilation of several meaningful ingredients added to database. This base gathers all points as overall repertory for “ascends map”. The base holds main table containing active places coordinates and height as well as tables with additional supplementary data and selection of correct interesting points. These is information about start place, route etc. Number of additional data can be modified depending on needs. It doesn’t affect on „ascend map” directly but only makes data presentation more interesting.

Presentation of map

Presentation of points on map can be shown with very popular Internet instrument - Google Earth. Selected group of points has to be processed to KMZ format. This is Google Earth standard. Created this way group of points give us static picture that show places recorded in database as thermal active points. It is example of web interface being general view to analyze before flight. Target interface for “ascend map” is the one that is dynamic application working on gathered data. The idea of this system is to direct pilot on closest lift zone based on his actual position. In basic form interface will generate digitally flight parameters to get to thermal active point in “ascend map”. These parameters are: coordinates of lift from “ascend map”, its height and distance to lift. Parameters will show to the pilot exact point where lift had been recorded in the map. System’s role is to be a flight assistant, pilot can not depend on it absolutely and final decision should bee made based on pilots experience. He chose flight path based on actual weather conditions. System only gives a hint where are places thermal active with high probability of finding a lift.

Conclusion

In thesis it was presented proposition of using database with flights on example of application being flight assistant. This application would be additional tool helping young pilots to choose right route during cross country flights. In the article there was suggested data converting process for ascend map and two examples for data presentation

Bibliography

[Burkhard, 2007] M. Burkhard, Thermal Flying - a guide for paraglider and hang glider pilots, Gaissach, Germany 2007

[Suchan, 2001], A. Suchanowski, Paralotnia. Sztuka żeglowania w przestrzeni, PIR, 2001

[Davies, 1995], E. Davies, EW Avionics IGC File Format, http://www.edavies.nildram.co.uk/ew/ewa-igc.htm

Authors' Information

Arkadiusz Rzucidło *– Editor , Department of Computer Science FMEA RUT; W. Pola 2, 35-959 Rzeszów, Poland ; e-mail: arzucidl@prz.edu.pl*

Industrial Control and Monitoring

APPLICATION OF GENETIC ALGORITHMS TO VECTOR OPTIMIZATION OF THE AUTOMATIC CONTROL SYSTEMS

Valery Severin

Abstract: *Methods for calculation of quality indexes for automatic control systems are presented. For the optimization of quality indexes defined only in a stability domain a vector objective function of varied parameters of the system is proposed. The stepwise principle of successive satisfaction of constraints for the passage into the definition domain of quality indexes is considered, as well as a rational mechanism of its realization in the form of the priority optimization of the vector objective function. For the optimization of the vector objective functions genetic algorithms as vector optimization methods are presented. Their application allows one to steer the optimization process from any initial point of the space of varied parameters into the stability domain of the system and to find the optimum of the quality indexes in this domain. The efficiency of the proposed application of vector genetic algorithms for the quality indexes optimization is confirmed by computational experiments.*

Keywords*: genetic algorithms, vector optimization methods, automatic control systems, quality indexes.*

ACM Classification Keywords*: G.1.6 Optimization - Nonlinear programming*

Introduction

For the synthesis of automatic control systems (ACS) quality of their functioning can be presented by different criteria [Besekerskiy, Popov, 2004]. There are many difficulties even at the synthesis of the linear control systems, and the results of synthesis substantially depend on the applied criteria [Poljak, Scherbako, 2005]. Yet the task of synthesis of the nonlinear systems is even more difficult.

Quality criteria for both linear and nonlinear ACS can be reflected by the direct quality indexes (DQI) and improved integral quadratic estimations (IQE) [Severin, Nikulina, 2004], [Severin, 2004], [Severin, 2005]. The optimization tasks of these criteria have an identical feature — their objective functions domain is limited by stability conditions [Severin, 2008]. Therefore the standard optimization methods can not be effectively used for the optimization of ACS quality indexes.

For the synthesis of the linear control systems by means of DQI and improved IQE optimization the vector objective functions are offered and for their optimization the direct methods of unconstrained minimization are modified [Severin, 2005]. The methods optimization laboratory OPTLAB is developed in MATLAB system [Severin, 2009]. However, the offered vector optimization methods do not allow to find global extremes.

The decisions search of optimization tasks in very large and difficult domains is executed by genetic algorithms which are evolutionary computations variety and behave to heuristic search methods [Voronovskiy, Makhotilo, Petrashev, Sergeev, 1997], [Setlak, 2004]. Genetic algorithms are used for optimization tasks decision based on the principles and mechanisms, reminding biological evolution [Panchenko, 2007], [Weise, 2008].

The purpose of this paper consists of conception development of synthesis for linear and nonlinear ACS on the basis of DQI, improved IQE and modifications of genetic algorithms for vector optimization.

The methods of DQI and improved IQE calculation are considered for linear and nonlinear systems. Statements of systems optimization tasks are presented with using vector objective functions. Modification of genetic algorithms is offered for vector optimization of ACS quality indexes.

Quality Indexes Calculation of Linear Automatic Control Systems

Let linear model of ACS, depending on a vector of varied parameters $x \in R^p$, in state space looks like:

$$\partial X(x,t)/\partial t = A(x)X(x,t) + B(x)u(t), \qquad y(x,t) = C(x)X(x,t), \tag{1}$$

where $X(x,t)$ is a state vector with an initial condition $X_0 = 0$; $u(t)$ is entrance influence, $y(x,t)$ is control output co-ordinate; $A(x)$, $B(x)$, $C(x)$ are matrices of the control system parameters. For the stable watching system at standard input step signal $u(t) = 1(t)$ the matrix of output $C(x)$ is set so that the condition of output coordinate scaling was executed: $y(x,\infty) = 1$. At the fixed parameters vector value x will build transient processes in model (1) on the quantum of time $[0, T_f]$. For this purpose at L integration steps of constant length $h = T_f / L$ with numbers $k = \overline{1, L}$ will enter denotations:

$$t_k = kh, \quad X_k(x) = X(x,t_k), \quad y_k(x) = C(x)X_k(x). \tag{2}$$

We will designate matrix exponent and it's integral:

$$\varphi(x) = e^{A(x)h}, \quad \Phi(x) = \int_0^h e^{A(x)\tau} d\tau, \quad g(x) = \Phi(x)B(x). \tag{3}$$

Then at an input signal $u(t) = 1(t)$ transient process in ACS with model (1) it is possible to build on the recurrent formulas of matrix method [Severin, 2008]:

$$X_k(x) = \varphi(x)X_{k-1}(x) + g(x), \quad k = \overline{1, L}. \tag{4}$$

For a deviation $z(x,t) = y(x,t) - y(x,\infty)$ there are its values

$$z_k(x) = y_k(x) - y(x,\infty), \qquad k = \overline{0, L}, \tag{5}$$

and their increments:

$$u_{lk}(x) = z_{k-2}(x) - z_{k-1}(x), \quad u_{rk}(x) = z_k(x) - z_{k-1}(x), \quad k = \overline{2, L}. \tag{6}$$

If the following condition, meaning that both successive increments are of the same sign, is met

$$u_{lk}(x)u_{rk}(x) > 0, \tag{7}$$

then with using of quadratic interpolation the extreme value is calculated $e_i(x)$:

$$d_{uk}(x) = [u_{lk}(x) - u_{rk}(x)]/2, \quad s_{uk}(x) = u_{lk}(x) + u_{rk}(x), \quad r_{uk}(x) = d_{uk}(x)/s_{uk}(x), \tag{8}$$

$$e_i(x) = z_{k-1}(x) - d_{uk}(x)r_{uk}(x)/2, \tag{9}$$

where $i = \overline{1, n_e(x)}$, $n_e(x)$ is extreme's number on segment $[0, T_f]$. By extreme's values of transient process the direct quality indexes are calculated: overshoot $\sigma(x)$, vibrations scope $\zeta(x)$, vibrations damping index $\lambda(x)$. Let $(v)_+ = \max\{v, 0\}$ is a cutting function of optional variable v. For watching system with $y(x,\infty) = 1$ direct indexes are determined on formulas:

$$\sigma(x)=\begin{cases}0, & n_e(x)=0,\\ [\max\limits_i e_i(x)]_+, & n_e(x)>0,\end{cases} \tag{10}$$

$$\zeta(x)=\begin{cases}0, & n_e(x)=0,1,\\ \max\limits_i\left|e_{2i-1}(x)-e_{2i}(x)\right|, & n_e(x)>1,\end{cases} \qquad \lambda(x)=\begin{cases}0, & n_e(x)=0,1,\\ \max\limits_i\{|e_i(x)|/|e_{i-1}(x)|\}, & n_e(x)>1.\end{cases} \tag{11}$$

For the stabilization system with $y(x,\infty)=0$, a process in which has even one extreme $e_1(x)$,

$$\sigma(x)=\max_i|e_i(x)|, \tag{12}$$

and at a calculation $\zeta(x)$ and $\lambda(x)$ in formulas (11) not taken into account $e_1(x)$.

For the calculation of ACS control time the entry times of deviation $z(x,t)$ in the set segment $[-\delta_z,\delta_z]$ of the steady-state value $z(x,\infty)=0$ are determined by verification of entrance condition:

$$|z_{k-1}(x)|\ge\delta_z\ \wedge|z_k(x)|<\delta_z. \tag{13}$$

At implementation of this condition taking into account denotations (5)–(8) auxiliary values are calculated:

$$u_i(x)=\delta_z\operatorname{sign}z_{k-1}(x)-z_{k-1}(x), \qquad v_{0i}(x)=r_{uk}(x)h, \tag{14}$$

$$s_i(x)=h\sqrt{r_{uk}^2(x)+2u_i(x)/s_{uk}(x)}, \qquad v_i(x)=\begin{cases}v_{0i}(x)+s_i(x), & v_i(x)\le 0,\\ v_{0i}(x)-s_i(x), & v_i(x)>0.\end{cases} \tag{15}$$

The moment of time, proper entrance of deviation function $z(x,t)$ in area of steady-state value, is determined:

$$t_i(x)=t_{k-1}(x)+v_i(x). \tag{16}$$

Control time $t_c(x)$ and its relative value $\tau(x)$ are calculated on formulas:

$$t_c(x)=\max_i t_i(x), \qquad \tau(x)=t_c(x)/T_f. \tag{17}$$

On formulas (3)–(17) for calculation of DQI $\sigma(x)$, $\zeta(x)$, $\lambda(x)$, $t_c(x)$, $\tau(x)$ the algorithms are obtained.

For the synthesis of watching ACS in place of few direct quality indexes it is possible to use their summarizing single index — improved IQE. On the model of kind (1) one can build a transfer function (TF)

$$W(x,s)=\beta(x,s)/\alpha(x,s), \quad \alpha(x,s)=\sum_{i=0}^n\alpha_i(x)s^{n-i}, \quad \beta(x,s)=\sum_{i=0}^m\beta_i(x)s^{m-i}. \tag{18}$$

For the watching systems a method is offered for forming of improved IQE $I(x)$ of error $e(x,t)$:

$$I(x)=\int_0^\infty[e(x,t)]^2dt, \qquad e(x,t)=\sum_{k=0}^l w_k z_t^{(l-k)}(x,t), \tag{19}$$

where l is an order of estimation, $l<n-m$; w_k are weighting coefficients:

$$w_k=\mu^{l-k}\gamma_k, \ k=\overline{0,l}; \quad \mu=t_e/t_s, \quad \gamma(s)=\sum_{k=0}^l\gamma_k s^{l-k}, \quad w(s)=\sum_{k=0}^l w_k s^{l-k}. \tag{20}$$

Here t_e and t_s are control times of etalon and standard processes, $\gamma(s)$ and $w(s)$ are standard and weighting polynomials. On TF (18) Laplace representation of error is formed $E(x,s)=\delta(x,s)/\alpha(x,s)$, where $\delta(x,s)=[\alpha(x,s)-\beta(x,s)w(s)]/s$. On the basis of this representation IQE calculation algorithm is developed [Severin, 2005].

Optimization Tasks of Linear Automatic Control Systems

Taking into account the high scope values σ_m, ζ_m, λ_m for DQI $\sigma(x)$, $\zeta(x)$, $\lambda(x)$ and requirements of maximal ACS response speed the system optimization task can be formulated as task of the constrained optimization which requires minimization of control time at implementation of limits on the other indexes:

$$\min_x \tau(x), \qquad \sigma(x) \le \sigma_m, \quad \zeta(x) \le \zeta_m, \quad \lambda(x) \le \lambda_m. \tag{21}$$

Using the improved integral estimation the task of control system optimization consists in minimization IQE:

$$\min_x I(x). \tag{22}$$

However, statements of optimization control system tasks (21) and (22) take into account neither priority of direct indexes nor limitation of their definitional domain and definitional domain of integral estimation.

The analysis of automatic control system requirements allows to set the following preference order of direct quality indexes: $\sigma(x)$, $\zeta(x)$, $\lambda(x)$, $\tau(x)$. The feature of these indexes as private quality criteria of the automatic control systems is the limitation of their definitional domain by stability conditions. On Routh criterion for stability of linear ACS with transfer function (18) there are necessary and sufficient conditions:

$$\alpha_i(x) > 0, \quad i = \overline{0,n}; \qquad \rho_k(x) > 0, \quad k = \overline{2,n-1}, \tag{23}$$

where $\rho_k(x)$ are elements of the first column of Routh table. The analysis of Routh criterion and research of properties of functions $\rho_k(x)$ justify the stepwise scheme of passage to the stability domain: if some from elements $\rho_k(x)$ is not positive, it is suggested to increase first of them to the positive value by the change of parameters values vector x, and then to increase subsequent elements. To simplify the scheme of passing to the stability domain and to meet the conditions of direct quality indexes (21), the parameter space R^p is divided into three domain sequences. The inequalities (23) and (21) are satisfied on the following domains of limitations:

$$\Omega_1 = \{ x \mid \alpha_i(x) > 0, i = \overline{0,n} \}, \quad \Omega_k = \{ x \mid \rho_k(x) > 0 \}, \quad k = \overline{2,n-1}, \tag{24}$$

$$\Omega_n = \{ x \mid \sigma(x) \le \sigma_m \}, \quad \Omega_{n+1} = \{ x \mid \zeta(x) \le \zeta_m \}, \quad \Omega_{n+2} = \{ x \mid \lambda(x) \le \lambda_m \}. \tag{25}$$

On these $m = n + 2$ domains the derived intersection domains D_k and domains of limitations levels H_k are formed:

$$D_1 = \Omega_1; \quad D_k = D_{k-1} \cap \Omega_k, \quad k = \overline{2,m}; \tag{26}$$

$$H_0 = R^p \setminus D_1; \quad H_k = D_k \setminus D_{k+1}, \quad k = \overline{1,m-1}; \quad H_m = D_m. \tag{27}$$

The domains of levels divide parameters space into the sequence of disjoint domains. The degree of violation of the first group of inequalities (23) is presented by penalty function

$$P(x) = \sum_{i=0}^{n} [-\alpha_i(x)]_+. \tag{28}$$

Stepwise principle of transferring to the stability domain and satisfaction of all limitations of direct quality indexes is based on the following: from any point x of parameters space R^p it is necessary to pass consistently to the level domain with greater index by minimizing in the current level domain using its corresponding penalty function. Taking into account the amount of levels domains there will be no more such steps of transition than the number of limitations m. For realization of stepwise principle of satisfaction of limitations in the task of ACS synthesis with optimization of direct quality indexes on the basis of levels domains, a vector objective function is introduced

$$F(x)=\begin{cases}(0;P(x)), & x\in H_0;\\ (k;-\rho_{k+1}(x)), & x\in H_k, \quad k=\overline{1,n-2};\\ (n-1;\sigma(x)-\sigma_m), & x\in H_{n-1};\\ (n;\zeta(x)-\zeta_m), & x\in H_n;\\ (n+1;\lambda(x)-\lambda_m), & x\in H_{n+1};\\ (n+2;\tau(x)), & x\in H_{n+2}.\end{cases} \tag{29}$$

Denote the first coordinate of this function as the function of level $F_1(x)$ and the second coordinate as the function of penalty $F_2(x)$. The vector objective function (29) can be calculated algorithmically.

Algorithm for calculation of the vector objective function for direct quality indexes optimization.

Input parameters: x is a vector of variable parameters, T_f is the upper limit of integration interval, L is a number of steps of integration, σ_m, ζ_m and λ_m — maximum acceptable values of DQI. Output parameter: F is a value of vector objective function. **1.** On model (1) calculate TF (18) with the characteristic polynomial $\alpha(s)=\alpha(x,s)$ of degree n. **2.** If the necessary stability conditions are violated, calculate penalty function (28), let $F=(0;P)$ and go to 12. **3.** Let $k=1$. **4.** On Routh chart calculate $\rho_{k+1}=\rho_{k+1}(x)$. **5.** If $\rho_{k+1}\le 0$ let $F=(k;-\rho_{k+1})$ and go to 12. **6.** If $k<n-2$ let $k=k+1$ and go to 4. **7.** On formulas (2)–(17) by numerical integration with quadratic interpolation calculate values of DQI $\zeta=\zeta(x)$, $\lambda=\lambda(x)$, $t_c=t_c(x)$, $\tau=\tau(x)$. **8.** If $\sigma>\sigma_m$ let $F=(n-1;\sigma-\sigma_m)$ and go to 12. **9.** If $\zeta>\zeta_m$ let $F=(n;\zeta-\zeta_m)$ and go to 12. **10.** If $\lambda>\lambda_m$ let $F=(n+1;\lambda-\lambda_m)$ and go to 12. **11.** Let $F=(n+2;\tau)$. **12.** Exit the algorithm.

Like function (29) a vector objective function is built for minimization of the improved IQE (19):

$$F(x)=\begin{cases}(0;P(x)), & x\in H_0;\\ (k;-\rho_{k+1}(x)), & x\in H_k, \quad k=\overline{1,n-2};\\ (n-1;I(x)), & x\in H_{n-1}.\end{cases} \tag{30}$$

The goal of control systems optimization using vector objective functions (29) and (30) can be presented as minimization of the function of penalty $F_2(x)$ with the priority condition of maximization of function of level $F_1(x)$, which in turn can be presented as a single task of vector optimization:

$$\min_x F(x). \tag{31}$$

Unlike the tasks of scalar optimization (21) and (22) the task of vector optimization (31) takes into account the stability conditions and order of preference of limitations. The process of optimal synthesis of ACS is grounded by minimization $F_2(x)$ with priority maximization $F_1(x)$ as optimization of vector functions (29) and (30) on the basis of comparison of their two arbitrary values $U=(U_1;U_2)$ and $V=(V_1;V_2)$ by the binary operations:

$$U<V=\begin{cases}1, & U_1>V_1\vee U_1=V_1\wedge U_2<V_2,\\ 0, & U_1<V_1\vee U_1=V_1\wedge U_2\ge U_2,\end{cases}\qquad U>V=\begin{cases}1, & U_1<V_1\vee U_1=V_1\wedge U_2>V_2,\\ 0, & U_1>V_1\vee U_1=V_1\wedge U_2\le U_2,\end{cases} \tag{32}$$

$$U\le V=\begin{cases}1, & U_1>V_1\vee U_1=V_1\wedge U_2\le V_2,\\ 0, & U_1<V_1\vee U_1=V_1\wedge U_2>U_2,\end{cases}\qquad U\ge V=\begin{cases}1, & U_1<V_1\vee U_1=V_1\wedge U_2\ge V_2,\\ 0, & U_1>V_1\vee U_1=V_1\wedge U_2<U_2.\end{cases} \tag{33}$$

These operations, allowing to determine which of the two values of vector objective function is «better», «worse», «not worse», or «not better», can be used in the numerical methods of unconstrained optimization.

Calculation of Quality Indexes of Nonlinear Control Systems

For nonlinear models the state vector and control coordinate will depend nonlinearly on the value of input influence $u = u(t)$. Unlike the linear model of ACS in state space (1), the nonlinear model can be presented as:

$$\partial X(x,u,t)/\partial t = f[X(x,u,t),u,t], \qquad y(x,u,t) = C(x,u)X(x,u,t). \tag{34}$$

For the stable watching system at an input step signal $u(t) = u_s 1(t)$ with magnitude $u_s \in [u_{\min}; u_{\max}]$ the output matrix $C(x,u)$ scales an output coordinate $y(x,u,\infty) = 1$. At a fixed value of parameters vector x let's build transient processes in model (34) on the quantum of time $[0, T_f]$ and calculate the Jacobian matrix of vector function of equation (34) by differentiating it on state vector coordinates:

$$A(x,u) = \partial f[X(x,u,t),u,t]/\partial X(x,u,t)\Big|_{x=0,\,u=0,\,t=0}. \tag{35}$$

Let's introduce notation similar to (2) and (3), but taking into account system nonlinearity:

$$\Phi(x,u) = \int_0^h e^{A(x,u)\tau} d\tau. \tag{36}$$

Transient process in the control system on a model (34) it is possible to build on recurrent formulas for $k = \overline{1, L}$:

$$X_k(x,u) = X_{k-1}(x,u) + \Phi(x,u) f[X_{k-1}(x,u),u,t_{k-1}]. \tag{37}$$

As a result of application of formulas, similar to formulas (4)–(17) but with functions depending both on x and u, we can calculate the direct indexes of quality $\sigma(x,u)$, $\zeta(x,u)$, $\lambda(x,u)$, $t_c(x,u)$, $\tau(x,u)$. Unlike linear ACS for the nonlinear systems an integral estimation (19) can be calculated only by numerical integration of the nonlinear system of differential equations (34) together with differential equation of estimation:

$$\partial I(x,u,t)/\partial t = \sum_{k=0}^{l} w_k z_t^{(l-k)}(x,u,t). \tag{38}$$

For the extended system Jacobian matrix (35) and integral of matrix exponent (36) are calculated. The improved IQE $I(x,u,T_f)$ will be a result of integration of such system of differential equations using formula (37) on the span of time $[0,T_f]$ required for the convergence of improper integral.

Optimization Tasks of Nonlinear Control Systems

In the first approaching stability of the nonlinear control system can be defined on a linearized model. For this purpose we differentiate the vector function of equation (34) on input action:

$$B(x,u) = \partial f[X(x,u,t),u,t]/\partial u\Big|_{x=0,\,u=0,\,t=0}. \tag{39}$$

Taking into account matrix (35) let's present the linearized model of the nonlinear system (34):

$$\partial X(x,u,t)/\partial t = A(x,u)X(x,u,t) + B(x,u)u, \qquad y(x,u,t) = C(x,u)X(x,u,t). \tag{40}$$

On this model let's build a transfer function

$$W(x,u,s) = \beta(x,u,s)/\alpha(x,u,s), \quad \alpha(x,u,s) = \sum_{i=0}^{n} \alpha_i(x,u)\, s^{n-i}, \quad \beta(x,u,s) = \sum_{i=0}^{m} \beta_i(x,u) s^{m-i}. \tag{41}$$

On a characteristic polynomial $\alpha(x,u,s)$ let's define the penalty function $P(x,u)$ of kind (28) and elements of the first column of Routh table $\rho_k(x,u)$. Vector objective functions for quality criteria optimization (29) and (30) also will depend both on the vector of the varied parameters x and on input action u. Let's designate these

functions through $F(x,u)$ and introduce n_u input step signals $u_i(t) = u_{si}1(t)$, $i = \overline{1, n_u}$ with magnitudes $u_{si} \in [u_{\min}; u_{\max}]$. Changing the value of input action u_{si} at the fixed value of vector x, we will get different values of vector function $F^{(i)}(x) = F[x, u_{si}1(t)]$ and using comparison operations (32) find the worst value

$$G(x) = \max_i F^{(i)}(x). \tag{42}$$

By analogy with task (31) for linear ACS the task of optimization of the nonlinear systems can be presented as:

$$\min_x G(x). \tag{43}$$

The solution of this task gives the optimal vector of the varied parameters, resulting to the best quality of transient processes for the specified set of input actions.

Modification of Genetic Algorithms for Vector Optimization of Control Systems

For optimization of vector objective functions we offer modifications of genetic algorithms. Initial population from M individuals is generated by introducing a set of random vectors $x^{(j)}$, $j = \overline{1, M}$ with real coordinates in the space of parameters R^p of the control system or vectors of binary values. In the second case it is necessary to represent every binary vector in space R^p and convert them to the vectors $x^{(j)}$, $j = \overline{1, M}$. Usually for this purpose a binary-to-decimal code or Gray code is used. The values of vector objective functions (29), (30) or (42) $F^{(j)} = F(x^j)$, $j = \overline{1, M}$ are calculated for all individuals using systems models equations (1), (34), (35), (39)–(41), quality indexes calculation formulas (2)–(20), (36)–(38), and defining expressions of vector functions (23)–(28). To rank individuals by the degree of fitness it is suggested to use the vector objective function sorting algorithms on the basis of operations of comparison of its values (32)–(33). The fitness level of individuals is subsequently scaled by the inverse square root $1/\sqrt{j}$ of their rank j in the sorted sequence. The scaled fitness level is used in the casual mechanism of selection. Application of genetic operators to the paternal individuals and generation of descendants is made, as well as in scalar genetic algorithms, with the use of different types of crossover, mutation, inversion. For all got descendants the values of vector objective function are calculated, and their ranks are obtained, similarly to the stage of forming the initial population. The new population is formed based on the results of sorting.

Conclusion

The researches results allow to formulate next conclusions.

1. The calculation methods of direct quality indexes and improved integral quadratic estimations have been studied for the linear automatic control systems. These quality indexes are defined only in stability domain of the systems.

2. The optimization tasks of quality indexes of the linear automatic control systems are presented as the tasks of optimization of vector objective functions, taking into account the conditions of stability of the systems, requirements to the quality indexes and priority of system requirements. For modification of optimization methods the set of comparison operations for vector objective functions is introduced.

3. The methods of quality indexes calculation have been also considered for the nonlinear automatic control systems. These quality indexes are the functions of not only varying parameters but also input action of control system.

4. Through dependence of quality indexes on input action of nonlinear control systems for one value of vector of varying parameters the several vector functions values are calculated at different values of input action. By the choice from these vector values the worst value of the vector objective function of nonlinear system is determined.

5. Vector modifications of genetic algorithms for optimization of vector objective functions, allowing to solve the tasks of synthesis for the linear and nonlinear control systems, have been developed.

The efficiency of the proposed application of genetic algorithms for the vector optimization of quality indexes of control systems has been confirmed by computational experiments on the test and applied tasks.

Bibliography

[Besekerskiy, Popov, 2004] V. A. Besekerskiy, E. P. Popov, Teorija system avtomaticheskogo upravlenija, St. Petersburg, Professija, 2004, (in Rus.), 752 p.

[Poljak, Scherbakov, 2005] B. T. Poljak , P. S. Scherbakov. Trudnie zadachi linejnoj theorii upravlenija. Nekotorie podhodi k resheniju // Automatika i telemehanika, 2005, № 5, (in Rus.), pp. 7-46.

[Severin, Nikulina, 2004] V. P.Severin, E. N. Nikulina. Algoritmi vichislenija prjamih pokazatelej kachestva phunkzij vesa sistem avtomaticheskogo upravlenija // Radioelectronica i informatica, 2004, №1, (in Rus.), pp. 52-59.

[Severin, 2004] V. P. Severin. Minimization of Integral Square Estimates of Automatic Control Systems. Part II. Step by Step Approach // Journal of Automation and Information Sciences, 2004, Vol. 36; N 9, pp. 1-9.

[Severin, 2005] V. P. Severin. Vector Optimization of the Integral Quadratic Estimates for Automatic Control Systems // Journal of Computer and Systems Sciences International, 2005, Vol. 44, N 2, pp. 207–216.

[Severin, 2008] V. P. Severin. Parametricheskij synthes sistem automaticheskogo upravlenija methodami vectornoj optimizacii // Tehnicheskaja elektrodinamika, 2008, Vol. 4, (in Rus.), pp. 47-52.

[Severin, 2009] Severin V. P. Struktura laboratorii methodov optimizacii OPTLAB v systeme MATLAB // Trudi IV Vserossijskoj nauchnoj konferencii «Proektirovanie ingenernih i nauchnih prilogenij v srede MATLAB», Astrakhan, Izdatelskij dom «Astrakhanskij universitet», 2009, (in Rus.), pp. 235–267.

[Voronovskiy, Makhotilo, Petrashev Sergeev, 1997] G. K. Voronovskiy, K. V. Makhotilo, S. N. Petrashev, S. A. Sergeev. Kharkov. Geneticheskie algoritmi, iskusstvennie nejronnie seti i problemi virtualnoj realnosti / Basis, 1997, (in Rus.), 112 p.

[Setlak, 2004] G. Setlak. Intelligent Decision Support System, Kiev, LOGOS, 2004, (in Rus.), 252 p.

[Panchenko, 2007] T. V. Panchenko. Geneticheskie algoritmi. Uchebnoe posobie / Pod red. J. J. Tarasevicha, Astrakhan, Izdatelskij dom «Astrakhanskij universitet», 2007, (in Rus.), 88 p.

[Weise 2008] T. Weise. Global Optimization Algorithms – Theory and Application, 2008, 818 p.

Authors' Information

Valery Severin *– Professor, Department of Computer Science and Control, National Technical University "Kharkov Polytechnical Institute", ul. Frunze 21, 31002 Kharkov, Ukraine; e-mail: severinv@mail.ru*

OPTIMAL SYNTHESIS OF INTELLIGENT CONTROL SYSTEMS OF ATOMIC POWER STATION USING GENETIC ALGORITHMS

Seyed Mojtaba Jafari Henjani, Valery Severin

Abstract: *The paper is devoted to the development of a perspective concept of atomic station power block intelligent automatic control systems synthesis on the basis of mathematical models and numeric methods of vector optimization of systems quality indexes using genetic algorithms. The methods for calculation of direct quality indexes and improved integral quadratic estimates have been created. The step-by-step principle of transition to the domain of system stability has been based. There have also been suggested vector objective functions including stability conditions and taking into consideration quality indexes priorities. The reliable genetic algorithms for vector objective functions optimization have been suggested. Mathematical models in the state space for intelligent automatic control systems of nuclear reactor and steam generator have been worked out. The quality indexes optimization of power block intelligent control systems has been carried out, which allowed to estimate various controller types efficiency.*

Keywords*: automatic systems, intelligent control, optimal synthesis, nuclear reactor, steam generator, genetic algorithms.*

ACM Classification Keywords*: G.1.6 Optimization - Nonlinear programming*

Introduction

The experience of research of work of the power units of the atomic electric stations (AES) shows that priority is an increase of efficiency of methods of analysis and synthesis of the automatic control systems (ACS) — ACS of power (ACSP) of the nuclear reactor, ACS of level (ACSL) of water in steam generator and other systems [Ivanov, 1982], [Denisov, Dragunov, 2002], [Nikulina, Severin, 2009].

For the improvement of dynamic properties of ACS of power unit it is suggested to optimize the improved integral quadratic estimations and direct quality indexes (DQI) of the control systems — overshoot, index of vibrations and control time [Severin, Nikulina, 2004], [Severin, 2005], [Severin, 2008], [Jafari Henjani Seyed Mojtaba, Severin, 2009]. The models of reactor WWER-1000 are built on the basis of the systems of differential equations (SDE) of neutrons kinetics and heat sink. The models of steam generator PGV-1000 take into account SDE of material and thermal balances, differential equation (DE) of circulation. Adding to the control objects DE of actuating mechanism and controllers, get the models of ACSP and ACSL [Nikulina, Severin, 2009]. ACS models for optimization of DQI must be simple, as they are repeatedly used by the methods of optimization at the calculation of objective function. The algorithms of calculation of DQI of ACS are considered [Severin, Nikulina, 2004]. The algorithms of vector methods of ACS optimization taking into account the conditions of stability are suggested [Severin, 2008].

For ACS of AES power unit with such control objects as a nuclear reactor and steam generator, the parameters of which change random in the process of functioning, development of synthesis methods is needed. In such control systems in place of standard PID-controllers it is expedient to use intelligent controllers, built on the basis of fuzzy logic and artificial neural networks with the using of genetic algorithms [Goldberg, 1989], [Voronovskiy, Makhotilo, Petrashev, Sergeev, 1997], [Rotshtein, 1999], [Sabanin, Smirnov, Repin, 2003].

The purpose of the article consists of analysis of perspectives of application of genetic algorithms for the synthesis of the intelligent control systems of power unit of nuclear power plant. The general method of synthesis of parameters of controllers of ACS is examined. The parameters of controllers of reactor and steam generator are optimized. The analysis of possibilities of intelligent controllers, built on the basis of fuzzy logic, neural networks and genetic algorithms is investigated.

Methods of Optimization of Controllers Parameters

We will consider the general optimization method of controllers parameters on the example of linear PID controllers, forming the control action u on control object by mistake ε and consisting of proportional (P), integral (I) and differential (D) controllers: P controller is reflected by a proportional law

$$u_P = K_P \varepsilon , \tag{1}$$

I and D controllers are formed the integral and differential control laws:

$$u_I = \frac{1}{T_I}\int \varepsilon dt , \quad u_D = \tau_D \frac{d\varepsilon}{dt} ,$$

where K_p, T_I and τ_D are parameters of controllers. Proportional, integral and realizable differential control laws are answered the transfer functions (TF) of controllers:

$$W_P = K_P , \quad W_I(s) = \frac{1}{T_I s} , \quad W_D(s) = \frac{K_D T_D s}{T_D s + 1} .$$

Last TF at $K_D = 10$ approximately forms a differential control law with $\tau_D = K_D T_D$. With designations

$$\lambda_I = 1/T_I , \quad \lambda_D = 1/T_D , \tag{2}$$

we will write down differential equation of I controller and equation of the realizable D controller:

$$du_I/dt = \lambda_I \varepsilon , \quad dv_D/dt = -\lambda_D (v_D + K_D \varepsilon) , \quad u_D = v_D + K_D \varepsilon . \tag{3}$$

With P, I and D controllers we will build PI, PD, ID and PID controllers. For optimization of parameters of controllers K_P, λ_I and λ_D will form from them the vector of the variable parameters $x \in R^p$ of vector length $p \in \{1,2,3\}$. For P, I and D controllers $p = 1$, for PI, PD and ID controllers $p = 2$, for PID controller $p = 3$.

Let the linear or nonlinear models of the control systems be presented as systems of differential equations:

$$\frac{dX(x,t)}{dt} = A(x)X(x,t) + B(x)U(t) , \quad \frac{dX(x,t)}{dt} = f(x, X(x,t), U(t)) , \quad y(x,t) = C(x)X(x,t) , \tag{4}$$

where $U(t) = U_s 1(t)$ is input action, $y(x,t)$ is an output variable, $X(x,0) = 0$, $y(x,0) = 0$. A constant U_s sets the size of step input action: for a linear model $U_s = 1$, and for a nonlinear model $U_s \in [-1;1]$. If $y(x,\infty) \neq 0$, will define $C(x)$ so, that $y(x,\infty) = 1$.

We will impose the boundary conditions of the variable parameters

$$a_i \le x_i \le b_i , \quad i = \overline{1, p} \tag{5}$$

with the domain of constraint satisfaction

$$G_1 = \{ x \,|\, a_i \le x_i \le b_i , \; i = \overline{1, p} \} \tag{6}$$

and will form a penalty function:

$$S(x) = \sum_{i=1}^{p} \left[\max\{0,\, a_i - x_i\} + \max\{0,\, x_i - b_i\} \right]. \tag{7}$$

At linear ACS (4) for a matrix $A(x)$, and at nonlinear ACS for Jacobian of vector function of right part (4) by the method of D. K. Faddeev will define a characteristic polynomial

$$\alpha(x,s) = \sum_{i=0}^{n} \alpha_i(x)\, s^{n-i},$$

where n is an degree of ACS. For this polynomial on algorithms from [Jafari Henjani Seyed Mojtaba, Severin, 2009] we will calculate the coefficients of Routh-Hurwitz $\rho_k(x)$, $k = \overline{0, n}$, will form domains

$$G_2 = \{x \mid \alpha_i(x) > 0,\, i = \overline{0,n}\}, \quad G_k = \{x \mid \rho_{k-1}(x) > 0\}, \quad k = \overline{3,n} \tag{8}$$

and will define a penalty function:

$$P(x) = \sum_{i=0}^{n} \max\{0, -\alpha_i(x)\}. \tag{9}$$

Integrating the systems of differential equations (4) on an interval $[0, T_f]$ with the number of steps L, will calculate the values of direct indexes of quality — overshoot $\sigma(x)$, scope of vibrations $\zeta(x)$ and time of control $t_c(x)$, proper the preset parameter δ_y of domain of the steady-state value of the output variable $y(x,t)$: at $t > t_c(x)$ $y(x,t) - y(x,\infty) \in [-\delta_y, \delta_y]$ [Severin, Nikulina, 2004]. We will set the upper bounds of overshoot σ_m, scope of vibrations ζ_m and will define domains

$$G_{n+1} = \{x \mid \sigma(x) \le \sigma_m\}, \quad G_{n+2} = \{x \mid \zeta(x) \le \zeta_m\}. \tag{10}$$

On the domains of constraints (6), (8), (10) we will form system of domains of simultaneous satisfaction of constraints $D_0 = R^p$, $D_k = D_{k-1} \cap G_k$, $k = \overline{1, n+2}$ and disjoint domains of levels of constraints: $H_k = D_k \setminus D_{k+1}$, $k = \overline{0, n+1}$, $H_{n+2} = D_{n+2}$. On these domains, functions (7), (9), $\rho_k(x)$, $\sigma(x)$, $\zeta(x)$ and the function of relative value of control time $\tau(x) = t_c(x)/T_f$, which is necessary minimize, will define a vector objective function:

$$F(x) = \begin{cases} (0;\, S(x)), & x \in H_0, \\ (1;\, P(x)), & x \in H_1, \\ (k;\, -\rho_k(x)), & x \in H_k,\, k = \overline{2, n-1}, \\ (n;\, \sigma(x) - \sigma_m), & x \in H_n, \\ (n+1;\, \zeta(x) - \zeta_m), & x \in H_{n+1}, \\ (n+2;\, \tau(x)), & x \in H_{n+2}. \end{cases} \tag{11}$$

This function, taking into account constraints of the variable parameters, necessary and sufficient conditions of stability of ACS and requirements to its DQI, will calculate by algorithms from [Severin, Nikulina, 2004], [Severin, 2008].

The first projection of function (11), representing the number of the satisfied constraints, it is necessary to increase, and the second projection — to diminish. The increasing of the first projection has priority, thus the values of this function $U = (U_1, U_2)$ and $V = (V_1, V_2)$ we will compare by a binary operation «better»:

$$U < V = \begin{cases} 1, & U_1 > V_1 \vee U_1 = V_1 \wedge U_2 < V_2, \\ 0, & U_1 < V_1 \vee U_1 = V_1 \wedge U_2 \geq V_2. \end{cases} \quad (12)$$

Optimization of vector objective function (11) will allow in a united computational process to satisfy constraints of the variable parameters (5), pass to the domain of stability of ACS and optimize DQI in this domain. For optimization of vector function (11) let's modify the methods of unconstrained minimization of scalar functions, replacing the operation of comparison of values of scalar objective functions by the operation of comparison of values of vector objective functions (12). For one-dimensional search we will use introduced by V. F. Korop the method, not requiring the calculation of scopes of interval of uncertainty and consisting in adaptation of step its multiplying by coefficient depending on the results of previous search. For optimization of function (11) of several variables the algorithms of vector methods of Hooke-Jeeves and Nelder-Mead are developed [Severin, 2005].

The Optimal Synthesis of Parameters of Controllers of Nuclear Reactor

We will consider the optimal synthesis of parameters of controllers of nuclear reactor WWER-1000. The ACS of reactor power includes an adder, the power controller (PC), actuating mechanism (AM), model of reactor and negative feedback (NFB) [Nikulina, Severin, 2009].

The input action of ACSP is setting of power U, an output is neutron power ν. A current value ν is measured by ionization chamber and with NFB given on adder, determining an error

$$\varepsilon = U - \nu . \quad (13)$$

An error acts on the input of PC, which forms control action u, given on AM. The AM shifts neutron-absorbing control rod and changes the component of reactivity ρ_d which is passed in linear or nonlinear models of reactor and changes the vector of its state X_R including ν:

$$\frac{dX_R}{dt} = A_R X_R + B_R \rho_d , \quad \frac{dX_R}{dt} = f(X_R, \rho_d), \quad \nu = C_R X_R . \quad (14)$$

Will present an actuating mechanism by DE

$$\frac{d\rho_d}{dt} = a_{dd}\rho_d + b_{du}u . \quad (15)$$

Values of parameters of model of reactor with six groups of delayed neutrons and AM given in [Nikulina, Severin, 2009].

On the models of nuclear reactor, actuating mechanism and controllers we will build the nonlinear and linear models of ACSP of reactor at the different laws of control in a kind (4). So, model of kind (4) with PI controller will build on (1), (3), (13), (15) at $x = (K_P, \lambda_I)$ and $u = u_P + u_I$:

$$X = \begin{pmatrix} X_R & \rho_d & u_I \end{pmatrix}^T, \quad C = \begin{pmatrix} C_R & 0 & 0 \end{pmatrix},$$

$$A(x) = \begin{pmatrix} A_R & B_R & 0 \\ -b_{du}K_P C_R & a_{dd} & b_{du} \\ -\lambda_I C_R & 0 & 0 \end{pmatrix}, \quad B(x) = \begin{pmatrix} 0 \\ b_{du}K_P \\ \lambda_I \end{pmatrix}, \quad f(x, X, U) = \begin{pmatrix} f_R(X_R, \rho_d) \\ a_{dd}\rho_d + b_{du}[K_P(U-\nu) + u_I] \\ \lambda_I(U-\nu) \end{pmatrix}.$$

We will get the models of ACSP with other PC similarly.

For optimization of DQI of ACSP of reactor will impose the values of boundary conditions of the variable parameters (5) with $a_i = 0$ and $b_i = 100$, $i = \overline{1, p}$. We will express through x the linear and nonlinear models

of ACSP (4). The degree of model of ACSP with P controller is $n=11$, with I, PI and PD controllers is $n=12$, with ID and PID controllers is $n=13$. For obtaining of the transient processes of power without overshoot and oscillation with the minimum time of control in the linear models of ACSP of reactor at input action $U=1(t)$ we will set the values of parameters of task of optimization of direct indexes: acceptable values of overshoot and scope of vibrations $\sigma_m=0$ and $\zeta_m=0$, parameter of domain of the steady-state value $\delta_y=0.05$. For I and ID controllers will set time of integration T_f 1000 s and for the other controllers — 100 s, number of steps of integration $L=200$. We will form a vector objective function (11) and will optimize it in the case of one variable by the vector method of step adaptation, and in the case of several variables — by the vector method of Nelder-Mead [Severin, 2005].

In table 1 for different PC the optimal values of parameters of PC are presented $K_P^*, \lambda_I^*, \lambda_D^*$, and also the proper by it's the values of projections of function (11) F_1^*, F_2^* and value of control time t_c^*. Values F_1^* show that all constraints of the task of optimization of the direct indexes of quality are executed in optimal points. At the optimal values of parameters of P and PD controllers a static error excels 10 %. The systems with the I and ID controllers have large value of control time. Efficiency of PI and PID controllers is identical, they allow to provide the fast response of ACSP. From the optimal values λ_I^* and λ_D^* it is possible to pass to the values of time constants on formulas (2). Data of table 1 show that for PI controller an optimal process on a fast-acting is provided on the high bound of parameter K_P.

Table 1. Results of optimization of parameters of PC

PC	K_P^*	λ_I^*	λ_D^*	F_1^*	F_2^*	t_c^*, s
P	45.8	—	—	13	0.232	23.2
I	—	0.083	—	14	0.343	343.3
PI	100	2.59	—	14	0.138	13.8
PD	100	—	0,044	14	0.114	11.4
ID	—	0.083	100	15	0.343	343.3
PID	100	2.59	100	15	0.138	13.8

For optimization of nonlinear models of ACSP with PI controller at the different values of setting of power U_s will put $y(x,t)=\nu(x,t)/U_s$. In table 2 at $b_i=25$, $T_f=200$ s, $K_P^*=25{,}0$ for different values U_s optimal values λ_I^* and control time t_c^* are given. With diminishing U_s λ_I^* diminishes and t_c^* increases.

Table 2. Results of optimization of nonlinear ACSP

U_s	λ_I^*	t_c^*, s	ν_s	λ_I^*	t_c^*, s
0.10	0.673	53.4	–0.20	0.659	54.7
0.05	0.671	53.6	–0.50	0.631	58.1
–0.05	0.667	53.9	–0.75	0.573	67.9
–0.10	0.664	54.2	–0.90	0.494	88.8

Thus, nonlinearity of mathematical model of nuclear reactor substantially influences on the optimal values of parameters of controller.

The Optimal Synthesis of Parameters of Controllers of Water Level in Steam Generator

The automatic control system of water level in steam generator (SG) PGW-1000 includes the level controller (LC), model of steam generator, NFB and adder [Nikulina, Severin, 2009]. On the input of the system control of

level a setting of level $\xi_{cs} = 0$ is given, an output is a coordinate of level ξ_c, which together with the increases of discharges of steam and water g_s and g_w form an error:

$$\varepsilon = \xi_{cs} - \xi_c + g_s - g_w .$$

We will express g_s and g_w through the vector of the state of SG X_G:

$$\varepsilon = -D_G X_G . \tag{16}$$

An error acts on the input of LC, forming control action u, given in the model of SG:

$$\frac{dX_G}{dt} = A_G X_G + B_{Gw} u + B_{Gs} U , \quad \xi_c = C_G X_G , \tag{17}$$

where U is disturbing influence of valve of control of turbine. We will build the models of the level control systems with different controllers in a kind (4). The model of ACSL with PI controller will get at $x = (K_P, \lambda_I)$ and $u = u_P + u_I$ on (1), (3), (16), (17):

$$X = \begin{pmatrix} X_G \\ u_I \end{pmatrix}, \quad A(x) = \begin{pmatrix} A_G - B_{Gw} K_P D_G & B_{Gw} \\ -\lambda_I D_G & 0 \end{pmatrix}, \quad B = \begin{pmatrix} B_{Gs} & 0 \end{pmatrix}^T, \quad C = \begin{pmatrix} C_G & 0 \end{pmatrix}.$$

We will impose on the values of the variable parameters constraints (5) with $a_i = 0$, $b_i = 100$, $i = \overline{1, p}$. We will express through x the linear models of the control systems of level of kind (4). The degree of model of the control system with P controller is $n = 9$, with I, D, PI and PD controllers is $n = 10$, with ID and PID controllers is $n = 11$. For obtaining of optimal transient processes without vibrations with minimum time of control will set the values of parameters of task of optimization of DQI: proper the possible increase of level $h_c = 15\,\mathrm{sm}$ value of maximal deviation of coordinate of level $\sigma_m = 1$, proper processes without vibrations legitimate value of index of vibrations $\zeta_m = 0$, time of integration $T_f = 500\,\mathrm{s}$, number of steps of integration $L = 200$, parameter of the domain of control time $\delta_y = 0.05$ at steady-state value of process $y(\infty) = 0$. We will form a vector function (11) which optimize by the vector method of step adaptation at $p = 1$ or by method Nelder-Mead at $p > 1$.

The results of optimal synthesis of parameters of controllers show that for ACSL with all considered types of controllers optimal processes are got without vibrations, inherent a process in real ACSL. The ASCL with P, D and PD controllers are static, and with the other ASCL — astatic. The best type of controller of level is PI controller with the optimal values of parameters, providing the most rapid transient process without vibrations.

Prospects of the Use of Fuzzy Controllers

Practical introduction of fuzzy controllers (FC) in the control systems of industrial sphere is intensively investigated. The investigated review on fuzzy controllers exposed the following:

— application of FC allows to use for control of technological processes information of qualitative type, which it is impossible to formalize during realization of traditional laws of control; an fuzzy controller is showed not sensitive to disturbances and demonstrates the best characteristics as compared to classic controllers;

— for composition of control rules of fuzzy controller intuition of developer and good knowledge of control object is required, methods for the direct synthesis of fuzzy controllers are practically absent;

— existent FR tune in to logic of user through the change of membership functions, thus a choice of membership functions is nontrivial procedure;

— there are not standardized recommendations on the choice of method of interpretation of fuzzy conclusion;

— possibility of application of fuzzy controllers is not certain for a multidimensional process.

Effective combination of methods of control theory and fuzzy logic theory allows to form the models of difficultly formalizable processes of control. The fuzzy systems are especially effective in the complex nonlinear processes of AES with parametric uncertainty.

Prospects of the Use of Neural Networks

The creation of artificial neural networks (NN) is based on development of principle new algorithms and methods of control for nonlinear dynamic objects. Most charts of neuron control are based on next approaches.

1. Successive chart of control. A neuron network will realize a reverse reflection in relation to a reflection «input-output» for the control object. Thus, if to set the supporting signal on NN, then the output signal of control object will be adopted by the same value. Teaching of neuron network is founded by back-propagation algorithm.

2. Parallel charts of control. NN of parallel type is used for tuning of input control signal which is the output signal of usual PID-controller. Tuning is executed so that an output signal of control object as possible more precisely corresponds the set supporting signal of ACS.

3. Control chart with self-tuning. A neural network is used for tuning the parameters of ordinary controller like tuning, to executable by a man-operator.

4. Control chart with an emulator and comptroller. Neural comptroller is taught on the inversion model of control object, and neural emulator on the ordinary model of object. Neural comptroller can be taught directly on the basis of back-propagation of error through neural emulator.

For the estimation of efficiency of neural comptrollers it is necessary to carry out computing experiments with the use of the developed models of objects of control of power unit and numerical methods of optimization.

Prospects of Application of Genetic Algorithms for the Systems Synthesis

For complex multiloop control systems and control systems with neural controllers it is characteristically large number of local extremes. At the synthesis of intelligent controller the surface of response is unknown, and the number of parameters which are necessary to be defined is large. According to ideology of work of genetic algorithm (GA) the form of surface of response does not matter for its successful work. The task of tuning of the parameters of intelligent controller, as a rule, is the multiextremal task of optimization. During optimization of the complex multiloop and multivariable control systems and intelligent systems with neural controllers genetic algorithms with high probability find global extreme. However the calculation of objective function on a time domain of transient process requires considerable computational resources that substantially influences on common work time of GA. It is expedient to develop the genetic algorithms in this direction.

Conclusion

The results of the investigated researches allow to present next conclusions.

1. The method of synthesis of parameters of controllers is considered on the base of optimization of vector objective function taking into account conditions of stability and direct indexes ACS.

2. The parameters of controllers of power of nuclear reactor WWER-1000 are optimized, that showed most efficiency of PI controller.

3. Optimization of parameters of controllers of water level of steam generator PGW-1000 showed most efficiency of PI controller with the optimal values of parameters.

4. The analysis of possibilities of intelligent controllers, built on the basis of fuzzy logic and neural networks, shows the wide prospects of their use in the automatic control systems of power units of nuclear power plants.

5. Application for the synthesis of the intelligent control systems of power unit AES the genetic algorithms will allow to raise quality of the systems.

Bibliography

[Ivanov, 1982] V. A. Ivanov. Regulirovanie energoblokov, Leningrad, Mashinostroenie, 1982, 311 p.

[Denisov, Dragunov, 2002] V. P. Denisov, J. G. Dragunov. Reactornie ustanovki WWER dlja atomnih electrostancij, Moscow, IzdAT, 2002. 480 p.

[Nikulina, Severin, 2009] E. N. Nikulina, V. P. Severin. Mnogocriterialnij sintez system upravlenija reactornoj ustanovki putem minimizacii integralnih kvadraticnih ocenok // Jadernaja i radiacionnaja bezopasnost, 2009, Vol 12, N 2, pp. 3–12.

[Severin, Nikulina, 2004] V. P. Severin, E. N. Nikulina. Algorithmi vichislenija prjamih pokazateley kachestva funkcij vesa system avtomaticheskogo upravlenija // Radioelectronica i informatika, 2004, N 1, pp. 52–59.

[Severin, 2005] V. P. Severin. Vector Optimization of the Integral Quadratic Estimates for Automatic Control Systems // Journal of Computer and Systems Sciences International, 2005, Vol. 44, N 2, pp. 207–216.

[Severin, 2008] V. P. Severin. Parametricheskij sintez system avtomaticheskogo upravlenija metodami vektornoj optimizacii // Tehnicheskaja elektrodinamika, 2008, N 4, pp. 47-52.

[Jafari Henjani Seyed Mojtaba, Severin, 2009] Jafari Henjani Seyed Mojtaba, V. P. Severin. Mnogokriterialnyj parametricheskij sintez system avtomaticheskogo upravlenija minimizaciej integralnyh kvadratichnyh ocenoc v sfere MATLAB // Astrakhan, Izdatelskij dom «Astrakhanskij universitet», 2009, (in Rus.), pp. 444–456.

[Goldberg, 1989] D. E. Goldberg. Genetic Algorithms in Search Optimizations and Machine Learning. Addison.Wesly, 1989.

[Voronovskiy, Makhotilo, Petrashev, Sergeev, 1997] G. K. Voronovskiy, K. V Makhotilo, S. N. Petrashev, S. A. Sergeev. Geneticheskie algoritmi, iskusstvennie nejronnie seti i problemi virtualnoj realnosti, Kharkov, Basis, 1997, (in Rus.), 112 p.

[Rotshtein, 1999] A. P. Rotshtein. Intelektualnye tehnologii identifikaccii: nechetkie mnogestva, geneticheskij algoritm, nejronnye seti, Vinnica: «UNIVERSUM-Vinnica», 1999, (in Rus.), 300 p.

[Sabanin, Smirnov, Repin, 2003] V. R. Sabanin, N. I. Smirnov, A. I. Repin. Optimizacija nastroennih parametrov regyliryuwih ystrojstv v ASR // Sbornic trudov konferencii Control 2003, MEI, 2003, pp. 144–148.

[Gostev, 2005] V. I. Gostev. Sintez nechetkih regyljatogov system avtomaticheskogo upravlenija, Kiev, Izdatelstvo Radioamator, 2005, 708 p.

Authors' Information

Seyed Mojtaba Jafari Henjani *– Ph.D Student, Department of Computer Science and Control, National Technical University "Kharkov Polytechnical Institute", ul. Frunze 21, 31002 Kharkov, Ukraine; e-mail: smjh79@gmail.com*

Valery Severin *– Professor, Department of Computer Science and Control, National Technical University "Kharkov Polytechnical Institute", ul. Frunze 21, 31002 Kharkov, Ukraine; e-mail: severinv@mail.ru*

OPTIMAL DESIGN OF INTELLIGENT CONTROL SYSTEMS OF STEAM TURBINE USING GENETIC ALGORITHMS

Khrystyna Fedyanyna, Ievgeniia Kucher, Valery Severin

Abstract: *One of the basic engineering problems of the automatic control systems synthesis for steam turbines is systems quality indices optimization task. The features of such task are defined by plenty of control systems structural parameters, complication of quality indices formalization and calculation, the systems models high order. The greatest difficulty of the control system synthesis is optimization models and methods design. The paper purpose is to develop models and methods for optimum design of the intelligent automatic control systems of atomic power station steam turbine using genetic algorithms. The steam turbine automatic control system is applied to stabilize turbine rotor frequency with high precision. The intelligent steam turbine control system includes a steam turbine, frequency sensor, intelligent frequency regulator, electro-hydraulic automatic drive and turbine adjusting valve. Input signals on an automatic drive can be given by the steam pressure regulator through the turbine management mechanism or by the electric power regulator. Assumptions substantiated to model the automatic control system and automatic drive as executive link of control system. The automatic drive diagram of principle includes electro-hydraulic transformer, sleeve valve, servomotor, position sensors and electronic part. In the paper were built the mathematical models of automatic drive and steam turbine, the models permanent parameters values were calculated, the mathematical models of the automatic control system were developed in state space with the intelligent frequency regulator, the regulator parameters were optimized with system quality indices using genetic algorithms.*

Keywords*: automatic control system, automatic drive, steam turbine, intelligent regulator, genetic algorithms.*

ACM Classification Keywords*: G.1.6 Optimization - Nonlinear programming*

Introduction

One of the basic engineering problems of creation of the steam turbines automatic control systems is a problem of optimization of their quality indices [Fragin, 2005]. Used engineering methods of calculation of the control systems are usually based on models substantial simplification and application of close indirect scalar quality criteria. The greatest lack of calculations of the control systems of nuclear power station turbines is slight using of models with optimal intelligent regulators in this area. Intelligent regulators based on fuzzy logic and artificial neuron networks allow to solve more intricate problems of compound objects control than standard regulators.

During last years works on practical introduction of fuzzy regulators, fuzzy expert and control systems in industrial and nonproductive spheres are intensively conducted [Terano, Asai, Sugeno, 1993]. Moreover fuzzy regulators excel traditional PID-regulators in transitional processes quality and in achieving management aims.

The use of intelligent regulators results in the nonlinear non-obviously set models of the automatic control systems. For the optimal synthesis of such systems parameters taking into account multiextremeness of their quality indices it is expedient to use the global optimization methods, in particular, methods of direct search and genetic algorithms [Nelder, Mead, 1964], [Sabanin, Smirnov, Repin, 2003], [Goldberg, 1989], [Voronovskiy, Makhotilo, Petrashev, Sergeev, 1997]. It is necessary to modify these methods, usually used for optimization of scalar objective functions, for optimization of vector objective functions taking into account all requirements to automatic control system [Kirillov, 1988], [Troyanovskiy, 1983], [Severin, 2007].

The purpose of this article is to develop models of the rotation frequency stabilizing systems of steam turbine rotor and its watching drive in state space with the using of standard and fuzzy regulators, and also to analyses optimization possibilities by genetic algorithms of the frequency stabilizing systems quality indices.

This purpose achievement will allow to do the steam turbines automatic control systems more effective taking into account all structural and technological requirements produced to them.

Possibilities of Genetic Algorithms and their Modification to Optimal by Synthesize Control Systems

Previous research experience of the automatic control systems with linear regulators has showed that the criteria of optimization, as a rule, were the same one-extreme. However, for the complex multicontour control systems and control systems with intelligent regulators presence of the large number of local extremes along with global ones is significant. Local extremes appear also in imposing restrictions on the area of search. For the solution of the same one-extreme task of optimization, there is a sufficient number of algorithms of gradient methods and methods of direct search. One of the most effective methods of direct search is the Nelder-Mead's method of the deformed polyhedron [Nelder, Mead, 1964].

Application of Nelder-Mead's method for the optimization of the automatic control systems with the fuzzy regulators of different structure does not result in an optimum decision. In every case results depend on the chosen coordinates of initial point of search. A conclusion result from this that objective functions of similar tasks are multiextremeness, and for their solution the global optimization methods are required [Sabanin, Smirnov, Repin, 2003].

The present time the most preferable methods of multiextreme optimization are genetic algorithms realizing the postulates of evolutionism and experience of selection of plants and animals [Goldberg, 1989]. The purpose of optimization of the control systems by genetic algorithms consists in finding the best possible decision of optimization task by one or few criteria. To realize a genetic algorithm it is needed at first to choose a suitable structure for these decisions presentation. In the search problem statement copy of this data structure presents a point in space of search of all possible decisions. To optimize a structure using a genetic algorithm it is needed to set some measure of quality for every structure in search space. The function of fitness is used for this purpose. During maximization the objective function turns itself into a fitness function. For the tasks of minimization it is necessary to invert an objective function and displace it after that to the area of positive values. Strategy of search of optimum solution in genetic algorithms relies on the hypothesis of selection: the more fitness of an individual is higher the higher is the probability that descendants, got with its participation, will have determining fitness features stronger expressed [Voronovskiy, Makhotilo, Petrashev, Sergeev, 1997].

If to accept that every individual of population is a point in coordinate space of optimization task $X_i[x_{1i}, x_{2i}, ..., x_{li}]$ and fitness of an individual is the proper value of objective function $f(X_i)$ then individual populations can be examined as a great number of coordinate points in space, and process of evolution — as motion of these points toward the improvement of values of objective function. A feature of the genetic algorithm as a global search method is that none of genetic operators in the descendants generation process leans on knowledge of local relief response surface for the objective function. Forming of descendants by genetic operators occurs in a random manner, and that is why there is no guarantee that found solutions will be better than paternal. The enumerated features restrain wide application of genetic algorithms in engineering practice. However necessity in such algorithms for the solution of the applied tasks of comparatively small dimension constantly grows, especially in connection with the planned tendency of introduction of fuzzy and neural network technologies in the control systems. For the optimum synthesis of the automatic control systems it is expedient to use vector objective functions taking into account all complex of system requirements in order of their preference [Severin, 2007]. As genetic algorithms are mainly used for optimization of scalar objective functions for optimization of vector objective functions these algorithms are modified with the using of operations of vector function values comparison [Severin, 2007].

Design and Optimization of Watching Drive

An electro-hydraulic watching drive is the main executive link of the automatic control system of frequency of rotation of steam turbines rotor. We will develop the mathematical model of watching drive intended for optimization of its quality indices.

The chart diagram of principle of electro-hydraulic watching drive on a fig. 1 includes an electro-hydraulic transformer (EHT), chopping off slide-valve (CHV), servomotor (SM), sensors of position (SP), electronic part (EP).

On the models of transformer, chopping off slide-valve, servomotor and sensor of position models of charts of drive with the different number of sensors are got. The flow diagram of watching drive with three sensors is presented on a fig. 2. The diagram includes summators, proportional link with a coefficient K, model of successive connection of electro-hydraulic transformer, chopping off slide-valve and servomotor, sensor of position of slide-valve of electro-hydraulic transformer, sensor of position of chopping off slide-valve, sensor of position of servomotor, three negative feed-backs according to the position of slide-valve of electro-hydraulic transformer, chopping off slide-valve and servomotor with coefficients k_y, k_s and 1. Signal

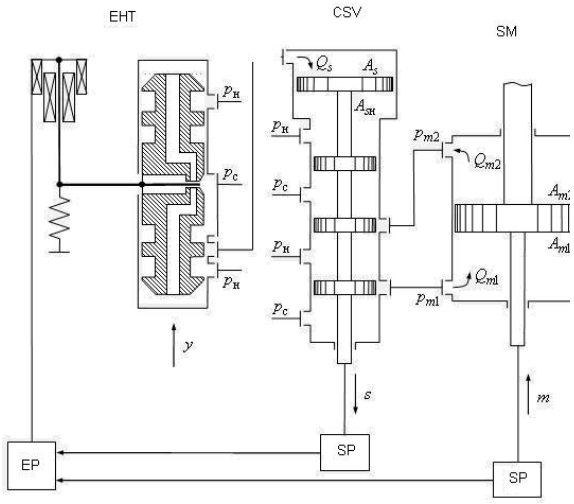


Figure 1. The chart of principle of watching drive

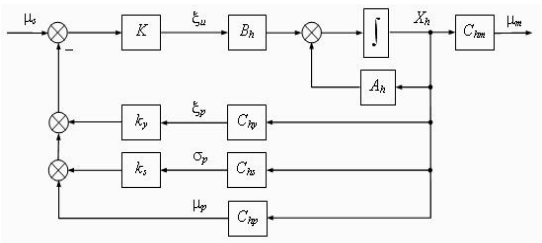


Figure 2. The flow diagram of watching drive

$\xi_u = K(\mu_s - k_y\xi_p - k_s\sigma_p - \mu_p)$ comes to the entrance of electro-hydraulic transformer.

For the reliability of the automatic control system of rotation frequency of turbine rotor a watching drive must keep a considerable reserve of stability, be sensible to control signals and fast-acting. We will consider watching drive quality indices optimization by the modified genetic algorithms.

For optimization of drive we form a vector from its variable parameters $x = (K, k_y, k_s)$. We impose limitations $a_i \le x_i \le b_i$, $a_i = 0$, $b_i = 100$, $i = \overline{1, p}$ on the values of the varied parameters and form a penalty function for them: $S(x) = \sum_{i=1}^{p} [\max\{0, a_i - x_i\} + \max\{0, x_i - b_i\}]$. On the penalty function of limitations $S(x)$, penalty function of necessary condition of stability $P(x)$, coefficients of the first column of Routh table $\rho_k(x)$, $k = \overline{2, n-1}$, overshoot $\sigma(x)$, vibrations scope $\zeta(x)$, maximum values of overshoot and vibrations scope σ_m and ζ_m, control time $\tau(x)$ and domains of levels H_k, $k = \overline{0, n+2}$ we form the vector objective function:

$$F(x)=\begin{cases}(0;S(x)), & x\in H_0,\\ (1;P(x)), & x\in H_1,\\ (k;-\rho_k(x)), & x\in H_k, k=\overline{2,n-1},\\ (n;\sigma(x)-\sigma_m), & x\in H_n,\\ (n+1;\zeta(x)-\zeta_m), & x\in H_{n+1},\\ (n+2;\tau(x)), & x\in H_{n+2}.\end{cases}$$

This function takes into account all limitations of the system parameters. The first its projection — the function of level corresponds to the number of executed limitations, and the second projection the function of penalty determines the penalty of limitation violation [Severin, 2007].

The vector objective function in a priority order takes into account direct limitations on the system parameters, necessary and sufficient conditions of stability, limitations of direct quality indices, requirement of time minimum for the adjustment of the control system.

For its optimization the genetic algorithms were modified with the use of operations of the vector function values comparison [Severin, 2007].

The results of drive optimization with the applied quality criteria — improved integral quadratic estimation and direct quality indices – are presented in a table 1. For the number of calculating experiments N optimum values of the varied parameters K, k_y^*, k_s^* and adjustment time t^* founded by the modified genetic algorithm are presented. The graphs of the proper transitional processes are shown on a fig. 3.

Table 1. Results of optimization of watching drive

N	K	k_y^*	k_s^*	t^*, s
1	2.6	0.45	0.75	0.080
2	1.8	0.60	0.90	0.112
3	0.7	0.68	1.26	0.161
4	2.4	0.44	0.77	0.082
5	3.0	0.46	0.69	0.076
6	1.5	0.57	0.98	0.118
7	2.0	0.61	0.99	0.147
8	2.5	0.63	1.02	0.174

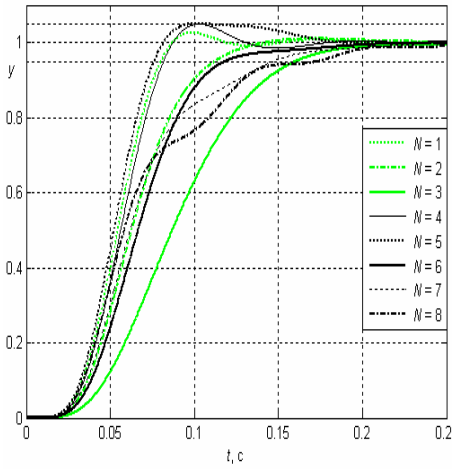


Figure 3. Optimum processes in a watching drive

These results, got by genetic algorithms, confirmed results that were got before by other optimization methods.

On a fig. 4 for $N=6$ the graphs of change of state variables of optimum watching drive are presented. Graphs turned out from the entrance step influence $\mu_s=1(t)$.

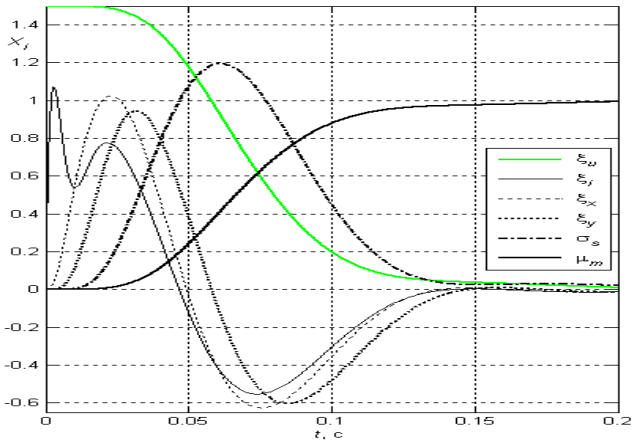


Figure 4. Change of state variables an optimal drive

All processes are smooth and quickly set without substantial vibrations. Coordinates of current ξ_i, displacement of managing spool ξ_x and slide-valve of electro-hydraulic transformer ξ_y from zero initial values increase at first, then diminished, reverse sign and aspire to zero eventual values. Coordinates of voltage ξ_u, chopping off slide-valve σ_S and servomotor μ_m are positive, ξ_u and σ_S aspire to zero eventual values, and μ_m aims at 1. Thus, watching drive quality indices optimization allowed to provide high quality of processes which flow in it.

Design and Optimization of the Control Systems of the Steam Turbine

We will consider the construction of mathematical models of the automatic control systems of the steam turbine K-1000-60/1500 in state space. The chart of principle of steam turbine is presented on a fig. 5 and includes the turbine adjusting valve (TAV), high pressure cylinder (HPC), volume before the high pressure cylinder, volume in a separator-superheater, separator-superheater valve (SSV), volume after the separator-superheater valve, middle pressure cylinder (MPC), low pressure cylinder (LPC), volume before the low pressure cylinder.

On the basis of the system of differential equations of watching drive, steam highway and rotor of steam turbine the mathematical model of turbine is got as a control object [Severin, 2007].

The automatic control system of frequency of steam turbine is intended for stabilization of the rotation frequency of steam turbine rotor. The heaviest test of the automatic control system of frequency of rotation is dropping of the nominal load of steam-turbine and passing to the no-load conditions. On developed model of the steam turbine K-1000-60/1500 we build models of the automatic control systems of rotation frequency of steam turbine rotor for the load dropping at the different laws of control with different regulators.

The flow diagram of the offered model of the real system of automatic control of steam-turbine rotation of frequency is presented on a fig. 6 and includes summarizing, regulator of frequency, steam turbine model and negative feed-back [Kirillov, 1988]. The signal of error is given on the entrance of standard regulator of frequency with a transmission function $W_C(t)$ and forms managing influence u. This influence and revolting signal ν_S come in model of steam turbine and change the vector of his state X_T. The coordinate of frequency φ is given by negative feedback on summarizing and with master influence φ_s forms the signal of error ε [Troyanovskiy, 1983]. The entrance perturbation action of the automatic control system of frequency is a signal of power change ν_S.

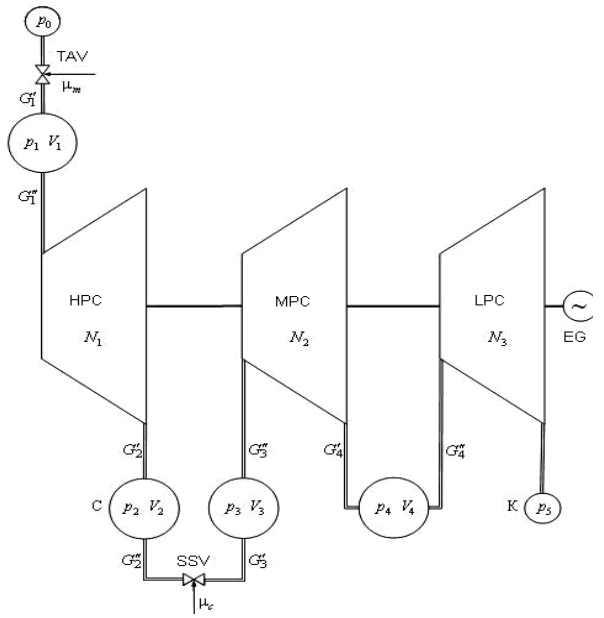


Figure 5. The chart of principle of steam turbine K-1000-60/1500

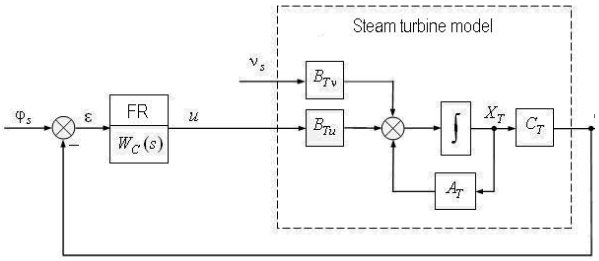


Figure 6. Flow diagram of the automatic control system of frequency of rotation

The output variable is frequency φ of steam turbine. On the basis of steam turbine model the mathematical models of its control systems are built with different regulators are built [Severin, 2007].

For optimization of parameters of the automatic control systems of rotation frequency we form a vector x from the varied parameters of PID-regulator of frequency K_P, λ_i and λ_D. We define the models of automatic control systems of rotation frequency at entrance perturbation action $\nu_s = -1(t)$ proper to the transition of steam turbine from the mode of nominal power $\nu = 0$ to the no-load conditions and to the output coordinate y of deviation of frequency in percents:

$$\frac{dX_c(x,t)}{dt} = f(x, X_c(x,t), \nu_s), \qquad y(x,t) = 100 \cdot C_c X_c(x,t),$$

where index C corresponds to the type of regulator P, I, D, PI, PD, ID or PID. Automatic control systems of rotation frequency with I, D and ID-regulators of frequency appeared uneffective as their optimum processes have frequency deviations which exceed 12 %. The results of optimization of vector objective function by the modified genetic algorithms are presented in table 2 and on fig. 7.

For the static automatic control systems of rotation frequency with P and PD-regulators $y(\infty) = 1.0$, and for the astatic automatic control systems of rotation frequency with PI and PID-regulators $y(\infty) = 0$. In the first four experiments the legitimate scope value of frequency vibrations is accepted $\zeta_m = 1$, and in the last two $\zeta_m = 0.5$. For $N = 5$ time of integration is set to $T_f = 50$, and in the other experiments to $T_f = 20$.

On fig. 8 the changes of some state variables of the optimum automatic control system of rotations frequency of steam turbine with a PID-regulator got at entrance step influence $\nu_S = -1(t)$ are presented. The coordinate of chopping off slide-valve $\sigma_s(t)$ changes in small limits and set in a zero value. The servomotor coordinate $\mu_m(t)$

diminishing takes on considerable negative values and passes to the value -1. The coordinate of power of turbine $\nu(t)$ from an initial zero value of nominal power after a few quickly attenuated oscillations aspires to the value -1 proper steam turbine idling. The frequency deviation coordinate $y(t)$ from an initial zero value after attenuated oscillations returns to zero nominal level. The signal of PID-regulator $u(t)$ diminishes and then after vibrations sets on a value -1. The electro-hydraulic watching drive output coordinate $\mu_m(t)$ repeats the signal change of frequency regulator $u(t)$ with a small delay.

Table 2. Results of optimization of the automatic control system of steam-turbine frequency

N	λ_I^*	λ_D^*	F_1^*	F_2^*	σ^*	ζ^*	t_c^*, s
1	—	—	20	2.0	3.6	3.0	3.56
2	27	—	22	0.3	3.5	1.0	5.19
3	—	1.1	21	1.7	3.5	2.7	3.16
4	31	1.1	23	0.2	3.4	0.8	3.50
5	10	—	22	0.4	3.6	0.5	19.42
6	23	1.1	23	0.3	3.4	0.5	5.04

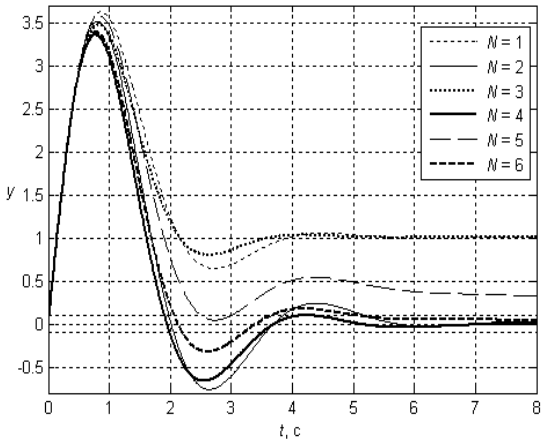


Figure 7 Change of frequency at different regulators

Thus, in this automatic control system of rotation frequency of steam turbine with a PID-regulator an electro-hydraulic watching drive and automatic control system of rotation frequency satisfy the requirements produced to them. Unlike the considered regulators a fuzzy PID-regulator allows to organize more flexible adjusting in PID control law with the automatic calculation of regulator adjustments for the objects with a proportional executive mechanism. Principle of action of fuzzy regulators differs from classic regulators by the fact that on the object entrance along with the regulator signal the additional trial sine wave signal of small amplitude is given. The calculation of regulator adjustments is carried out on amplitude and phase of harmonic constituent in the object output signal [Terano, Asai, Sugeno, 1993].

The fuzzy regulators use will allow to automatize better the process of contours adjusting and also to give up the use of object dynamics authentication ordinary methods and adjustment regulators optimum parameters calculation. As practice shows adaptive regulators allow to save up to 15% of raw material and energy resources as compared with a hand management or about 5% as compared with the not optimally adjusted classic PID-regulator. Moreover application of the adaptive adjustment leads to reduction of terms and costs of starting-ups and adjustment works.

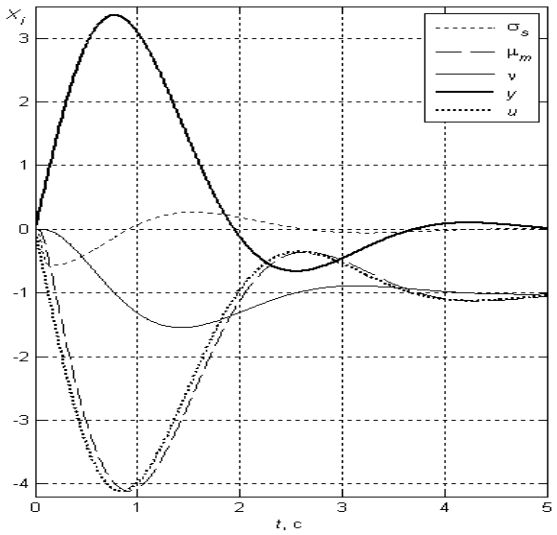


Figure 8. A state variables change in the optimal automatic control system of frequency

Conclusion

The results of the conducted researches allow to formulate following conclusions.

1. On the basis of models of electro-hydraulic transformer, chopping off slide-valve, servomotor and sensors of position a mathematical model in state space of steam turbine watching drive with three sensors is developed. Parameters optimization of this model provides high quality of monotonous process in a watching drive at the optimum value of amplification factor.

2. The mathematical models of the stabilizing systems of rotation frequency of steam turbine rotor are developed in state space with different standard and fuzzy regulators of frequency. The direct quality indices optimization tasks for the rotation frequency stabilizing systems are solved with the use of the modified genetic algorithms. The most effective standard type of frequency regulator at the turbine load dropping is a PID-regulator which provides the most rapid transitional process with the least deviation of frequency. Possibilities of fuzzy PID-regulators are analysed.

3. For optimization of vector objective functions of the automatic control systems the direct methods of unconstrained minimization and genetic algorithms which allowed to find the optimum values of indices are modified. Calculable experiments confirmed high efficiency of application of the modified methods.

Bibliography

[Fragin, 2005] M. S. Fragin. Regulirovanie and maslosnabzhenie of steam-turbines: nastoyaschee and nearest prospects. St. Petersburg, Energotekh, 2005, (in Rus.), 248 p.

[Terano, Asai, Sugeno, 1993] T. Terano, K. Asai , M. Sugeno Applied fuzzy systems, Moscow, Mir, 1993, (in Rus.), 212 p.

[Nelder, Mead, 1964] J. A. Nelder, R. A. Mead. A Simplex Method For Function Minimization // Computer Journal, N 7, 1964, pp. 308-313.

[Sabanin, Smirnov, Repin, 2003] Sabanin V. R., Smirnov N. I., Repin A. I. Optimization of settings of regulative devices in ASR // Collection of labours of conference of Control 2003, MEI, 2003. pp. 144-148.

[Goldberg, 1989] D. E.Goldberg. Genetic Algorithms in Search Optimizations and Machine Learning., Addison, Wesly, 1989.

[Voronovskiy, Makhotilo, Petrashev, Sergeev, 1997] G. K. Voronovskiy, K. V. Makhotilo, S. N. Petrashev, S. A. Sergeev. Genetic algorithms, artificial neuron networks and problems of virtual reality, Kharkov, Basis, (in Rus.), 1997.

[Kirillov, 1988] I. I. Kirillov. Avtomatic adjusting of steam-turbines and gas-turbine options, Leningrad, Engineer, 1988, (in Rus.), 447 p.

[Troyanovskiy, 1983] B. M. Troyanovskiy. Steam turbines. Successes, unsolved problems // Teploenergetika, 1983, N 1, pp. 6-11.

[Severin, 2007] V. P. Severin. Modeli i metodi optimizacii pokaznikiv yakosti system automatic upravlinnya energobloku atomnoї elektrostanciї: Avtoref. dis... d-ra tekhn. sciences: 05.13.07 / NTU «KPI», Kharkiv, 2007, (in Ukr.), 36 p.

Authors' Information

Khrystyna Fedyanyna *– Student, Department of Computer Science and Control, National Technical University "Kharkov Polytechnical Institute", ul. Frunze 21, 61002 Kharkov, Ukraine; e-mail: fedyanyna@gmail.com*

Ievgeniia Kucher *– Student, Department of Computer Science and Control, National Technical University "Kharkov Polytechnical Institute", ul. Frunze 21, 61002 Kharkov, Ukraine; e-mail: kucher_ievgeniia@ukr.net*

Valery Severin *– Professor, Department of Computer Science and Control, National Technical University "Kharkov Polytechnical Institute", ul. Frunze 21, 31002 Kharkov, Ukraine; e-mail: severinv@mail.ru*

Intelligent Agents and Multi-Agent Systems

AGENT ORIENTATION AS A TOOLBOX FOR ORGANIZATIONAL MODELING AND PERFORMANCE IMPROVEMENT

Jacek Jakieła, Bartosz Pomianek

Abstract: *The main goal of the paper is to convince that agent orientation may be considered as a powerful paradigm for organization modeling and the reference architecture for Management Information Systems, that if properly applied, would lead to firm's overall performance improvement. The scenario of proving these theses consists of three steps. Firstly the basic concepts of agency have been defined. Then the agent and multi-agent system are presented as playing the roles of very natural and intuitive modeling constructs and complexity management tools. Finally the benefits of agent oriented software application are described from the perspective of gaining competitive advantage by improving intra- and interorganizational efficiency due to basic characteristics of agents such as constant monitoring of the environment (where they are situated), reactivity, pro-activeness and social ability.*

Keywords*: Software Agent, Intelligent Agent, Agent Oriented Approach, Organization Modeling, Multi-agent System, Complexity Management, Business Performance Improvement, Supply Chain, Demand Chain, Management Information Management System*

ACM Classification Keywords*: I. Computing Methodologies; I.2 Artificial IOntelligence I.2.11 Distributed Artificial Intelligence; Multiagent systems*

Introduction

Business organizations today are becoming increasingly complex systems. In order to manage this emerging complexity proper toolbox is needed. The paper presents two aspects of agent oriented toolbox. The first one considers support for organization modeling which is currently crucial to improve business processes performance as well as ICT development effort. As North et al. suggest, nowadays organizations are facing several problems which are mainly related to fragmented consumer markets, more interwoven industrial supply chains, sophisticated transportation systems and growing interdependency of infrastructures [North, 2003]. Better understanding of how these would affect specific organization requires the business model that may be analyzed from different angles and on several levels of abstraction. The model development process is conducted with the use of modeling language that must have modeling constructs which enable to fully express the characteristics of business problems. As there is shown in the paper, agent orientation is a new way to capture and analyze the structure and behavior of enterprises. Multi-agent system as a modeling metaphor can be considered very natural and intuitive paradigm for building business models.

The second aspect of the toolbox regards the agent based software support for process and knowledge oriented business organizations. As paper implies, performance improvement efforts should take into consideration business computing paradigm shift – from "direct manipulation metaphor" to "do not navigate – delegate!" way of

using software systems. In such scenario, the societies of software agents work autonomously on behalf of business actors, who delegated to them sophisticated tasks, helping organization to effectively achieve its goals.

Basic concepts of agent orientation

Before advantages of agent oriented organization modeling and performance improvement will be presented, it seems advisable to explain the essence of agency.

The agent

Over the last two decades the concept of an intelligent agent has become really popular. A number of researchers dealing with artificial intelligence focused on agency. Consequently numerous definitions of an agent have been coined. Two of them have been mentioned below.

Michael Wooldridge and Nicholas R. Jennings describe an agent as: "a hardware or (more usually) software-based computer system that enjoys the following properties:

- autonomy: agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- social ability: agents interact with other agents (and possibly humans) via some kind of agent-communication language;
- reactivity: agents perceive their environment, (which may be the physical world, a user via a graphical user interface, a collection of other agents, the INTERNET, or perhaps all of these combined), and respond in a timely fashion to changes that occur in it;
- pro-activeness: agents do not simply act in response to their environment, they are able to exhibit goal-directed behavior by taking the initiative [Woolbridge, 1995].

Another definition has been proposed by S. Franklin and A. Graesser in their paper attempting to distinguish software agents from regular computer programs [Franklin, 1996]: An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.

Basing on this definition few vital attributes of an agent can be noticed:

- an agent exists in a certain environment and thus it ceases to be an agent when extracted from such environment,
- an agent senses its environment, acts on this environment and its actions can affect what an agent will sense in the future,
- an agent operates over time and acts whenever it "feels" it's necessary; unlike regular program which terminates once its mission is accomplished,
- an agent operates autonomously pursuing its own goals.

All these basic characteristics constitute conceptual framework that will be used later when trying to show how agent oriented applications may lead to organization performance improvement.

Multi-agent systems

A Multi-Agent System may be defined as a set (society) of decentralized software components (where every component exhibits the properties of an agent, mentioned in the previous section), that are carrying out tasks collaboratively (often in parallel manner) in order to achieve a goal of the whole society. Later in the paper this definition has been disaggregated and all the properties are used to show why the multi-agent system can be considered as a very intuitive organization modeling metaphor.

As presented definition reveals, software agents have the ability to collaborate with each other which enables the creation of multi-agent systems. Collaboration is defined as a process in which society coordinate its actions in

order to achieve common goals. Software agents are able to collaborate with one another as well as human agents.

The corner-stones of inter-agent collaboration are: communication and knowledge sharing. Communication is basically an exchange of information among agents (agents can send messages to each other, observe each other's state and behavior, however, communication takes place on the so called knowledge level). To enable knowledge sharing agents must have common goals and decompose the process of achieving these goals into sequence of actions providing that every agent is capable of performing task assigned to it.

Inter-agent collaboration requires also a communication language. Currently the most popular agent communication languages are: Knowledge Query Manipulation Language (KQML) developed in early 90's and FIPA-ACL developed by Foundation for Intelligent Physical Agents. Both rely on speech act theory and define a set of performatives, their meaning and protocol for perfomatives exchange.

To enable agents to understand each other they must not only speak the same language, but also have a common ontology. According to Gruber "…An ontology is an explicit specification of a conceptualization… A body of formally represented knowledge is based on conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them" [Gruber, 1993].

Within an agent community agents have different knowledge, capabilities and responsibilities. Each one can have access to different resources and can perceive certain matters differently. Developing a successful multi-agent system demands effective coordination of autonomous agents thus creating a synergy enabling this system to solve problems in dynamic environment despite imperfect data and information. To achieve it one must answer few critical questions [Farhoodi, 1993]:

- What are appropriate agent architectures for different classes of problem-solving?
- How does an agent acquire its knowledge and how should it be represented?
- How does a complex task get decomposed and allocated to different agents?
- How should agents cooperate and communicate with each other?
- Can an intelligent agent be trusted?

Agent orientation as a modeling paradigm

When considering agent orientation as a modeling paradigm it is essential to answer two fundamental questions. The first one is why agent and multi-agent system characteristics make agents so natural and intuitive organizational modeling constructs? The second asks why agent orientation is optimal choice for complexity management? The following sections answer them one by one.

Agent as a modeling metaphor

One of the fundamental assumptions for this article is that organization modeling process bases on multi-agent system metaphor, that leads to perceiving and understanding of organization in the way typical for multi-agent system software engineering, but also takes under consideration business aspects along with basic organization characteristics.

The use of a metaphor during the process of organizational analysis and understanding is of great significance. Morgan says that metaphor is frequently understood as a way to make an argument more appealing, however its meaning is much more important. The use of a metaphor is a consequence of a way one thinks and perceives which penetrate our understanding of the world that surrounds us [Morgan, 2005].

We use metaphor when we try to understand some portion of reality using another its portion. Thus we formulate a theorem that "A is B" or that "A is similar to B". Concerning this article an assumption has been made that in qualitative terms the characteristics of modern organizations are really close to the characteristics of multi-agent systems both in structure and in behavior.

The appearance of obvious similarities between multi-agent systems and business organizations is particularly visible in the case of companies which adopted such business strategies as decentralization of operations and process orientation, that led to the creation of specific organizational forms [Drucker, 2000].

The table 1 shows observed similarities. In the first column are basic structural and behavioral characteristics of multi-agent systems, and in the second characteristics of modern organizations extracted from key titles which deal with organization design and process orientation.

Table 1. Similarities between structural and behavioral characteristics of modern organizations and multi-agent systems [Jakieła, 2006].

Multi-agent system	Business organization
Multi-agent system is a set of **decentralized** software components.	Centralized model and functional decomposition reflect precisely decomposition of workload into smaller tasks, that are distributed among particular company departments and are accomplished in sequences [Peppard,1997]. Modern organizations operate according to decentralized business process patterns which are accomplished by distributed organizational actors. Decentralization causes, that these processes are moved from companies headquarters to local offices. The operational model of modern companies is a highly decentralized.
Multi-agent system is a set of **autonomous** software components.	Decentralization requires in turn autonomy delegation, that has drastically changed the role of organizational actors, because "controlled positions" have been replaced by positions which give full competence [Drucker, 2000]. This trend is really similar to that which takes place in software engineering due to control encapsulation into distributed components, which therefore poses operational autonomy. According to Champy and Hammer people working within processes that are being reengineered must be equipped with delegations. As members of a process team they are both allowed and obliged to make decisions. [Hammer, 1996]. In case of process orientation it is impossible to avoid situation when organizational actors, who perform process oriented jobs, are fully autonomous entities.
Multi-agent system is a set of **goal-oriented** software components.	In modern organizations functional departments have been replaced by process teams. A set of organizational actors, who cooperate in order to achieve particular goals of certain process is a natural form of modern firm's organization [Hammer, 1999].
Multi-agent system is a set of software components, which may carry out tasks in **parallel manner**.	In the company organized around processes, subsequent work-stages are performed in natural order. Instead of artificial operations order, natural operation order is used. Processes' de-linearization allows task performance acceleration due to two factors. Firstly, lots of tasks are performed in the same time. Secondly, shrinking of time between initial and final stages of processes causes that the necessity for serious changes, which can undermine or make incoherent work done so far is less probable [Hammer, 1996].

Agent orientation and complexity management

Thank to the use of an agent as a modeling construct designer/analyst can cope better with problem domain and design process complexity.

Complexity management problem is a really vital issue, because each human is in this field inherently limited. When analyst/designer sees the problem domain, he is trying to solve, for the first time, he sees vast variety of components, which interact with one another in many various ways. Trying to structuralize the model during the analysis she is forced to take under consideration plenty of factors. Unfortunately, according to research conducted by cognitive scientists, each human being can work effectively with 7±2 portions of information at the same time [Miller, 1956]. Simon claims that the velocity of information processing by the human, which totals approximately 5 sec for the perception of each new portion of information, is an additional restriction [Simon, 1996].

Complexity management is also a serious problem in case of organization modeling. Each organization is undoubtedly a complex system. According to Carley "Organizations are heterogeneous, complex, dynamic nonlinear adaptive and evolving systems. Organizational action results from interactions among adaptive systems (both human and artificial), emergent structuration in response to non-linear processes, and detailed interactions among hundreds of factors." [Carley, 1999].

Booch [Booch, 2007] relying on Simon's work [Simon, 1996] has defined the following set of basic characteristics of complex systems:

Complexity frequently takes a form of hierarchy, where the system is composed of sub-systems connected with each other, which have their sub-systems, which in turn have their sub-systems and so on until the elementary level is reached. This hierarchy does not mean the superior-subordinate relation. Thank to the fact, that complex systems are nearly decomposable we can fully understand them, describe or even perceive. Simon claims that it is highly probable that in reality only the systems that have a hierarchical structure can be understood [Simon, 1996]. Looking at process oriented organizations from this perspective, it is possible to distinguish such levels of hierarchy as organization actors level, business process level, singular organization level and specific configuration of few organizations in a form of extended enterprise or virtual organization (See Fig. 1.)

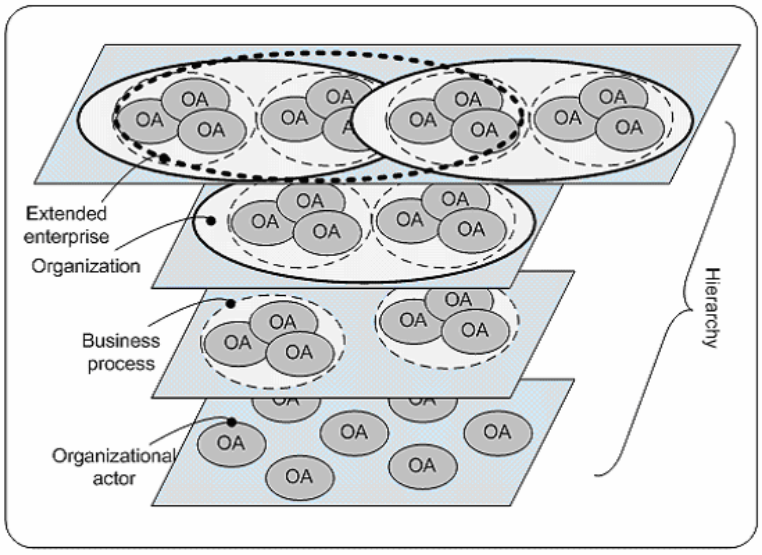


Figure 1. Business organization as a complex system

The choice which components of a system should be treated as elementary is arbitrary and depends on the system observer's decision.

It is possible to identify interactions taking place between sub-systems as well as inside sub-systems between their components, however interactions of the second type have one row higher frequency and are more predictable. The interaction frequency will differ depending on the level of hierarchy. For example, within a business organization more interactions will take place between employees working on the same process than between teams of employees working on different processes. The differences in interaction frequency within and

between sub-systems allow decomposition and lead to the clear division between domains of analysis. In case of social systems, and undoubtedly every organization can be seen as such system, nearly decomposable character is clearly visible, therefore it is possible to exploit advantages of the decomposition method.

Complex systems are mostly sets of similar elements composed in various combinations. In other words there are certain common templates created on the basis of reuse of similar elementary components or more complex structures in the form of sub-systems.

Systems organized hierarchically tend to evolve over time, and hierarchical systems evolve faster than non-hierarchical ones. Simon claims that complex systems will evolve out of simple systems, if certain intermediary forms exist [Simon, 1996].

Taking under consideration basic characteristics of complex systems as well as agent approach described above we can start our argumentation, purpose of which is to show advantages of agent approach in the context of complexity management in the organization as well as information system modeling process.

As the first argument it can be noticed, that agent oriented decomposition of a problem domain is an effective way to division of the problem space, while modeling organizations and information systems. It can be concluded from a number of factors.

Firstly, hierarchical structure of complex systems causes, that modularization of organization components in terms of goals, that are to be achieved is a really intuitive solution. As Jennings and Wooldridge claim hierarchical organization of complex systems causes that at each level of the hierarchy the purpose of the cooperation between sub-systems is achieving a functionally higher level. Whereas within sub-systems components, which these sub-systems are composed of cooperate in order to achieve total functionality of a sub-system. As a consequence, decomposition oriented on goals that are to be reached is very natural division [Jennings, 2000]. Applying this schema to an organization the situation emerges where organization actors cooperate in order to achieve goals of the process, in turn processes are realized in order to achieve the goal of the organization, and organizations combine their inherent competences in order to achieve goals of the extended enterprise or virtual organization. It is worth to remember that goal orientation is one of the main characteristics of an agent and thus agent concept can be used without any additional effort.

Another vital issue is presentation of such characteristic of a modern organization as decentralization in the area of information processing and control. In this case agent oriented decomposition seems to be an optimal solution due to such characteristics of an agent as thread of control encapsulation in the form of autonomy property. The distributed organizational components may be thus modeled with autonomous agents as a basic modeling constructs.

Agent oriented approach allows also to solve problems connected with the design of interactions taking place between system components. It is a serious issue due to the dynamics of interactions between organization components. It is really frequent, that organization components enter an interaction in difficult to predict time and for unknown at the stage of design reasons. As a consequence it's really challenging to predetermine parameters of such interactions. The solution to this problem is existence of system components with characteristics thank to which they can make decisions concerning the type and range of interaction not before the system is running.

Another argument for an agent oriented approach is that it allows to eliminate semantic gap between agent abstraction used during the information system design phase and structures used during organization modeling. It is directly connected with similarities which appear between structural and behavioral characteristics of a multi-agent system and organization (table 1). Continuing this thread it is advisable to point out the following conveniences:

Mutual interdependencies present among organization actors and organization sub-systems can be naturally mapped into the system architecture in terms of high-level social interactions which take place among agents.

In most organizations dependencies of this type are present in the form of really complex network of dynamically changing relations. Agent based approach includes mechanisms which allow to describe such relations. For example, interaction protocols such as Contact Net Protocol can be used in order to dynamically create business

process teams, which can be in case of such need activated and after reaching particular goals deactivated. What is more, there are off-the-shelf structures, which can be used during the community modeling, what is really useful when modeling organization actors and sub-systems [Jennings, 2000].

The process of organization modeling and system design frequently requires to perceive modeled object from the perspective of various abstraction levels, treating set of elements as atomic modeling structure. The idea of an agent is flexible enough to be used in an elementary component role on any detail level depending on the analyst's needs. For example, an agent could be organization actor, process or whole organization and components treated as elementary interact only in an integrated form omitting details concerning intra-interactions.

Organization modeling and system design with agent oriented approach leads to the structure, which has numerous stable intermediary forms, what is really important concerning complexity management. Among others it means that system components in the form of agents can be created rather independently and in case of such a need added to the system providing a smooth functionality growth.

Besides advantages of agent oriented approach mentioned above, additional benefits can be pointed out, which are connected with agent oriented approach during information systems development. They are analyzed in details in monograph [Cetnar, 1999]. As its author claims, agent oriented approach should lead to improvement and enrichment following characteristics of information systems:

- *Flexibility*. System can be rather easily adopted to the performance of new tasks in changing environment. The adaptation process is much more flexible with the use of an agent approach.
- *Reliability*. Particularly in case of heavily developed systems, which can cause un-stability problems understood as permanent or temporary loss of the ability to perform tasks of the system.
- Ability to combine functions of various systems flexibly, as well as ability to create new properties such as self-organization or self-adaptation.

What is more, application of agent orientation may lead to the simplification, improved reliability, and as an effect decrease in costs of analysis, design and implementation of decentralized information systems.

Agent oriented vs. object oriented modeling

It can be seen that agents are really similar to objects and as a consequence agent based modeling is similar to object oriented approach. However agents have some vital advantages over regular objects which can bring modeling of contemporary enterprises one step higher.

Agents are intelligent, have the ability to learn. They can use regular objects in pursuit of their own goals. Their behavior may vary depending on the circumstances and environment due to their experience. They pursue their goals actively cooperating with each other and influencing one another. As Farhoodi suggests "…business objects make a major contribution to modeling information in the enterprise. Intelligent business agents extend this capability to provide the breakthrough in modeling knowledge in the enterprise" [Farhoodi, 1993].

Object oriented notation is really well suited for software engineering, but can be difficult to understand by business people. As a consequence models created with the use of OO techniques are difficult to validate for them. Whereas models created with agent oriented approach are much more comprehendible, because the same perspective and vocabulary is used during analysis, design and construction phases in software development cycle. The example of how agent orientation may significantly reduce the semantic gap between organization model and management information system architecture has been presented in [Jakieła, 2006]] in the form of AROMA methodology.

To emphasize once again the potential of agent oriented approach let's quote Farhoodi and Fingar: "…Intelligent agents can facilitate the incorporation of reasoning capabilities within the application logic (e.g. encapsulation of business rules within agents or modeled organizations). They permit the inclusion of learning and self improvement capabilities at both infrastructure (adaptive routing) and application (adaptive user interfaces) levels.

Intelligent user interfaces (supporting task centered user interfaces and intelligent assistance to end-users) can be a boon to productivity in a network-centric world" [Farhoodi, 1993].

Agent as a tool for performance improvement

When evaluating the ways agents can be used for performance improvement it is advisable to look at the organization from the perspective of value chain model (see Fig. 2.). Porter introduced the value chain concept as a systematic way of examining all activities a firm performs and how applications of ICT interact to provide competitive advantage. A firm gains competitive advantage by performing these strategically important activities in more efficient way then competitors [McCormack, 2003]. It is possible when properly chosen ICT are applied to the processes which have the highest information density and/or are used as activities coordination mechanism.

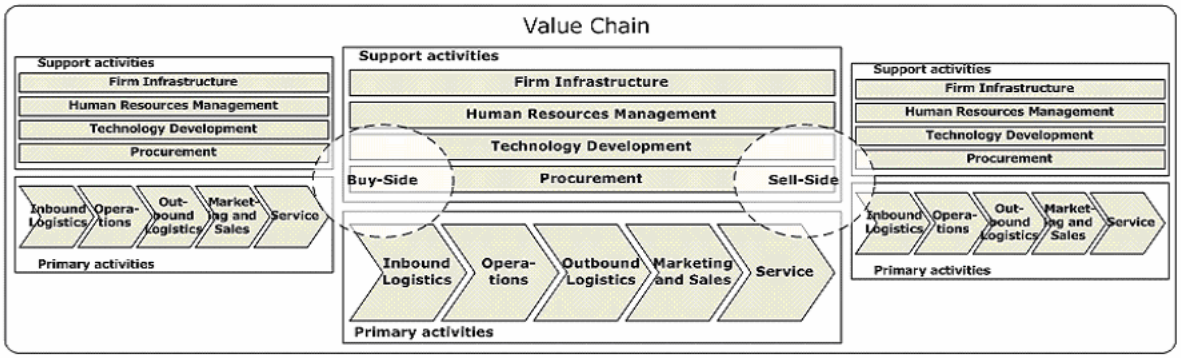


Figure 2. The value chain model

Besides the basic activities that organization performs individually, it is also important to think about links that connect it with its business partners. These "market interfaces" are known as buy-side and sell-side of firms operations. The analysis that follows, showing advantages of using agent oriented software in the process of optimizing inter- and intra-organizational effectiveness, has been divided into two parts: Agents in Supply Chain Management and Agents in Demand Chain Management. In order to reveal the superiority of agent technology over the traditional solutions, the references have been made to most important agent's characteristics (See section entitled The Agent).

Agents in Supply Chain Management

Contemporary enterprises integrate sell-side and buy-side activities into sophisticated Supply Chain Management (SCM) systems. Figure 3 depicts the generic structure of typical supply chain.

As in [Singh, 2005] was shown, agents may work as Management Information Systems, enabling integration of information flows across multiple electronic marketplaces and thus support performance improvement of the critical SCM activities. In this case the delegation of tedious tasks to agent society enables the availability of market information to all business partners as well as effective and efficient coordination of the supply chain interactions. The coordination effort is mainly related to dynamic supply and demand planning but the aggregated information flowing through e-marketplaces is extensively used also for such tasks as real-time planning, buyer-supplier selection and transaction facilitation.

It is easy to observe that agent orientation of such solution provides significant advantages to all participants over traditionally automated (systems not using agent-oriented approach) supply chains of competitors. The argumentation may be based on the analysis of basic characteristics of agents in this specific context.

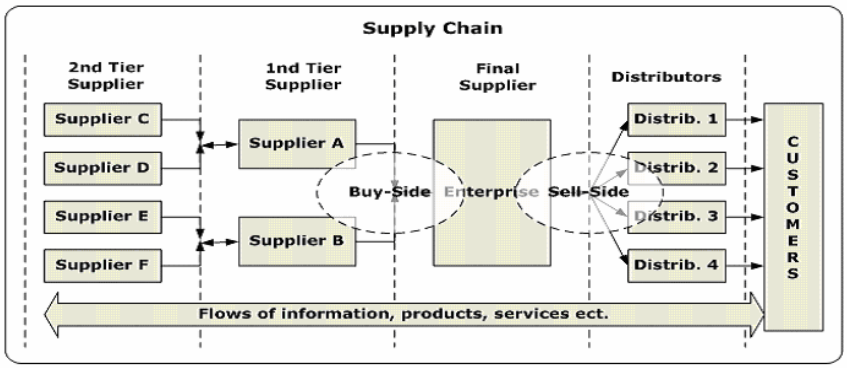


Figure 3. The generic model of Supply Chain

The SCM system architecture described in [Singh, 2005] has been developed with the use of agent oriented conceptualization of problem domain. As has been already shown in the paper, agents can be very natural and intuitive modeling constructs. Mentioned architecture consists of basic agents for buyers and suppliers as well as several auxiliary agents (discovery agents, transaction agents, monitoring agents). Agent oriented decomposition significantly reduced the semantic gap between the high-level description of the business requirements and system architecture.

The agents are situated in dynamically changing environment consisted of e-marketplaces related to business actors roles in the whole supply chain. Because agents are inextricably linked with their environment – what is an essence of agency – this is quite natural mapping between business context and the information architecture enabling incorporating all important flows into the system. Agents are monitoring role specific data such as demand requirements, matching suppliers' properties, transaction details and level of satisfaction using their sensors and affecting the flows through effectors.

Every agent operates autonomously and knows its role and behavior related to it. It is working on behalf of the user, and what is important, it performs goal oriented tasks, without direct intervention of human agent.

Roles are connected with goals, which agents are trying to proactively achieve. For example transaction agent's goal is optimal transaction facilitation, buyers and suppliers agents' goal is to maximize the level of satisfaction related to executed transactions, discovery agent is responsible for matching buyers and suppliers, maximizing the value of utility and monitoring agents fulfill the goal of marketplace synchronization.

Finally, the social abilities of agents enable the cross e-marketplace information transparency. Communication among agents plays the role of coordination mechanism and allows for dynamic and transparent planning of demand and supply requirements through real-time information integration across agent-oriented supply chain.

Agents in Demand Chain Management

On the other value chain side enterprises develop Demand Chain Management systems. The role of Management Information Systems again can be taken over by agents. Such agents are mainly used by companies to enhance the performance of Customer Relation Management Systems enabling constant assistance throughout the whole buying decision making process. According to Turban [Turban, 2008] this process includes such stages as need identification, information search, product or merchant brokering, purchase terms negotiation, payment and delivery facilitation and even post-purchase support (See Fig. 4.). On every stage customer has a goal and tasks needed for its achievement may be smoothly mapped into agent functionality.

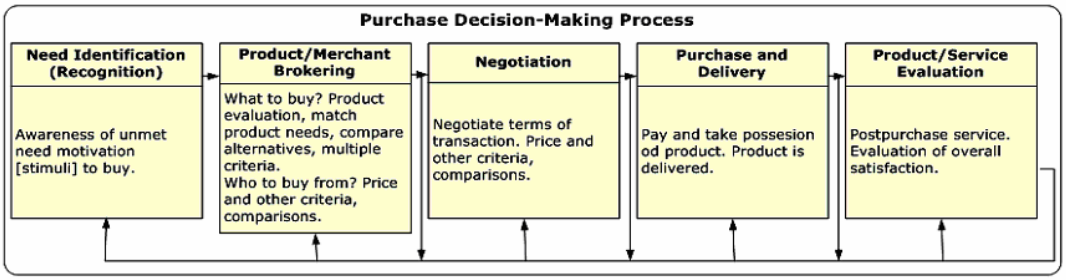


Figure 4. Buying Decision Making Process

Agents in such scenario allow unobstructed information exchange between vast numbers of actors (sellers and buyers) participating in numerous e-marketplaces, and are of invaluable help to buyers who are forced to deal with the problem of information overload thus greatly improve companies relations with their customers.

It is obvious that the application of agents in such context can bring vital advantages for all the participants (buyers and sellers) in terms of performance improvement. Let's refer again to the basic characteristic of agents. The fact that agents are highly decentralized entities enables smooth decomposition of a problem domain thus creating numerous agents responsible for various tasks related to all the stages in decision making process. For example we can observe agents that support need identification, reduce information overload, provide customers with comparison shopping engine, facilitate payments and offer after-sales support.

Agents are proactive what enables them to aggressive solution searching, active monitoring of vast amounts of information dispersed among numerous sources and as a consequence bringing the edge over traditional passive systems, which are much less effective concerning present rapidly changing environment. An agent can monitor a given information source without being dependent on the system from which it originates. Agents can wait for certain kinds of information to become available. It is often important that the life spans of monitoring agents exceed or be independent of the computing processes that created them [Maes, 1999]. There are agent implementations that are using comparable techniques to recommend complex products based on multi-attribute utility theory and assist customers in the merchant-brokering and negotiation stages. Some agent oriented solutions use automated word-of-mouth recommendation mechanism called "collaborative filtering." Whenever someone would like to buy something the system first compares a shopper's product ratings with those of other shoppers. After identifying the shopper's "nearest neighbors," or users with similar taste, the system recommends the products the neighbors rated highly but which the shopper may not yet have rated [Maes, 1999]. Proactive agents can also acting on their own initiative negotiate terms of transactions, organize product/service delivery as well as evaluate consumers satisfaction.

Agents are autonomous what allows them to work effectively on behalf of the customers, taking initiative, identifying the need of their "bosses", independently searching for the best product as well as merchant choice, making decisions related to their goals and form valid contracts. In automated negotiation for instance, agents find and prepare contracts on behalf of the real-world parties they represent. This automation saves negotiation time, and agents are often better in finding deals in combinatorially and strategically complex settings. Agents also make it possible to provide customers with dynamic pricing models. The main benefit is that the burden to determine *a priori* the price of a good is pushed into the marketplace. As a result the limited resources are allocated to those who value them most. Using dynamic pricing models in the real world may be to costly to customers (geographical collocation, time spent for offers monitoring etc.). This is where agent technologies come in. Customer can delegate the task by creating the auction, specify auction parameters (reservation price, clearing times and method for resolving tie bids) and let agent to do the rest for her. Autonomous agents can also

participate during purchase and delivery stages as well as during post-purchase stage taking care of post-purchase support.

Each agent is goal oriented and each goal depends on the role of a particular agent. For example need recognition agents “keep an eye” on information sources and basing on consumer preferences inform her when the specific product is available, other agents analyzing vast quantities of information select best products and vendors, transaction agents negotiate purchase terms, delivery support agents facilitate purchase and delivery tasks, call center agents solve consumer problems and in case it is necessary connect them with appropriate human assistant. As can be seen the whole decision making process may be significantly supported by pro-active agents’ services, increasing consumer satisfaction level and creating switching costs.

What is more, agents can form societies thus enable smooth flow of information, strengthen cooperation between various actors and through the exchange of experience improve consumer trust. Agent societies sharing information can significantly enhance consumers need recognition as well as product and merchant selection. Agents organized in societies can not only negotiate terms of transaction and arrange product delivery, but also exchange information concerning post-purchase satisfaction thus helping each other making better choices in the future and greatly improving overall consumer’s loyalty.

Conclusion

Organization modeling is still a niche, but it is developing rapidly. A progression from proprietary models and tools to new standards can be seen. The growing complexity of contemporary firms requires the tools enabling us to better understand distributed and knowledge oriented business processes. It is highly likely that in ten years, business modeling will be the mainstream. Using business models, will become the natural and ordinary way for interdisciplinary teams to communicate, much as software modeling is mainstream today for software engineers.

More and more sophisticated business architectures require also flexible ICT infrastructure that will properly coordinate their operations. The paper shows that agent-orientation is a very promising solution. As Maes et al. predict “…in the near future the agent oriented solutions will explore new types of transactions in the form of dynamic relationships among previously unknown parties. At the speed of bits, agents will strategically form and reform coalitions to bid on contracts and leverage economies of scale” [Maes, 1999].

As a summary the following conclusions have been drawn:

1. Taking into consideration that characteristics of decentralized and process oriented organizations are semantically very close to properties of multi-agent systems, using agent oriented conceptualization seems to be very natural and intuitive modeling metaphor. This direct mapping eliminates semantic gap between business requirements and management information systems architecture.
2. Agent orientation can be considered as a toolbox for complexity management. It has been shown by analyzing the relationships between the classical definition of complex system provided by Simon and structures and mechanisms offered by agent paradigm. The paper shows why agent approach is especially well suited for dealing with inherently complex systems such as contemporary business organizations.
3. The application of agent–oriented software may lead to performance improvement of firms due to software agents’ properties. What can be easily observed, agent technology started to transform the way companies conduct business but the real prominent changes will occur as agent technologies mature. The predictions say that the next wave of agent solutions will be able to better manage ambiguous content, personalized preferences, complex goals, changing environments, and disconnected parties. However it is important to remember that the full adoption of agent technologies will occur after standards are widely accepted and used. Unfortunately it may take some time as in case of object oriented technologies where it took over three decades.

Bibliography

[Booch, 2007] G. Booch, R. A. Maksimchuk, M. W. Engel, B. J. Young. Object-Oriented Analysis and Design with Applications. 3rd Edition. Addison-Wesley, 2007

[Carley, 1999] K. M. Carley, L. Gasser. "Computational organization theory", in Weiss G. "Multiagent systems. A modern approach to distributed artificial intelligence", MIT Press, 1999.

[Cetnarowicz, 1999] K. Cetnarowicz. "Problemy projektowania i realizacji systemów wieloagentowych", Uczelniane wydawnictwa naukowo-dydaktyczne AGH, Kraków, 1999.

[Drucker, 2000] P. F. Drucker. „Zarządzanie w XXI wieku", Muza S.A., W-wa 2000.

[Farhoodi, 1997] F. Farhoodi, P. Fingar. Competing for the Future with Intelligent Agents; DOC Magazine; 1997.

[Franklin, 1996] S. Franklin, A. Greasser. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents; University of Memphis; 1996.

[Gruber, 1993] T. A. Gruber. A Translation Approach to Portable Ontology Specifications; 1993.

[Hammer, 1996] M. Hammer, J. Champy. Reengineering w przedsiębiorstwie, Neumann Management Institute, Warszawa, 1996.

[Hammer, 1999] M. Hammer. Reinżynieria i jej następstwa, Wydawnictwo Naukowe PWN, Warszawa, 1999.

[Jakieła, 2006] J. Jakieła. AROMA – Agentowo zorientowana metodologia modelowania organizacji; WAEiI Politechnika Śląska, Gliwice, 2006.

[Jennin00] N. R. Jennings. Wooldridge "Agent-oriented software engineering", Proceedings of the 9th European Workshop on Modeling Autonomous Agents in a Multi-Agent World : Multi-Agent System Engineering , 2000.

[Maes, 1999] P. Maes, R. H. Guttman, A. G. Moukas. "Agents that buy and sell. Communication of the ACM, Volume 42 , Issue 3, 1999

[McCormack, 2003] K. McCormack, W. C. Johnson. Supply chain networks and business process orientation. Advanced Strategies and Best Practices, CRC Press LLC, 2003.

[Miller, 1956] Miller, G. March. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", Psychological Review vol. 63(2), p. 86, 1956.

[Morgan, 2006] G. Morgan. Images of organizations, Sage Publications, 2006.

[North, 2003] M. North, C. Macal, P. Campbell. Oh Behave! Agents-Based Behavioral Representations in Problem Solving Environments, LNCS, Springer Verlag, 2003.

[Peppard, 1997] J. Peppard, P. Rowland. „Reengineering", Gebethner i S-ka, Warszawa, 1997.

[Singh, 2005] R. Singh, A. F. Salam, L. Iyer, Agents in e-supply chains, Communication of the ACM, vol 48, No. 6, 2005.

[Simon, 1996] H. Simon. "The Sciences of Artificial", MIT Press, 1996.

[Turban, 2008] D. King, P. Marshall, J. K. Lee, E. Turban, J. McKay. Electronic Commerce 2008, 4/E, Prentice Hall, 2008

[Wooldridge, 1995] M. Wooldridge, N. R. Jennings. Agent Theories, Architectures, and Languages: A Survey; Berlin: Springer-Verlag; 1995.

Authors' Information

Jacek Jakieła, Ph.D., Eng. *– Department of Computer Science FMEA RUT; W. Pola 2, 35-959 Rzeszow, Poland; e-mail: jjakiela@prz.edu.pl*

Bartosz Pomianek, M.Phil. *– Department of Mathematical Economics and E-business, University of Information Technology and Management, Sucharskiego 2, Rzeszow, Poland ; e-mail: bpomianek@wsiz.rzeszow.pl*

AGENT TECHNOLOGIES FOR WEB MINING FROM A NON-EXTENSIVE THERMODYNAMICS PERSPECTIVE

Franciszek Grabowski, Marek Zarychta, Przemysław Hawro

***Abstract**: Motivated by the works that reveal the small-world effect and scale-free property of various real-life networks, many scientists devote themselves into studying complex networks. One of the ultimate goals is to understand how the topological structures of networks affect the dynamics upon them. In this paper, we give a brief review on the studies of agent technology appliances for Web mining which is performed in complex network environment ruled by non-extensive thermodynamics.*

***Keywords**: non-extensive thermodynamics; web mining; complex networks; multi-agent technology.*

***ACM Classification Keywords**: I.2.11 Distributed Artificial Intelligence - Multiagent systems*

Introduction

The study of complex systems has became one the most active areas of research. This subject has interest in natural sciences as well as in social and artificial systems. Typical features present in complex systems are long-range interactions, long-term memory, fractal phase-space structure, scale-free network structure, or even combinations of these characteristics. These systems are common in computer science and cannot be correctly described by the well-established Boltzmann-Gibbs statistics. Such systems lack linear relationship between the input and the output, or between a cause and its effect. Formally, such relationships are commonly described by a power law. Scientists are particularly interested in two types of dynamics in complex systems and complex networks:

- event-driven processes, a dynamic process is apparently caused by external interferences, for instance, a fire caused by a dropped match
- self-organized criticality, a system in which dramatic changes may take place in the absence of major causes at the macroscopic level [Chen, 2003].

Despite current complex network theory have not yet reached a level that one can specifically identify the cause–effect relations associated with the dynamics observed over a complex network, we should consider it as new approach to well known areas of research including web mining and multi-agent systems.

Network dynamics and topology of World Wide Web

Robust development in Information Technology and it's still gaining popularity has impact on the World Wide Web acting as an information super highway. Millions of people all around the world have access to miscellaneous information scattered over the Web. The distributed and dynamic nature of the Web has thrown down the gauntlet to information retrieval on this complex information space. The software agent technology seem to be appropriate to pick up this gauntlet. The system consistent of autonomous, collaborative and adoptive software entities is suitable for solving complex problems and coping with huge amount of information, but this software agents must be aware of their surrounding and run strictly connected to it.

Existing in real world networks, reveal more complex structures than were firstly considered, so the sense of topology evolved and acquired importance [Dorogovtsev, 2002]. Knowledge of topology is necessary to understand phenomena occuring in the network.

In the simplest random network model, vertex pairs are connected with each other with some probability value. Traditional approach to physical networks expose them as real finite dimensional regular lattices, field fully connected graphs or random graphs. Watts and Strogatz have developed model that interpolates between

a regular ordered graph and a random graph calling it small-world phenomenon [Watts, 1998]. Their model is appropriate for systems having a high degree of local clustering combined with a short finite path length between any pair of vertices. This feature is common among many social networks of human relationships, electric power grid, etc. A small-world network can be constructed starting from a regular lattice, in which each vertex is joined to its neighbours or fewer lattice spacings away, and then adding or moving a portion of the edges. The moving can be done by examining each vertex in turn and with some probability moving the other end of the edge to a different vertex chosen at random. The result is a lattice with shortcuts.

To study the degree distribution of a network, let p_k denote the fraction of vertices with degree k, or, the probability that a vertex chosen at random has degree k. Random network models produce usually a Poisson distribution for p_k , whereas most real-world networks have highly right-skewed degree distributions, meaning that there are lots of vertices having a few connections, and some vertices have many connections — highly-connected vertices are practically absent in random and small-world networks. The networks with right-skewed degree distributions have no characteristic scales for the degrees, thus the networks of this kind are called scale-free. Their degree distribution follows a power law $p_k \sim k^{-\gamma}$. Barabási and Albert built the model and found that the exponent in the power law has been approximated for many different real-world networks having values in range $2 < -\gamma < 4$ [Albert, 2002]. The World Wide Web topology based on preferential linking and revealing small-world phenomenon is currently one of the largest global social networks for which topological information is available.

Web mining

Web mining is the use of data mining techniques to automatically discover and extract information from web documents and services [Dunham, 2006]. There can be seen huge grow and impressive evolution of the Web uncovering scalability problems of actual Web search engines. The Web structure can be compared to graph structure where pages appear as vertices and hyperlinks as edges. Collecting miscellaneous useful informations from this structure is called Web mining and can be divided into three categories:

1. Web content mining,
2. Web structure mining,
3. Web usage mining.

Web content mining aims at the knowledge discovery, where the objects are common collections of text documents, sets of multimedia documents such as images, videos, which are incorporated or linked to the Web pages. Web content mining from the agent-based approach aims on improving the information finding and filtering and could be divided into the following three categories [Cooley, 1997]:

- Intelligent Search Agents. These agents search for relevant information using domain characteristics and user profiles to organize and interpret the discovered information.
- Information Filtering/ Categorization. These agents use information retrieval techniques and characteristics of open hypertext Web documents to automatically retrieve, filter, and categorize them.
- Personalized Web Agents. These agents learn user preferences and discover Web information based on these preferences, and preferences of other users with similar interest.

There is another approach to the Web content mining which aims on modeling the data on the Web into more structured form in the purpose of applying standard database querying mechanisms and data mining applications to analyze it. This approach is categorized to Multilevel databases and Web query systems.

Web structure mining concentrates on revealing the structure of the Web. The effort is focused on discovering the model underlying the hyperlink structures of the Web. This model can be used to categorize the Web pages and is useful to generate information such as similarity and relationships between Web sites containing important information which can help in filtering or ranking Web pages. Divergent nature of the Web and its objects creates new challenges and even obstacles, because there is not possibility to strictly make use of existing techniques such as database management or information retrieval. Since the Web is full of loops and traps, to generate a Web structure, the circuits and repetitive cycles should be detected and removed, without which a client may be

lost in Cyberspace through the complex cycles. Multi-agent systems seem to be well suited and have sufficient capabilities to perform such task.

Web usage mining covers the prediction techniques of the users behavior and their interactions with the WWW. The most useful source containing information for such analyses are logs from Web servers. The data collected from Web log records allows to discover user access patterns of Web pages and their interests. Typical applications generated from this analysis can be classified as personalization, system improvement, site modification, business intelligence and usage characterization [Cooley, 1997]. Web usage mining sometimes encounter obstacles for instance, due to the collaboration lack of the users or webserver administrators, who tend to keep the Web log records as jealously-guarded secret. Due to this fact, privacy plays a key role in Web usage mining, because users should be at least warned about privacy policies before they admit to reveal their personal data. There are no straight boundaries between Web structure mining and Web usage mining, hence all of them could be used in combined applications, which will not be discussed here.

Non-extensive thermodynamics and its impact on web mining techniques.

Since most of the systems in nature are in nonequilibrium states, not even tending toward equilibrium, over the years considerable effort has been devoted to the development of nonequilibrium statistical mechanics and nonequilibrium thermodynamics. The fluxes of energy, matter, etc. in nature are irreversible [Prigogine, 1961]. The second law of thermodynamics fixes the direction of these irreversible processes by specifying that the accompanying entropy production should be always positive. This is so indifferent of whether the system is open or closed, and independently of whether entropy flows into or out of the system to its surroundings. For an isolated system the second law therefore indicates that entropy can never decrease, but it does not affect open systems, where entropy can either increase or decrease. The Shannon information entropy defines entropy in terms of the probability distribution of observations, or the information applying to a set of observations. In nonequlibrium systems, irreversible behavior occurs as information is lost in the observation process. Driving a system away from equilibrium breaks symmetry and consequent emergence of organization but can also affect stability of the system by producing turbulences. Questions related to the statistical thermodynamics of irreversibility and self-organization are important in a wide range of new and cross-disciplinary fields, such as epidemiology, evolutionary dynamics, artificial life, agent technologies and web mining [Dewar, 2003].

Most systems, including Web, are not isolated they are open in various ways and they experience fluxes of energy, mass, information, etc. across their boundaries. Although the total entropy of a system plus its environment must increase, it does not follow that the entropy of an open system must increase. Instead, there are many remarkable instances of self-organization of such systems into coherent structures, ranging from tropical cyclones through individual biological organisms to human civilizations. Constantino Tsallis proposed new entropy definition to cope with all these phenomena [Tsallis, 1998]. So the question is not: whether self-organization of the Web will occur or not, but when it will occur and what are general principles to determine it?

In the last few years, there has been a great interest in understanding the topological properties of multi-mesh peer-to-peer networks which are capable of rearranging topology in case of node or link failure. This studies help us understand the behavior of systems such as the Internet and the World Wide Web. Whilst studies driven by traditional approach assume that once a link is created between two nodes, it is never deleted, research devoted to dynamic communication networks show that links are being constantly rewired. An important issue is to discover the topology that, given a search algorithm, optimizes the search process, optimality is defined as the minimization of the average time to perform a search. Clearly, being able to obtain such topology structures seems to be a useful guide to drive the evolution of dynamic communication networks [Cholvi, 2005].

Conclusion

Table1. Reductionism vs. System-Oriented Perspective.

	Reductionistic, Object-Oriented Approach	System-Oriented Approach
Principle	Behavior of a system can be explained the properties of its constituent parts	Multipartite systems have emergent properties that are only possesed by whole system and not the isolated parts
Model characteristics	Linear, predictable, deterministic	Non-linear, stochastic, sensitive to initial conditions, chaotic
Agent interactions	Homeostasis principle, normality, cooperation, self-regulation	Adaptation, robustness, percolations, self-organization
Network topology	Erdös-Rényi graphs, random linking, static wiring	Small-worlds, preferential linking, scale-free networks, dynamic, self-organizing topologies
Probability distribution	Poissonian or Gaussian distribution	Heavy tailed, self-similar, Pareto, Zipf distributions
Entropy definition	Classical Boltzman-Gibbs, Shannon	Tsallis non-extensive entropy
Information flows	Laminar, deterministic fluxes	Turbulent, self-similar fluxes

Still evolving network of hyperlinks and Web pages needs integration of different mining methods taking advantage of multi-agent architecture to permit the discovered knowledge to be verified, reliable and updated automatically. Automatic and non-invasive web personalization seems to be a challenge for nowadays search engines. As we can see in the table 1, underlying principles of web mining techniques should be revised in order for a system perspective to be fully appreciated. The reductionism nature of current paradigm manifests in many aspects leading to wrong perception of the Web. The developing fields of chaos theory, non-extensive thermodynamics and complex system science has not yet sufficient contribution to the Web mining. What becomes evident from these analyses is that the behavior of the system arises from the active interactions of its components appreciating emergent phenomena of coexistent entities. As it was said, the Web is kind of rapidly changing, uncertain environment with indeterministic information, thus for better understanding the phenomena occurring there, we need non-equilibrium thermodynamics and non-extensive statistics, which brings us adequate description of the stability in the open, distributed system in the state far from the thermodynamic equilibrium. Multi-agent systems are naturally suited to run in uncertain environments, for example in networks where there may be a connection or node failure, or someone can sabotage calculation by sending an incorrect data. Multi-agent systems do not require a synchronized clock, they easily adapt to the environment dynamically increasing and decreasing consumed memory size and processor usage (without any impact on performed calculations). They also tolerate a stopovers and frequent delay of communication as well as are capable of taking advantage of heterogeneous environments. The World Wide Web is an interactive and dynamic network in which the properties of single Web site is contingent on its relationships to other sites. Thus Web mining is performed with high risk of mistakes due to cheats and abuse done by unreliable Web-users, -masters etc. Widely used web mining techniques are susceptible to such unfair tricks because have not shifted to system oriented perspective yet.

Bibliography

[Albert, 2002] R. Albert and A.-L. Barabási: Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 2002.

[Chen, 2003] Ch. Chen, N. Lobo: Semantically Modified Diffusion Limited Aggregation for Visualizing Large-Scale Networks, IV, Seventh International Conference on Information Visualization (IV'03), 2003.

[Cholvi, 2005] V. Cholvi, V. Laderas, L. Lopez, A. Fernandez: Self-adapting network topologies in congested scenarios. Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), Vol. 71, No. 3. (2005).

[Cooley, 1997] R. Cooley, B. Mobasher, J. Srivastava: Web mining: information and pattern discovery on the World Wide Web, Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference, 1997.

[Dewar, 2003] R. Dewar: Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states, J. Phys. A: Math. Gen. 36, 631 (2003)

[Dorogovtsev, 2002] S. N. Dorogovtsev and J. F. F. Mendes: Evolution of networks. Advances in Physics, 51(4), 2002.

[Dunham, 2006] M. H. Dunham and S. Sridhar: Data Mining, Introduction and Advanced Topics, Prentice Hall Publication, 2006.

[Prigogine, 1961] I.Prigogine: Thermodynamics of Irreversible Processes, New York Interscience, 1961

[Tsallis, 1998] C. Tsallis: Possible generalization of Boltzmann-Gibbs statistics, Journal of Statistical Physics 52, 1998.

[Watts, 1998] D.J. Watts and S.H. Strogatz: Collective dynamics of „small-world" networks, Nature, 393, 1998.

Authors' Information

Franciszek Grabowski *– Editor, Department of Distributed Systems FECE RUT; W. Pola 2, 35-959 Rzeszów, Poland ; e-mail: fgrab@prz.edu.pl*

Marek Zarychta *– Editor, Higher Vocational State School in Jarosław; Czarnieckiego 16, 37-500 Jarosław, Poland ; e-mail:zarychta@pwszjar.edu.pl*

Przemysław Hawro *– Editor, Higher Vocational State School in Jarosław; Czarnieckiego 16, 37-500 Jarosław, Poland ; e-mail: przemyslaw@pwszjar.edu.*

THE ROLE OF MULTI-AGENT APPROACH IN BUILDING INFORMATION INFRASTRUCTURE FOR A MODERN COMPANY AND CARRYING OUT MANAGEMENT TASKS

Elena Serova

Abstract: *Companies applying information, communication technologies and computer simulation modeling tools are paying more and more attention to the ability to adaptive and hybrid architectures when building and developing information infrastructure. This paper studies an approach to forming an external information infrastructure for a company with a developed clientele by joining opportunities of CRM (Customer Relationship Management Systems) technologies and intelligent agents. Companies can succeed in carrying out their development strategies by following modern trends, adding intellectual information tools to CRM systems and creating an adaptive hybrid for external infrastructure based on the multi-agent approach.*

Keywords*: Modeling, Multi-Agent systems, Management, Customer Relationship Management systems.*

ACM Classification Keywords*: K. Computing Milieux - K.6 Management of Computing and Information Systems – K.6.4 System Management – Centralization/Decentralization*

Introduction

In modern Russia, using information and communication technologies (ICT) in management, including computer simulation modeling methods and tools, is a key driver of business efficiency. They do this by: helping improve quality of products and services, save labor and material costs, increase productivity, and improve production management. In industrially developed countries, despite a much higher degree of IT penetration in businesses, issues of choice and implementation of modern information systems and business simulation modeling tools that would meet the market demand and business strategies are also vital. Development of companies' external information infrastructure (II) and improvement of customer relations management (CRM) are growing in importance, with intellectual information technologies (IT) developing.

When solving business problems, simulation modeling tools should ensure mutual understanding at all organizational levels and bridge the gap between strategic vision and its implementation. One of the solutions is multi-agent systems (MAS) that have been developing rapidly in the last decade. Modern business simulation modeling tools use special software, programming languages and systems to develop models demonstrating structure of business processes, relations between people and room for optimization in the organizational structure as a whole.

This research studies the application of mobile agent technology to expand CRM possibilities for a growing clientele and comprehensive adaptive information infrastructure. Whereas a client-oriented approach is widely used as an element of business strategy, the agent concept is innovative and developing, and the very idea of integrating MAS and CRM methodologies is quite new.

This issue is undoubtedly important for more and more companies facing the necessity to improve customer relations management and recognizing their need for modern information and communication technologies and approaches.

This paper aims at analyzing applicability and adaptability of the joint CRM and MAS technologies, as well as the formation of comprehensive external information infrastructure for companies with growing clientele.

Thus, the research should study the concept of adaptive infrastructure, its modern technologies, and construction of adaptive infrastructure with multi-agent systems and customer relations management systems.

MAS for Management

Multi-agent systems as systems of distributed artificial intelligence have the following advantages:

- They speed up task fulfillment by parallelism and save the volume of data transmitted by passing high-level partial solutions to other agents;
- They are flexible by using agents of various capacity to carry out a task dynamically in cooperation;
- They are reliable by passing functions from agents unable to carry out a task to other ones.

MAS integration into a company structure can bring the following results:

- An information system specifically adapted to the enterprise's needs.
- More flexibility and ability to adapt to the external environment, especially in uncertainty.
- Ability to search and get unorthodox solutions.
- Confirmation of suppositions that previously lacked information.
- Faster decision-making when modeling negotiations.
- Finding and resolving potential conflicts of interests in both external and internal environment.
- More reliable decisions made owing to the agents' ability to pass functions to one another and redistribute responsibilities, which is not always possible in real life.
- Optimized access to information for all employees.

Major advantages of the multi-agent approach relate to the economic mechanisms of self-organization and evolution that become powerful efficiency drivers and ensure enterprise's stable development and prosperity [Chekinov G., Chekinov S., 2003]. Based on the multi-agent approach, a brand-new intellectual data analysis can be created, open and flexibly adaptive to solve problems, and can be deeply integrated in other systems.

The published feedback on MAS application shows that there are the following areas of agent application:

- distributed or network enterprise management;
- complex and multi-functional logistics;
- virtual organizations and Internet portals that sell products and services;
- academic management in distance-learning systems;
- companies with developed distribution and transportation networks (e.g., Procter&Gamble);
- distribution channels management;
- users' preferences simulation modeling (e.g., Ford).

Big companies can see advantages to the multi-agent approach such as: faster problem solving, less data transmission by passing high-level partial solutions to other agents, faster agreements and order placements.

Distributed companies find primary advantages in improved supply, supervision and coordination of remote divisions and structures. Companies with wide and quickly changing varieties of products can flexibly react to clients' changing preferences and foresee periodic changes. Service companies can preserve their experience of interaction and problem solutions with MAS technologies.

Integrating CRM and Multi-Agent Approach

Methodologies of the client-oriented approach to organization of company operations and the multi-agent approach can be integrated. In other words, CRM strategy can be carried out with multi-agent systems:

- To simulate and forecast clients' behavior, both returning and potential ones';
- To coordinate dealers and remote divisions with a multi-agent system;

- To automate and improve the Customer Support process within the CRM concept;
- To preserve knowledge and skills of marketing and sales specialists in the relevant agents' databases;
- To develop an integrated multi-agent Internet portal for agents to keep users' personal contents;
- To create a search agent to monitor outside information;
- To organize a distance-learning portal.

Here are principal provisions of the methodologies considered in this paper as well as major mechanisms for their implementation that can underlie integration of the approaches:

The CRM concept methodology provisions:

- Systemic approach to customer relationship management;
- Business strategy to efficiently manage customer relationship;
- Client identification, profiling and personalization;
- Assessment of clients and their needs with data analysis and sorting;
- Long-term customer relationship;
- Meeting client's needs;
- Cutting-edge management and information technologies to collect information about clients at every stage of their life-cycles;
- Automation of the three key divisions that are a principal interface between the enterprise and its clients: marketing, sales and service;
- New products based on customers' feedback;
- Every contact with clients fixed and stored in the contact history.

The CRM technology can be implemented with the following mechanisms:

- collecting and processing partners' data in a unified database;
- automating and controlling managers' work;
- timely analyzing efficiency of the enterprise.

The multi-agent approach methodology includes:

- Distributed artificial intelligence methods;
- Human or software agent's impact on environment.
- Program's ability to react to external events and choose relevant actions on its own;
- Forming action plans, forecasting environment changes;
- Social aspect in agents' behavior and their interaction within a multi-agent system;
- Opportunities to transfer data, knowledge, responsibilities and tasks;
- Systemic approach (agents are parts of a single system and carry out a single task);
- Decentralized data, access to them and agent management;
- Negotiation simulation modeling and finding an optimum solution from a conflict of interests.

Mechanisms used by the multi-agent approach include:

- databases on a certain area of life with models of primitive values and relations as well as analysis, learning and situational orientation algorithms;
- agents' cooperation, conflict of interests, economic cooperation methods;
- object-oriented approach;

- agent design standardization, special agent programming languages (e.g., the ACL group – Agent Communication Language).

Interestingly enough, experts note that CRM systems are most efficiently applied, among others, by high-tech and distribution companies. At the same time, distribution and new high-tech services are leading in using multi-agent systems.

Please note that the suggested agent approach is not the only one possible, principal or most efficient to build external II for a company specializing in software development and integration. Yet, it can help the company get the following advantages in management, strategy planning and development:

- more efficient database maintenance gives more clients;
- consistent users of the system get better service;
- lower technical support specialists' workload;
- distance-learning portal attracts more users.

Let us consider an example of a software company whose external infrastructure consists of the following blocks: potential clients; technical support and consultancy; education; sales and dealership. The company strives for long-term contracts, customer loyalty and client-oriented approach. It has a distributed client network all over Russia.

The basic stages to design a MAS for the company are:

1. *To formulate the mission (goal) for the MAS.*

MAS's goal is to build the company's external II to get certain advantages when working with the client network.

2. *To determine MAS agents' principal and additional functions.*

The company should analyze its operations and find the key ones to efficiently manage the client network. The following operations (components of key business processes) impact the client network management and are its major, most important part:

- Working with potential clients and customers;
- Keeping a database of potential and returning clients;
- Working with remote clients;
- Controlling dealers;
- Technical support and consultancy for users;
- Trainings;
- Online user support (keeping users' personal contents on the company's Internet portal);
- Efficiently measuring marketing actions.

3. *To specify agents and distribution of their functions.*

Upon the previous item, the following agents are needed to build a MAS:

- Incoming requests processing agent;
- Potential clients agent;
- Regular customers DB agent;
- Remote clients agent;
- Consulting agents;
- Distance learning agents;
- User agent on the Internet portal;
- Information monitoring search agent.

4. *To determine basic correlations (relations) between MAS agents*

Agents' basic correlations include:

- exchange of information received from external environment and acquired knowledge;

- information transfer from the potential clients agent to the regular customers DB and remote clients agents;
- distribution of incoming requests by the incoming requests processing agent among consulting, potential clients, regular customers DB and remote clients agents;
- distance learning agents' interaction;
- search agent's communication with Internet portal user agents—transfer of appropriate material (user agents create requests to the search agent).

5. *To define possible agents' actions (operations):*
   - communications;
   - interaction with users through a special interface;
   - interaction with users during distance learning;
   - cooperation and efficient distribution of information.
6. *To analyze current or potential changes in the external environment (functioning conditions).*

At present, the most efficient MAS scheme for the company under consideration is a system with an active human involvement. This will be more effective for coordination of actions, efficient management of the client network and better customer relationship is of primary importance. At the same time, such an important aspect as releasing employees' time is not critical, although the issue exists due to a high employee turnover. Among future modification of the MAS suggested, there may be widening the agent network and transferring some functions currently performed by humans to them.

Conclusion

Businesses and ITC are more and more closely interrelated. Like all over the world, IT in Russia is becoming a critical element of the product/service and profit generation chains. Companies' profits are growing with both cost and management optimization, and more clients brought by client-oriented strategies. It is an international II development trend that the CRM technology is getting more and more popular, and companies are allowing customers to form their requirements on their own. Yet, customer relationship management systems are most often built on standard solutions based on CRM modules as connection tools between the ERP system and external environment. CRM technologies in their essence cannot be a comprehensive tool to form information infrastructure of a company developing information interactions in the external environment. That is why it seems very interesting and perspective to study how external II can be built to organize comprehensive client networks with the widely used CRM technology along with a rare but perspective MAS approach. This area is as of yet understudied, however it is developing quickly.

It is no less important that the ability to adapt and hybrid architectures are becoming essential when building II. A company can successfully implement its development strategy by relying on these modern trends, adding intellectual information tools to the CRM, and creating a hybrid adaptive external infrastructure based on the multi-agent approach.

Once again, the suggested approach to building the external information infrastructure for a company with a developed client network, based on joint possibilities of the CRM technology and multi-agent approach, is neither the only one possible nor the most efficient one. Based on the methodologies and mechanisms of the system described, one can find ways to integrate them, although it would not be appropriate to presume the integration full.

The CRM and MAS technologies can be mutually complementary. Both of them offer a certain specific approach to structuring business operations rather than just automating certain single processes. Thus, multi-agent systems are a radical concept that starts an era of network organizations with intellectual robots' collective interaction by offering to switch from powerful centralized systems to fully decentralized ones, with hierarchical structure replaced with network organization, rigid bureaucratic "from top to bottom" management (based on

bosses' commands for subordinates) with negotiations, and planning with flexible agreements. As a result, production volumes, profitability, competitiveness and mobility are growing. The CRM technology also offers a flexible approach to building the whole company's business. The CRM methodology should not be considered just a concept of interaction with clients; it is rather a system that helps build a long-term client-oriented business.

Bibliography

[Chekinov G., Chekinov S., 2003] G. Chekinov; S. Chekinov. Primenenie tekhnologii multiagentnykh sistem dlya intellektualnoy podderzhki prinyatiya resheniy (Applying the multi-agent system technology to intellectual support of decision-making // *Sistemotekhnika*, No. 1, 2003.

[Gavrilova, Muromtsev, 2007] T.A.Gavrilova; D.I.Muromtsev. Intellektualne tekhnologii v menedzhmente: instrumenty i sistemy (Intellectual technologies in management: tools and systems). St. Petersburg: GSOM SpbSU, 2007.

[Gavrilova, Khoroshevskiy, 2001] T.A. Gavrilova; V.F. Khoroshevskiy. Bazy znaniy intellektualnykh sistem (Databases of intellectual systems). St. Petersburg: Piter, 2001.

[Gorodetsky, 1998]. V. Gorodetsky; M. Grushinsky; Khabalov. Mnogoagentnye sistemy. Obzor (Multi-agent systems. Review). *Novosti iskusstvennogo intellekta (Artificial Intelligence News)*. No. 2, 1998.

[Gorodetsky, 2008] V.Gorodetsky; O. Karsaev; O. Konyushy; V. Samoylov. Modelirovanie protsessov upravleniya vozdushnym dvizheniem na osnove mnogoagentnykh tekhnologiy (MAS-based simulation modeling for flight traffic management). *MSTU CA Scientific Journal, series "Navigation and Flight Traffic Management",* 2008.

[Katkalo, 2006] V. Katkalo. Evolutsia teorii strategicheskogo upravlenia (Strategic management theory evolution). St. Petersburg: St. Petersburg University Publishing House, 2006.

[Kazantsev, 2007] A. Kazantsev; E. Serova; L. Serova; E. Rudenko. Informatsionno-tekhnologicheskie resursy rossiyskoy ekonomiki (Information technology resources of Russia's economy). St. Petersburg: St. Petersburg University Publishing House, 2007.

[Kleyner, 2008] G. Kleyner. Strategiya predpriyatiya (Enterprise strategy). Moscow: Delo Publishers, 2008.

[Kouchtch, 2002]. S. Kouchtch. Setevoy podkhod v marketinge: rossiysky opyt (Network approach in marketing: Russia's experience). *SpbSU Vestnik, series "Management",* Vol. 1, No. 8, pp. 81-107, 2002.

[Pospelov, 1998] D. Pospelov. Mnogoagentnye sistemy – nastoyaschee i buduschee (Multi-agent systems – present and future). *Information Technologies and Computing Systems*, No. 1, 1998.

[Serova, 2009] E. Serova. Kontseptsia CRM i ee mesto v obshchey strategii kompanii (The CRM Concept and Its Role in the Corporate Strategy). *Papers of the 4st Workshop Conference "Modern Management: Problems and Perspectives"*. ENGECON, p. 301-306, 2009.

[Serova, 2008] E. Serova. Modern Methodological and Tool Approaches for Business-modeling tasks. *Information Technologies and Knowledge, Decision Making and Business Intelligence Strategies and Techniques*. Bulgaria, Sofia, Volume 2, 2008.

[Serova, 2006] E. Serova. Rasshirenie vozmozhnostey sistem upravleniya otnosheniyami s klientami na osnove Internet-tekhnologiy (Internet-extended functions of CRM systems). *Papers of the 1st Workshop Conference "Modern Management: Problems and Perspectives"*. ENGECON, p. 142-143, 2006.

Authors' Information

Elena Serova *– PhD (Economics), Deputy Chair, Dept. of IT in Management, Graduate School of Management, St. Petersburg State University; Per. Dekabristov 16, St. Peterburg 199155, Russia; e-mail: serovah@som.pu.ru; serovah75@gmail.com*

APPLYING THE CONCEPTS OF MULTI-AGENT APPROACH TO THE DISTRIBUTED AUTONOMOUS EXPLORATIONS

Vadim Golembo, Alexey Botchkaryov

***Abstract**: The problem of development and application of the multiagent techniques for organizing mobile exploring agents' cooperative behavior is examined. The main problems related to design and operating of distributed decentralized homogeneous multi-agent systems with restricted local interaction between agents are analyzed. A number of collective behavior models (algorithmic, interpolational, entropic), which discover the problems of distributed decentralized exploration are proposed.*

***Keywords**: multi-agent systems, distributed autonomous explorations, self-organization.*

***ACM Classification Keywords**: I.2.11 Distributed Artificial Intelligence*

Introduction

The basic scenario of distributed autonomous explorations is the following [[Ögren, Fiorelli, Leonard, 2004] - [Botchkaryov, Golembo, 2003]. The set of autonomous explorers is located in some environment. Then each of explorers starts gather information and transmits it to the center (the case of global user) or to some actuators located in the same environment (the case of local user). Together explorers form the system of distributed autonomous explorations, which main objective is to gather more precise and complete information by less cost. Main features of the distributed autonomous explorations' system are the following: 1) explorers are constructively and operationally autonomous; 2) system is spatially distributed; 3) explorers do in-situ measurements in environment; 4) system performs long-term explorations; 5) explorers are mobile. Examples of the systems of distributed autonomous explorations are 1) mobile wireless sensor networks [Ögren, Fiorelli, Leonard, 2004] - [Howard, Mataric, Sukhatme, 2002], 2) spatially distributed radar systems, 3) autonomous oceanographic sampling network [Turner, Turner, 1998] , [Curtin, Bellingham, Catipovic, Webb, 1993], etc.

The current state of the art in the area of organizing mobile explorers' cooperative behavior requires a fresh self-organization perspective look on the possibility to adapt the exploration instrument to the environment characteristics. Thus in our research work on distributed autonomous exploration systems in the Laboratory of Multiagent Systems at Computer Engineering Department of Lviv Polytechnic National University [Botchkaryov, Golembo, 2001] - [Botchkaryov, Golembo, 2003] we try to develop corresponding multiagent techniques. Here explorer station (sensor node, measuring device) is thought as mobile explorer (explorer agent) and corresponds to intelligent agent [Weiss, 2000] , [Wooldridge, 2002]. In addition, the distributed autonomous exploration system is thought as mobile explorers' team and corresponds to multiagent system [Weiss, 2000] , [Wooldridge, 2002] .

Problem statement

There are two main problems related to the operating of the distributed autonomous explorations' system. The first problem is the problem of placement. It can be described in the following way. The number of explorer agents located in environment (explored object space) is limited. This limitation naturally results from principle of minimization of exploration instrument's influence on explored object. Hence, explorer agents' team can obtain only the limited information about explored object at one moment of time. Based on this fact one can make the following statement: different placements of explorer agents in explored object space give us images with different amount of information about this object. Thus, the problem of placement arises: how one can place explorer agents in object space to achieve image with maximum amount of information?

The second problem is the problem of control. It can be described in the following way. The global (or local) user is usually remote from the explorer agents. The user also has no or has a little a priori information about processes in object under exploration. Hence, user cannot solve problem of placement precisely and in time. The quality of decisions generated by user will be always limited by uncertainty about real conditions of corresponding task. Based on this fact one can make the following statement: a user cannot eliminate the uncertainty because of its remoteness while the explorer agents potentially have such ability. Thus, the problem of control arises: how one can delegate the initiative in making rational decisions to the explorer agents?

Considering problem of placement and problem of control together, one can see that behavior of the explorer agents' team must be in some way strictly related to the processes in explored object. In other words we need to develop such a multiagent exploration system, which can autonomously (solving the problem of control) find the best according to the specified criteria way of exploration (solving the problem of placement). Here the most difficult case is non-linear dynamics of environmental processes (especially the so-called synergetic processes). Thus, we can state the following proposition: if synergetic environmental processes will be explored by self-organizing multiagent system, then we obtain the qualitative rise in autonomous explorations. Multiagent exploration system must be capable to assimilate "order" of environmental processes (this "order" in fact is the source of multiagent system's self-organization). In this way, multiagent exploration system obtains similar to the environmental processes dynamics eliminating corresponding uncertainty.

The main properties of multiagent exploration system under consideration are 1) homogeneity of explorer agents (all agents have the same structure (Fig.1) and embodiment and each agent perform the same set of control & AI algorithms), 2) decentralized control (each agent makes and implements decisions independently, the control center is absent), 3) local limited communication between explorer agents (each agent can detect and possibly communicate only with neighbor agents within limited detection range).

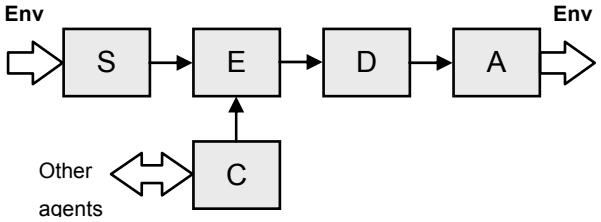


Figure 1. Functional structure of mobile explorer agent:

S – sense the environment state, C – communicate with other agents,
E – estimate the appropriateness of previous action, D – decide about next action, A – actuate decision

Under these conditions, we consider the main problem of multiagent systems' design: how one can transform the Global Utility Function of multiagent exploration system to local utility functions of agents? Or else how one can make the desired collective behavior emerges from individual agents' actions? In this case, the best way of exploration is equal to emergent collective behavior and corresponds to global extremum of the Global Utility Function (Fig.2).

Here the following problems arise. How one can decide which way of exploration is better? How one can estimate current success in finding proper way of exploration? How one can develop the effective algorithms of collective behavior of explorer agents?

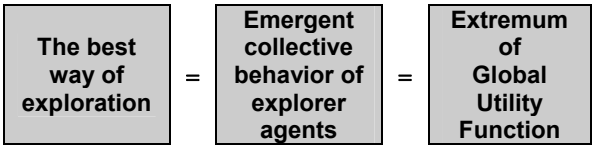


Figure 2. The common framework

Research and Development

Our approach (Fig.3) consists in implementing key problems (problem of control, problem of placement, and eventually problem of self-organization) in designed task environments (models of collective behavior) and developing the algorithms of collective behavior (collective decision making, multiagent reinforcement learning, autonomous exploration heuristics) in framework of these task environments [Botchkaryov, Golembo, 2001] - [Botchkaryov, Golembo, 2003], [Wooldridge, 2002] - [Botchkaryov, 2002], [Botchkaryov, 2005], [Botchkaryov, Golembo, 2005]. Here we use the following theories and methods: real-time search [Ishida, 1997], learning automata [Tsetlin, 1973], and reinforcement learning [Weiss, 2000] , [Sutton, Barto, 1998], [Kaelbling, Littman, Moore, 1996].

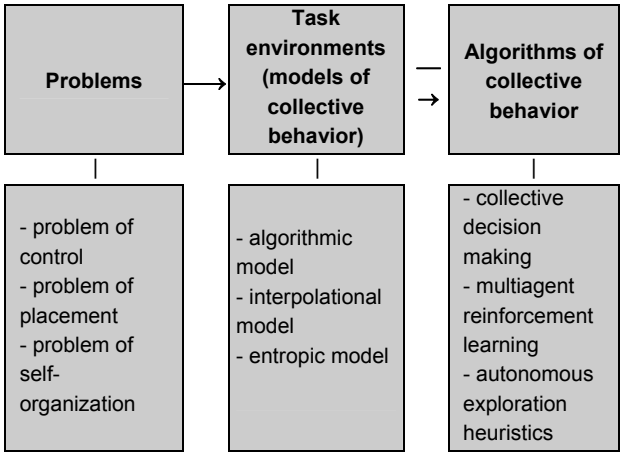


Figure 3. Research & development flow

Algorithmic model. Algorithmic model is used to find the appropriate methods of decentralized control (i.e. solutions to the problem of control) corresponding to the collective exploration specific. In the framework of this model procedures of decentralized collective measurements are developed. In algorithmic model, the problem of control of distributed autonomous explorations is interpreted in terms of load balancing problem [Botchkaryov, 2002] by analogy with algorithmic theory of measurements [Stahov, 1979]. Here each agent is thought as independent balance weight. The agent can change his individual weight in some range. The collective of agents must find the common weight equivalent to the unknown load (i.e. measured value). The agents do it step-by-step manner being reinforced by current difference between their common weight and weight of unknown load. Here the first question is solvability and the second question is quality of solution (minimization of number of steps). The key problem in this case is uncertainty about the actions of other agents. An agent cannot decide explicitly about appropriateness of his previous action because overall system reacts to the collective action. The analog problem from multiagent RL is credit assignment problem [Weiss, 2000]. Different techniques can be used to eliminate the uncertainty about other agents' actions. Even the most difficult case with lack of inter-agent communication is solvable. We develop the number of collective behavior algorithms based on learning automata [Tsetlin, 1973] and reinforcement learning techniques [Sutton, Barto, 1998], [Kaelbling, Littman, Moore, 1996]. We plan and conduct the number of numerical experiments with the algorithmic model [Botchkaryov, 2002]. The main results are the following: 1) speed of balancing increases with communication limitations decrease; 2) speed of balancing decreases with number of agent increase; 3) collective behavior algorithms based on reinforcement learning techniques shows the best results.

Interpolational model. Interpolational model is used to find the appropriate methods of autonomous explorations (i.e. solutions to the problem of placement). In the framework of this model procedures of autonomous

explorations are developed. Here the discrete environment with function f(X) of some parameter realized over points is considered. Each explorer agent can locate in one environment point, sense the function value in this point and report this value to the center (parameter value and point coordinates). The agents can move through the environment in any direction. The center builds the image F(X) of environment function using some interpolation method. Here explorer agents play the role of mobile interpolation nodes. Deviation between f(X) and F(X) can be estimated and taken as value of Global Utility Function of explorer agents' team. The key problem in this case is uncertainty about the environment function and other agents' actions. Thus agents must find the best placement in environment (to minimize deviation between f(X) and F(X)) or organize convergence of their movements to the best placements. At first, we develop relatively simple "relaxation" algorithms with low inter-agent communication rate [Wooldridge, 2002]. The more sophisticated approach to develop corresponding collective behavior algorithms is based on S-transform (Vallee-Poussin algorithm) and R-transform (R-algorithm) of placements (case of Chebyshev's interpolation) and reinforcement learning techniques [Botchkaryov, 2005] [Botchkaryov, Golembo, 2005] . Another approach is based on computer geometry methods, extremum search heuristics, and reinforcement learning techniques [Botchkaryov, Golembo, 2005]. We plan and conduct the number of numerical experiments with the interpolational model [Botchkaryov, Golembo, 2003], [Botchkaryov, 2005], [Botchkaryov, Golembo, 2005]. The main results are the following: 1) quality of explorations increases with communication limitations decrease; 2) collective behavior algorithms based on heuristics shows the best results.

Entropic model. Entropic model is used to find the correlation between process of self-organization in explorer agent's team and amount of information gathered in environment. Self-organization techniques are main subject of interest here. According to the model environment is a network of stationary event sources (nodes) with different Shannon entropy values (number of different events is limited and equal to all sources). Each agent can move through the network, observe events in the node where he is currently located and report this information (node id and event id) to the center. Center builds the statistical image of environment based on information gathered by agents. Initially center supposes all event sources have the maximum entropy. Thus if an event source has maximum entropy then an agent located in corresponding node gives no new information to the center. The key problem in this case is uncertainty about the real entropy values of nodes and other agents' actions. Under these conditions, agents must collectively decide about next placement over environment nodes. Here some heuristics based on the statistical methods can be used. In this framework, we developed collective behavior algorithms based on idea of probabilistic automata and reinforcement learning techniques [Botchkaryov, Golembo, 2005]. We plan and conduct the number of numerical experiments with the interpolational model [Botchkaryov, Golembo, 2005]. The main result is the following: dependence between self-organization quantitative parameter and amount of gathered information shows increase of self-organization quantitative parameter with increase of amount of gathered information.

Results and Future research

The main results of our work are 1) proposition to use self-organizing multiagent exploration system; 2) task environments (models of explorer agents' collective behavior); 3) Global Utility functions in the context of corresponding task environments (models of explorer agents' collective behavior); 4) algorithms of explorer agents' collective behavior; 5) research and development software.

Complementary outcomes of our work are 1) algorithms of collective formation (inter-agent detection); 2) self-synchronization methods for explorer agents' team; 3) self-organization of explorer agents' team in space; 4) algorithms of keeping the communication connectivity while moving in space; 5) modifications and new variants of reinforcement learning techniques.

Extra outcomes of our work are 1) model of explorer agents' collective behavior based on poly-probe exploration method; 2) game models of collective behavior with emphasis on exploration techniques; 3) intelligent agent architectures adapted to the collective exploration domain.

Our future research will cover the following topics: 1) considering the multipurpose exploration case (for example exploration of several functions in interpolational model); 2) integrating information obtained by multiagent exploration system with data of satellite monitoring (including case where explorer agents play role of control points for correcting data of satellite monitoring); 3) combination of different methods of explorations (point, tail, bearing, center-probe, poly-probe); 4) emergent languages of inter-agent communication (problem of finding most appropriate language for communication and explored object description during the process of exploration); 5) most common case: collective perception (including the task of collective visual exploration of unknown object using the image processing techniques).

Conclusion

Development of the systems of distributed autonomous explorations is an actual problem for many applications. Multiagent approach can be successfully used for this purpose. The problem of effective environment exploration is important in multiagent systems design too. Main problems related to the distributed autonomous explorations were analyzed. Two of them are critical: problem of placement (how agent' team can find autonomously the best way of exploration?) and problem of control (how one can organize team of explorer agents without centralized control?). The way to use self-organization principle is proposed. Three models of collective behavior (algorithmic, interpolational, and entropic) were developed. The algorithms of collective behavior of explorer agents in the framework of these models were developed.

Bibliography

[Ögren, Fiorelli, Leonard, 2004] P. Ögren, E. Fiorelli, N. E. Leonard. “Cooperative control of mobile sensor networks: adaptive gradient climbing in a distributed environment”, IEEE Transactions on Automatic Control, vol. 49, no. 8, pp. 1292–1302, 2004.

[Chung, Gupta, Burdick, Murray, 2004] T. H. Chung, V. Gupta, J. W. Burdick, R. M. Murray. "On a decentralized active sensing strategy using mobile sensor platforms in a network”, in IEEE Conf. on Decision and Control, (Paradise Island, Bahamas), pp. 1914-1919, 2004.

[Cortés, Martínez, Karatas, Bullo, 2004] J. Cortés, S. Martínez, T. Karatas, F. Bullo. "Coverage control for mobile sensing networks”, IEEE Transactions on Robotics and Automation, vol. 20, no. 2, pp. 243-255, 2004.

[Howard, Mataric, Sukhatme, 2002] A. Howard, M. J. Mataric, G. S. Sukhatme. "Mobile sensor network deployment using potential fields: A distributed scalable solution to the area coverage problem”, in International Conference on Distributed Autonomous Robotic Systems (DARS02), (Fukuoka, Japan), June 2002, pp. 299–308.

[Turner, Turner, 1998] R. M. Turner and E. H. Turner. "Organization and reorganization of autonomous oceanographic sampling networks”, Proceedings of the IEEE Int. Conf. on Robotics and Automation, Leuven, Belgium, pp. 2060-7, May 1998.

[Curtin, Bellingham, Catipovic, Webb, 1993] T. B. Curtin, J. G. Bellingham, J. Catipovic, D. Webb. "Autonomous oceanographic sampling networks”, Oceanography, Vol. 6, No. 3, pp. 86-94, 1993.

[Botchkaryov, Golembo, 2001] A. Botchkaryov, V. Golembo. "Distributed contact sensing system based on autonomous intelligent agents”, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, no. 437, pp.14-20, 2001. (Ukr.)

[Botchkaryov, Golembo, 2002] A. Botchkaryov, V. Golembo. "Models of collective behavior of measuring agents”, Transactions on Computer systems and networks, Lviv Polytechnic National University Press, no. 463, pp.19-27, 2002. (Ukr.)

[Botchkaryov, Golembo, 2003] A. Botchkaryov, V. Golembo. "Collective behavior of mobile agents solving distributed sensing task", Proceedings of the 1st International Conference "Advanced Computer Systems and Network: Design and Application" (ACSN'2003), Lviv, Ukraine, pp. 123-124, September 2003. (Ukr.)

[Weiss, 2000] G. Weiss (Editor). Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence, MIT Press, 2000.

[Wooldridge, 2002] M. Wooldridge. An Introduction to MultiAgent Systems, John Wiley & Sons, 2002.

[Botchkaryov, Golembo, 2003] A. Botchkaryov, V. Golembo. "Interpolational model of collective behavior of mobile agents of smart sensor network", Transactions on Computer systems and networks, Lviv Polytechnic National University Press, no. 492, pp.21-27, 2003. (Ukr.)

[Melnyk, Golembo, Botchkaryov, 2003] A. Melnyk, V. Golembo, A. Botchkaryov. "The new principles of designing configurable smart sensor networks based on intelligent agents", Transactions on Computer systems and networks, Lviv Polytechnic National University Press, no. 492, pp.100-107, 2003. (Ukr.)

[Botchkaryov, 2002] A. Botchkaryov, "Solving the problem of mechanical balancing by collective of mobile agents", Transactions on Computer systems and networks, Lviv Polytechnic National University Press, no. 463, pp.14-18, 2002. (Ukr.)

[Stahov, 1979] A.P. Stahov. Algorithmic theory of measurements, Moscow, Znanie, 1979. (Rus.)

[Botchkaryov, 2005] A. Botchkaryov. "The collective behavior of mobile intelligent agents solving the tasks of distributed autonomous explorations", Transactions on Computer systems and networks, Lviv Polytechnic National University Press, no. 546, pp.12-17, 2005. (Ukr.)

[Botchkaryov, Golembo, 2005] A. Botchkaryov, V. Golembo. "Self-organization of collective of mobile explorer agents solving the task of distributed in-situ sensing", Artificial Intelligence, Donetsk, no.3, pp. 723-731, 2005. (Rus.)

[Ishida, 1997] T.Ishida. Real-Time Search for Learning Autonomous Agents, Kluwer Academic Publishers, 1997.

[Tsetlin, 1973] M. L. Tsetlin. Automaton Theory and Modeling of Biological Systems. Academic Press, New York, 1973.

[Sutton, Barto, 1998] R. S. Sutton, A. G. Barto. Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998.

[Kaelbling, Littman, Moore, 1996] L.P. Kaelbling, M. L. Littman, A. W. Moore. "Reinforcement learning: A survey," Journal of AI Research, N4, 1996, pp. 237-285.

Authors' Information

Vadim Golembo *– Associate professor, Ph.D., Department of Computer Engineering, Lviv Polytechnic National University; Bandery Str., 12, 79035, Lviv, Ukraine; e-mail: golembo@polynet.lviv.ua*

Alexey Botchkaryov *– Senior lecturer, Department of Computer Engineering, Lviv Polytechnic National University; Bandery Str., 12, 79035, Lviv, Ukraine; e-mail: alb@polynet.lviv.ua*

Intelligent Manufacturing Systems

DESIGN CONCEPT OF INTELLIGENT MANAGEMENT SYSTEMS

Galina Setlak, Sławomir Pieczonka

Abstract: *In this paper a concept of designing and building intelligent decision support systems in production management is introduced. The new approach to the design of intelligent management systems is proposed based on integration of artificial intelligence technologies (fuzzy logic, artificial neural networks, expert systems and genetic algorithms) with exact methods and models of decisions search and simulation techniques. The proposed approach allows for creating intelligent decision support systems of complex, unstructured management problems in fuzzy conditions. The systems learn based on accumulated data and adapt to changes in operation conditions.*

Keywords*: artificial neural networks, fuzzy inference systems, classification, decision support system, intelligent manufacturing systems*

ACM Classification Keywords*: I. Computing Methodologies, I.2 Artificial Intelligence, J. Computer Applications J.6 Computer-Aided Engineering*

Introduction

Globalization of the world economy and growth of competition on the market impose increasingly greater demands on modern entrepreneurs. Currently, management and control of production enterprises is impossible without an application of appropriate tools supporting decision making at each stage of a company's functioning from designing through to product exploitation. CIM (Computer Integrated Manufacturing) Systems are an example of such available tools that enable composite automatization of technological and organizational preparation for manufacture, current supervision, technological process control, organization and management. The development of CIM Systems has, in recent years, been directed at applying the methods of artificial intelligence to support decision processes and production control as well as monitoring, simulation and technological process diagnosis. IM (Intelligent Manufacturing) [Chlebus, 2000], [Zuomin Dong, 1994], [Ladet, Vernadat, 1995] is the most recent idea in the development of automatization and computer integration of production systems. According to the definition given in [Chlebus, 2000], Intelligent Manufacturing is:

> "a set of methods, procedures and Cax tools (eg. CAD, CAP, CAM) equipped with artificial intelligence tools and supporting designing, planning and manufacturing."

The following, among others, are the basic constituent elements of Intelligent Manufacturing Systems mentioned in publications:

- intelligent machines and tools, i.e. numerically controlled machines and robots,
- intelligent manufacturing systems, and
- intelligent management systems.

The concept of intelligent manufacturing combines the ability of decision-making support systems in generative systems to obtain knowledge, to learn and to adapt to a changing environment and to the actual arrangement of system components. The nature of intelligent manufacturing is system's possibility to learn and its self-development as well as the possibility to generate information necessary to control the integrated production system.

The aim of the work is to present a conception of designing and creating one of the components of Intelligent Management Systems as Intelligent Decision Support Systems alongside with the basic methods and tools of artificial intelligence which are necessary to support decision making in these systems.

Basic Assumptions Behind the Conception of Designing Intelligent Management System

Intelligent Management System of a production enterprise denotes an information system which provides necessary information, enables its analysis and use of analytical and simulation - based decision making models in order to assist decision making at each stage of decision process, as well as it is capable of learning and adapting to the dynamically changing environment and the current arrangement of system components. In other words, it is a decision supporting system based on the applied methods and tools of artificial intelligence able to solve complex decision problems, semistructurised or non-structurised, requiring the processing of incomplete, unreliable, contradictory, or difficult to formalize knowledge.

The demands towards Intelligent Management Systems in a production enterprise are as follows:

- A possibility of collecting and processing different types of information from all sources, both internal and external, in order to acquire and model knowledge necessary to make decisions at all levels of decision process in an enterprise. At the same time a possibility of modeling knowledge and processes, based on human thinking, is required.
- In a decision process, at decision selecting, the decision maker's subjective evaluation based on his experience and intuition should be taken into account in IMS.
- There should be a possibility of preliminary information handling and analysis with analytical methods as well as modern artificial intelligence technologies.
- A possibility of detecting emergency and critical situations and of prompt reaction to them. There must be a possibility of situational data analysis in real time, necessary in an emergency inside the production system or in its surroundings.
- A possibility to allow for complexity and comprehensiveness of decision-making issues in strategic management support.
- Taking into account the lack of stability and change dynamics, both in the surroundings and inside the enterprise, the IMSs under design should have the capability for learning from experience and adapting the experience to intensive alteration of working conditions.

In the conditions of modern economy, information and data bases which were fundamental for integration processes of design, planning and production control within CIM, are not sufficient; what is essential is a broadly understood knowledge of all these processes. Data and information collection, modeling and processing, currently evolve into a process of knowledge gaining, modeling and processing. The concept of knowledge is not clearly defined in the literature. Knowledge does not only encompass very extensive and dispersed resources of different types of information, but also, and above all, it is a complex structure of links between pieces of information and it involves information that is difficult to formalize. Experience, qualifications, human intuition, and models of different processes (including discreet, dynamic and stochastic processes) are all knowledge. Therefore, the basic demand that an intelligent management support system should meet is the necessity for

collecting and processing all types of knowledge from all sources, external and internal, in order to gain and model the knowledge necessary to make optimum decisions.

Considering the above, what should be the base in methodology of intelligent systems design is an approach that would address knowledge management issues and would combine the existing manufacture engineering systems and subsystems with artificial intelligence technologies in order to create integrated environment for comprehensive decision-making solutions.

Class ERP (Enterprise Resource Planning) information systems created in the late 1990s work in all enterprise management areas: storage, production, finances (including management accounting), distribution, transport and servicing. Additionally, they enable firms to cooperate with clients and partners effectively and in a modern way. Currently, integrated systems assisting class ERP economic activity are commonly introduced, chiefly in large production enterprises. In recent years, more and more integrated software packages (from e.g. Cognos, Connect Distribution, Comarch, Microsoft, Oracle, Sybase, and SAP) have been appearing, both in the world and on the home market, including packages that apply Business Intelligence technologies. Systems based on Business Intelligence technologies are used to collect information, mainly in data warehouses, and to analyze the data to support decision making in business processes. BI technologies enable users to conduct a detailed data analysis with the aid of different types of analytical tools (e.g. OLAP or data mining). The effective use of Business Intelligence tools is very much dependent on data warehouses creation, which makes it possible to unify and relate the data collected in various information systems of an enterprise.

It needs to be emphasized here that both class ERP systems and systems with Business Intelligence technologies, despite many advantages, do not solve all the problems related to knowledge processing and modeling such as, for instance, those pertaining to linguistic data or the natural language.

The analysis of the existing methods and approaches to the creation of enterprise management support systems indicate that data warehouses currently constitute the basis for modern information systems and guarantee effective use of the information included. Thus, implemented ERP systems and systems with Business Intelligence technologies can be utilized as sources of knowledge while creating intelligent management system of an enterprise.

IMS Design Approach

The suggested frame of IMS design uses the basic rule of object methods in which modeling of information and processes is concurrent. It is also assumed that the design process is based on a single conceptual category of ‘an object’. The frame works on some assumptions made in the methodology of design and implementation of open systems for computer integrated enterprises CIMOSA (Open Systems Architecture for CIM). CIMOSA was developed by AMICE (European Computer Integrated Manufacturing Architecture) Syndicate within European Union ESPRIT research projects in the years 1986-1996 [CIMOSA Association, 1996], [Ladet, Vernadat, 1995]. CIMOSA introduced an integrated enterprise modeling methodology based on processes.

In CIMOSA architecture an enterprise is defined as a set of domains consisting of a set of organizational processes realizing elementary objectives of the enterprise activity. In the approach presented an assumption was made that the domain may be defined for one or several spheres of the enterprise activity, implementing one or several management functions, while the enterprise organizational processes are linked to the market surroundings.

Based on the problem analysis conducted and the existing methods, and taking into account the afore listed demands from IMS, the following basic rules for designing intelligent systems for decision-support in management can be formulated:

1) Intelligent Management Systems should be created as open architecture systems of modular structure that would allow for their evolution, thus extending a range of possibilities in both the design process and the functioning.

   The rule will also secure, in IMS, the use of different methods of modeling and processing of knowledge gained from different sources as well as updating the knowledge in real time. The open modular architecture enables IMS to quickly adapt to the changing environment and according to the condition of the enterprise components.

The Intelligent Management system can be described as a distributed system in the following way:

$$IMS = <M, R(M), F(M), F(IMS)>,$$

where

M={Mi} – set of formal or logic-linguistic models performing specific intelligent functions;

R(M) – choice function of required models (set of models) in a given current situation;

F(M)={F(Mi)} – set of model modification functions;

F(IMS) – modification function of IMS and its basic structural elements M, R(M), F(M).

2) The designed IMSs should be made into adaptable systems and capable of learning from experience. A designed IMS should be able to aid the organization of a production process at any given time. The reconstruction of knowledge models and decision processes should be conducted immediately every time the need for process reorganization arises due to changes in the system. Learning systems design will ensure effective system functioning in real complex conditions.
3) An IMS of distributed architecture should be developed to be equipped with possibility of parallel information processing in order to increase its efficiency in collecting and exchanging a large amount of information, its analysis as well as in making group decisions. It is essential to utilize contemporary information technologies including local and global computer networks.
4) Widespread application of multimedia technology, computer graphics and hypertext in information presentation.

Conceptual Model of Intelligent Management System

In the presented approach to IMS design it is recommended that IMS's should be created as open architecture and modular structure systems, enabling the application of several methods of knowledge presentation and the integration of different knowledge processing schemes in the inference process, as well as the application of several learning methods. In intelligent manufacturing systems, the following selected contemporary methods and techniques of knowledge and decision process modeling should be integrated:

- Artificial neural networks – the most fascinating tool of artificial intelligence, capable of modeling extremely complex functions and, to some extent, copying the learning activity in the human brain.
- Fuzzy logic – technologies and methods of natural language formalization, linguistic and quality knowledge processing and fuzzification.
- Genetic algorithms and methods of evolutionary modeling – learning algorithms based on theoretical achievements of the theory of evolution, enriching the artificial intelligence techniques above.

The combination of these tools, in which knowledge is represented symbolically, with the traditional expert system will make it possible to create complex programmatic tools for solving difficult decision-making problems at each stage of enterprise functioning.

The conceptual structure of Intelligent Management System is illustrated in fig. 1.

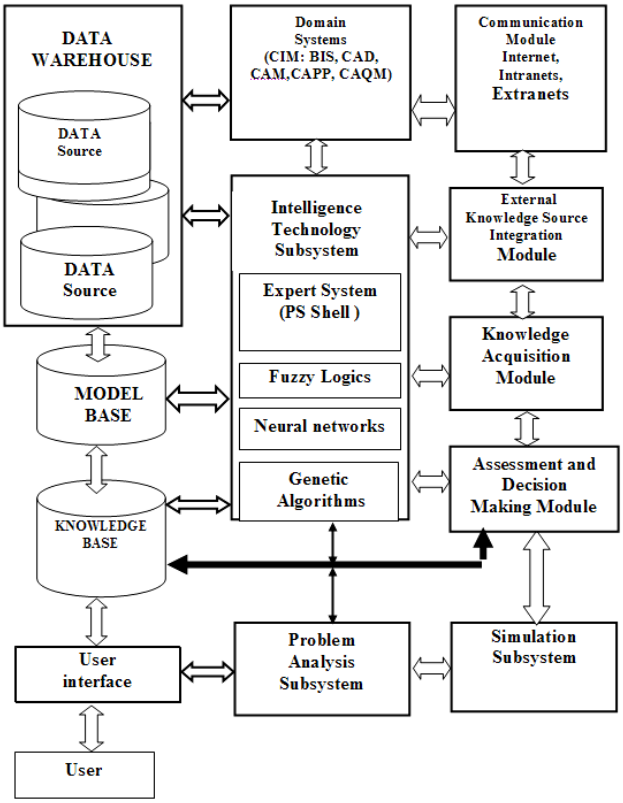


Figure 1. A conceptual structure of Intelligent Management System

The structure of IMS consists of the following subsystems and modules:

1) Modules which are very common in Intelligent Systems architecture
   - Databases and knowledge base
   - Inference Engine
   - Knowledge Acquisition Module
   - Explanation Engine
   - User Interface (graphical or command line; preferably with natural language understanding functionality)
2) The subsystems of Intelligent Technologies
   - Expert subsystem (eg. Expert System Shell)
   - Fuzzy Logic Subsystem
   - Neural Networks Subsystem (eg. NeuroSolutions or Statistica Neural Networks)
   - Genetic Algorithms Module (eg. Genetic Library)
   - AI Technologies Integration module
3) Domain-oriented Information Systems
   - Management Information System (MIS)

- Computer Aided Design System (CAD)
- Computer Aided Process Planning System (CAPP)
- Computer Aided Quality Management System (CAQM)
- Management Support System (MSS)

4) Problem Analysis Module
   - Problem Simulation Module
   - Monitoring Module
5) Decision selection and assessment subsystem. It includes practically verified analytical procedures as well as problem solving probabilistic methods that turned out to be very effective in several real applications.
6) Communication subsystem. It is based on computer networks technologies and Internet infrastructure that enable to acquire information from external sources.

Expert systems enabling decision support in product design and production process, planning and production control as well as management have been widely used for several years in integrated production systems CIM [Kacprzyk, 2001], [Ladet, Vernadat (ed.), 1995]. The classic expert systems are based on symbolic representation and knowledge processing. The symbolic processing of knowledge is a characteristic feature of most expert systems that are known [Kisielnicki, Sroka,1999]. In such expert systems, the following kinds of knowledge base can be indicated: database, rule base, text base, model base, and common sense knowledge base. The symbolic representation of knowledge had prevailed in research until very promising results were achieved in a research on artificial neural networks by Rumelhart and McMlelland (published 1986). What can be observed since that time is an intensive development of the application of these most fascinating artificial intelligence tools to solve very difficult problems in different areas, including monitoring, controlling and supervision of technological processes for manufacturing. Artificial neural networks are modern computational systems parallel processing data and, above all, having the capability of adaptation and learning and a high defect and fault tolerance.

Knowledge processing in neural networks is conducted dynamically [Mulawka, 1996]. This basic quality of neural networks has been widely used by researchers for over twenty years to create so called hybrid expert systems. In the research here presented, neural networks are used to classify and cluster data for marketing analysis [Setlak 2004], [Setlak, 2001] as well as to classify constructional and technological modules; they are, then, the basic tool for data exploration, such as data mining, etc.

Hybrid expert systems were created as a result of integrating traditional expert systems, artificial neural networks, genetic algorithms and fuzzy systems in different configurations [Rutkowska, Piliński, Rutkowski, 1997]. In such systems, thanks to neural networks, the process of knowledge gaining, which is, admittedly, the most difficult stage of creating an expert system, is largely facilitated (instead of loading rules, the neural network is trained based on a representative pattern). Fuzzy systems are built using the fuzzy set theory and fuzzy logic [Kacprzyk, 2001], [Zielińskiego, 2000], in which dependencies are expressed as fuzzy rules of the 'IF/THEN' type and which include linguistic and qualitative variables. The possibility of processing, by fuzzy systems, qualitative knowledge representing information which cannot be precisely described using traditional methods of mathematics or binary logic, enables users to use these tools in intelligent decision support systems. As any artificial intelligence method, fuzzy logic is an attempt to describe reality in a way akin to human reasoning. In fuzzy systems knowledge is stored in structures of symbolic character. Fuzzy reasoning can be easily implemented as a neuron-like numerical procedure, which made it possible to develop many effective methods of teaching fuzzy systems.

Consequently, the theories of fuzzy sets and that of fuzzy logic has become an indispensable tool in designing hybrid intelligent systems, including those supporting management.

The term of evolutionary algorithms refers to computational systems of problem solving, which work based on the rules observable in natural evolution of living organisms [Rutkowska, Piliński, Rutkowski, 1997]. To evolutionary algorithms also belong such methods as genetic algorithms, evolutionary programming and evolutionary strategies. However, due to the limited scope of this paper, the differences between them were not addressed here.

The idea of genetic algorithms is based on processes observable in nature, such as natural selection, evolution of species, mechanisms of reproduction and heredity. Simplicity and versatility are two important positive features of evolutionary algorithms. In IMSs, genetic algorithms are successfully used, especially in solving optimization tasks based on one or many criteria. Apart from that, genetic algorithms are used to find the most effective neural network structure or neural-fuzzy system and as procedures teaching neuron-like structure. As a result of the research, genetic algorithms were recognized as very effective tools for solving complex practical optimization tasks including NP-hard problems such as scheduling manufacturing tasks.

In intelligent systems supporting management of integrated enterprises, simulation modeling methods must be used to solve various problems, both at the stage of designing and that of production system functioning. Petri Net model is currently the universal tool for modeling production processes.

Conclusions

In the work, a certain conception of designing intelligent systems for enterprise management was presented. Based on the conception, a methodology of creating the IMS is being developed based on the integration of artificial intelligence technologies with exact methods, well-known in the decision making theory, as well as with simulation modeling methods. The approach proposed will open up a possibility to build an IMS of open structure, combining existing information systems with the information sub-systems in production engineering using artificial intelligence technologies in order to create an integrated environment for comprehensive solving of decision-making problems in the system of intelligent manufacturing.

Intelligent manufacturing is the most promising and future-oriented of production system developments aiming at further automatization, optimization and integration of manufacturing processes. Intelligent processing of data from different sources will enable the integration of the data as well as strategic analyses and correct decision making. Flexible analysis, diagnosis and reporting scenarios will make it possible to create a basis for evaluation of the current situation and to facilitate and accelerate decision processes at each stage of the enterprise functioning. Fully integrated intelligent manufacturing systems will, in the near future, enable enterprises to function effectively and reliably on the global market, complying with its rising requirements.

Bibliography

[Chlebus, 2000] Chlebus E. (2000) Techniki komputerowe CAx w inżynierii produkcji, WNT, Warszawa.

[CIMOSA Association, 1996] CIMOSA Association (1996): CIMOSA Technical Baseline. Boblingen, Germany, Stockholmerstr.7, D-70731.

[Zuomin Dong (ed.), 1994] Zuomin Dong (ed.) (1994) Artificial Intelligence in optimal design and manufacturing, PTR Prentice Hall, Englewood Cliffs, New Jersey.

[Kacprzyk, 2001] Kacprzyk J.(2001) Wieloetapowe sterowanie rozmyte, WNT, Warszawa, s.416

[Kisielnicki, Sroka H.,1999] Kisielnicki J., Sroka H.(1999) Systemy informatyczne biznesu, Agencja wyd. Placet, Warszawa, s. 320.

[Kusiak, 1990] Kusiak A. (1990) Intelligent Manufacturing Systems, Prentice Hall, Englewood Cliffs, New Jersey, 532p.

[Ladet, Vernadat F.B.(ed.), 1995] Ladet P., Vernadat F.B.(ed.) (1995) Integrated Manufacturing Systems Engineering, Chapman & Hall, London.

[Mulawka, 1996] Mulawka J.(1996) Systemy ekspertowe", WNT, Warszawa, s.235.

[Rutkowska, Piliński, Rutkowski, 1997] D. Rutkowska, M. Piliński, L. Rutkowski (1997) Sieci neuronowe, algorytmy genetyczne i systemy rozmyte, PWN, Warszawa-Łódż ,.

[Setlak, 2004] G.Setlak (2004) Inteligentne systemy wspomagania decyzji, Wyd. LOGOS, Kijów, 252 str.

[Setlak, 2000] Setlak G. (2001) Neural networks in the Intelligent Information Systems of Management, Journal of Automation and Information Sciences, Kiev, 2000, N1, pp. 112-119.

[Setlak, 2001] G.Setlak (2000) Fuzzy-Neuro System for Decision-making in Management, Proceedings of the 7th IEEE International Conference on Methods and Models in Automation and Robotics, Wyd. Uczelniane Politechniki Szczecińskiej, Szczecin, 2001, Vol.1, p. 505-510.

[Zieliński (ED.), 2000] J. Zielińskiego (ED.) (2000) Inteligentne systemy w zarządzaniu, PWN, Warszawa, s.348.

Authors' Information

Galina Setlak, *D.Sc, Ph.D., Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science , W. Pola 2 Rzeszow 35-959, Poland, and The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland, e-mail: gsetlak@prz.edu.pl*

Sławomir Pieczonka *- PhD Student, M.A. The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland: e-mail: slawomir.pieczonka@pwszjar.edu.pl*

INTELLIGENT SYSTEM FOR ASSEMBLY PROCESS PLANNING

Galina Setlak, Tomasz Kożak, Monika Piróg-Mazur, Wioletta Szajnar

***Abstract**: The flexibility, customization and localization offered by computer integrated manufacturings are attractive but generate a new class of management problems. Intelligent support systems are needed for the managing: planning, implementation, configuring and operating of intelligence manufacturing systems. This paper presents a project of intelligent system for assembly process planning. The paper includes a coincidence description of the chosen aspects of implementation of this intelligent system using technologies of artificial intelligence (neural networks, fuzzy logic, expert systems, neuro-fuzzy systems).*

***Keywords**: Artificial intelligence, flexible assembly systems, neural networks, fuzzy logic, group technology formatting rules.*

***ACM Classification Keywords**: I. Computing Methodologies, I.2 Artificial Intelligence, J. Computer Applications, J.6.Computer Aided Engineering*

Introduction

Computer integrated manufacturing (CIM) provides manufacturing industry with the means to produce a variety of products efficiently. Effective planning, scheduling and control in the CIM environment depend largely on proper design of the decision support system (DSS). A manufacturing system is driven by input stimuli from the market in the form of direct product demand, market conditions and feedback on production with a variety of information perspectives. The activities of this system may be broadly classified as management (including strategic planning), design, production planning and production operation.

The planning of the modern manufacturing systems is a very complicated and responsible task. It assumed that the modern assembly systems are universal enough to be able to connect a high production capacity with the small quantities of production lots and short cycle time. It should ensure a production under the conditions of dynamical and sudden changes of the product range, the planed fixed dates for order realization and also the possibility of fast introducing of product design change into production [Boothroyd, Knight, 1994]. According to the opinions of many assembling specialists the module assembly engineering is the fundamental and most promising direction of the development of the modern assembly technology [Grabmeier, 2002], [Szabajkowicz, 1998]. The module technology is based on rules of the group production technology, which dominated in last dozen or so years, it improves and develops it [Szabajkowicz, 1998]. In addition, the module assembly technology enables a production adjusting according to market requirements, an easy adjusting oft the assembly system to every change of the product design, adding new engineering assembly modules. The planning of the flexible assembly modules in accordance with the modular engineering is the NP-hard problem at the centers for research and science as well as in design offices of the leading production companies in the last years.

This paper presents the application of the intelligent system Computer Aided Assembly Process Planning (CAAPP) for aiding of the module assembly technology planning, which was in [Setlak,1999] described. The CAAPP was developed in order to aid the decision making in the designing and functioning of the flexible assembly systems. In order to fulfill the identification and clustering tasks for product, parts and assembly unit groups, additional program modules are used, which include Self Organizing Map (SOM) of Kohonen and neuro-fuzzy systems.

The approach of the modules assembly technologies

The module assembling engineering consists in presenting of the production process as a set of technological modules. The technological module is considered as a structurally closed part of the processing, which conforms to the functionality, integrity and universality requirements. The module assembly means, that the assembly system has a modular structure and each module realizes a defined function or a limited function range, which are part of a general assembly process. According to the definition [Szabajkowicz, 1998] a technological assembly model composes “an integral set of the main and auxiliary activities of assembling, which are realized in a defined sequence at one station and uses a defined tool set for connecting of surfaces, parts, subassemblies, assemblies”. The connection of the elementary technological modules lies in a proper development and selection of technological modules. Each of them realizes a proper design module of construction.

During the planning of the flexible assembly systems with the modular assembly engineering the following stages can be selected:

- Analysis of the construction of the assembled product and the assembling technologies.
- Identification and classification of objects into groups and subgroups of the processed parts and (technological similar) assembly sets. The working out of a typical flow chart (based on common assembly sequences, similar to the manipulation activities, duration, etc.).
- Separation of autonomic, integrated assembly activities from the flow charts, then assembling the separated assembly units into groups depending on equipment with instrumentation to carry out these operations.
- Planning of structures and functions of the constructional modules.
- Preliminary planning of elementary technological modules.
- Assembling of the elementary modules and selection of proper, possible variants of the technological and constructional modules.
- Optimization of the technological module structure and the structure of the constructional module realized.
- Clustering of the elementary technological assembly modules.
- Final planning of the technological assembly modules, the modular technological complexes and of the corresponding constructional modules.

The analyses of the construction of the assembled product concerns first off all the analysis of a producibility for the product construction, which is in the present generally made using the DFA methodology (Design For Assembly). The analysis of the producibility for a construction must be carried out in order to simplify the product constructions, reducing the part forms and subassemblies number. The questions concerning the producibility of product constructions assembled automatically were investigated among other in [Łunarski,1993]. In this work the fundamental quality and quantity characteristics for producibility of product constructions for automatically assembling are presented (these are such features, as: interchangeability, regulation possibility, easy controlling and tool accessibility etc.). Planning products for assembly using the modular technology the constructional product modularization principle is to be kept. That means that by planning of units, subassemblies and parts following steps must be taken:

- Identification, separation of parts and basic surfaces;
- Use of typical assembly diagrams and methods;
- Aspiration to adjust a new product to such a construction, that the existing constructional modules and technological modules can be used.

By working out the expert system for modular assembly aiding system planning the necessity of integration of the constructional planning process with the processing planning was taken into consideration in order to utilize better the existing production equipment and eventually expansion or modernizing of it.

Concept of intelligent system for assembly process planning

The intelligent system Computer Aided Assembly Process Planning (CAAPP), which was described in [Setlak,1999], uses PC-Shell 4.0. – domain independent expert system shell, having strong hybrid properties. The PC-Shell has been implemented in Artificial Intelligence Laboratory (AITECH, Katowice). The PC-Shell 4.0 system integrates the expert systems shell using blackboard architecture elements and the simulator of the neural network. It assures the knowledge representation as declarative expressed rules, facts and distributing knowledge in the neural network. The expert knowledge can contain in some knowledge sources. A concept of model of the system CAAPP for aiding of assembly module planning was shown in [Setlak, 2008].

For aiding of the planning of the modular assembly technologies and to solve the problems of identification and classification of products groups, parts and units, the program modules have been developed, which complete the expert system CAAPP. These are the program module KLASGRUP and the module PKTMT. The program module KLASGRUP includes all procedures, which are necessary to carry out the constructional analysis of the planned or modernized product, and procedures to clustering the processed parts and (technological similar) assembly units of the mounted parts in order to separate and work out the constructional modules. The module PKTMT contains procedures for classification and grouping of the technological assembly modules. The details of the working out and structure of the expert system CAAPP, which technologically aids the assembly production preparation, are shown in the work [Setlak,1999].

To realize and test the program modules form aiding of the planning of the flexible assembly systems using the modular assembly engineering the knowledge base must be completed with following data:

- typical constructions;
- constructional features of the product parts;
- typical assembling flow charts and assembly methods;
- machinery data, technical equipment of the production system data;
- production costs for representatives of products from technologically similar groups.

In form of algorithms the constructional product analysis methods and assembly technology are formalized. The intelligent system CAAPP has been expanded by two additional modules; in addition a user interface has been introduced, which enables a presentation of a quality, verbal information in form of referring to adequate primary fuzzy sets. It enables a use of fuzzy inference engine in the program modules KLASGRUP and PKTMT the neural networks are used to classify the assembly parts and group the products.

Application of neural networks and neuro-fuzzy system for classification of assembly parts and units

The application of the clustering procedure can be classified into one of the following techniques [Jang, Sun, Mizutani, 1997]:

- hierarchical form trees in which the leaves represent particular objects, and the nodes represent their groups. The higher level concentrations include the lower level concentrations. In terms of hierarchical methods, depending on the technique of creating hierarchy classes (agglomerative methods and divisive methods);

- graph-theoretic clustering,
- fuzzy clustering,
- methods based on evolutionary methods,
- methods based on artificial neural networks.

The basic algorithms of the classification methods of machine elements are presented in [Ramachandran, 1991], [Ed. by Knosala, 2002], [Zolghadri Jahromi, Taheri, 2008].

Neural networks are widely used as classifiers; see e.g. [Jang, Sun, Mizutani, 1997], [Moon, Divers, 1998]. Classification and clustering problems has been addressed in many problems and by researchers in many disciplines like statistics, machine learning, and data bases. The basic algorithms of the classification methods are presented in [Nauck, Klawonn, Kruse, 1997], [Setlak, 2004]. In the literature various classification methods have been proposed (see e.g. [Grabmeier, Rudolph, 2002],).

In this work two approaches have been applied to clustering of parts and assembly units. As basic method it was used Self Organizing Map (SOM), which were introduced by T. Kohonen in the early '80s. It is a class of unsupervised learning neural networks, to perform direct clustering of parts families and assembly units. This type of neural network is usually a two-dimensional lattice of neurons all of which have a reference model weight vector (is shown in Fig. 1). SOM are very well suited to organize and visualize complex data in a two dimensional display, and by the same effect, to create abstractions or clusters of that data. Therefore neural networks of Kohonen are frequently used in data exploration applications as well [Kohonen, 1990], [Takagi, 2000]. The SOM can learn to recognize clusters of data, and can also relate similar classes to each other. SOM networks can also be used for classification when output classes are immediately available - the advantage in this case is their ability to highlight similarities between classes. SOM have been applied to classification of machine elements in group technology [Malave,1992], [Ed. by Knosala, 2002], [Setlak , 2004], [Setlak, 2008].

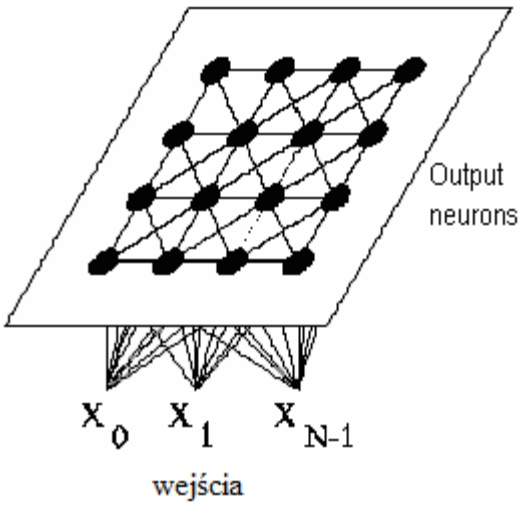


Figure 1. Self Organization Maps (Kohonen Networks)

The training of the SOM is achieved through a competitive learning process which consists of two steps that are applied iteratively.

- In the first step each input vector is compared to all the neurons' codebook vectors. The neuron s that has its codebook vector at the shortest geometric distance to an input vector becomes the winner for that input vector.

- In the second step, each winning neuron and its surrounding neurons, i.e., neurons within a neighbourhood N_s gradually change the value of their codebook vectors in an attempt to match the input vector for which it has won.

This cycle of competition and learning processes is repeated. At each cycle the size of the neighborhood of the winning neuron is decreased. The whole process terminates when each codebook vector has reached a satisfactory approximation of their corresponding input vector.

In the examples below the presented algorithm have been used the method of geometrical description of the units of machine engines described in [Knosala,2002].

Geometrical features of structural elements were presented in the form of the matrix of properties. This method consists in exploiting geometrical primitive conditions which basic geometric features of similar are describing. Next made coding of geometrical features which consists in using wood is B-Rep method in order to receive the structure of the model in the three-dimensional space (3D). As a result of the division of the model of the element in three dimensions with the determined resolution to layers a matrix image of the element is received.

The format of input data is being presented as follows:

<x> <y> <z>

<nr element> <nr layer > < the number of layers>

$<x_{11}>$ $<x_{12}>$ … $<x_{1n}>$

….

$<x_{n1}>$ $<x_{n2}>$ … $<x_{nn}>$

Where three first values means the resolution of the division of the 3D element into classes.

Grouped elements were written in the digital form at the 16x16x16 division in harmony with the accepted accuracy of the description of elements. The training data set includes 16x16X16 data items. The Kohonen neural networks composed of 16 neurons.

The other approach applies fuzzy logic and neuro-fuzzy systems for classification of parts and assembly units. However, neural networks work as a "black box", which means that they produce classification results but do not explain their performance. Thus, we do not know the rules of classification. Neural network weights have no physical interpretation. Fuzzy and fuzzy-neural systems can be employed in order to solve classification problems [Rutkowska, 2002]. Some of the major woks in this area are ANFIS (Adaptive Neuro-Fuzzy Inference System, [Jang 1997]), NEFCLASS (Neuro-Fuzzy CLASSification system, [Nauck, Klawonn, Kruse, 1997]), CANFIS (Co-Active Neuro-Fuzzy Inference System), [Lin, 1996].

A neuro-fuzzy systems for classification of parts and assembly units can be presented and is shown in Fig.2.

It is connectionist multi-layer architectures of neuro-fuzzy systems. The neuro-fuzzy systems are rule-based systems that realize fuzzy IF-THEN rules, described as follows:

$$R^{(k)}: \text{IF } x_1 \text{ is } A_1^k \text{ and } x_2 \text{ is } A_2^k \text{ and. .... and } x_n \text{ is } A_n^k \text{ THEN (y is } B_l), \quad (1)$$

where $X = (x_1, x_2, \ldots, x_n)^T$, x_i and $y \in Y \subset R$ for $i=1,2,...,n$ - are linguistic variables, $A_i^k \subseteq X_i \subset R$ $(i=1,2,...,n)$ are fuzzy sets characterized by membership function $\mu_{A_i^k}(x)$, B_l -for $l=1,2,...m$ – are classes, N denotes the number of rules $R^{(k)}$, for k=1,2,...,N.

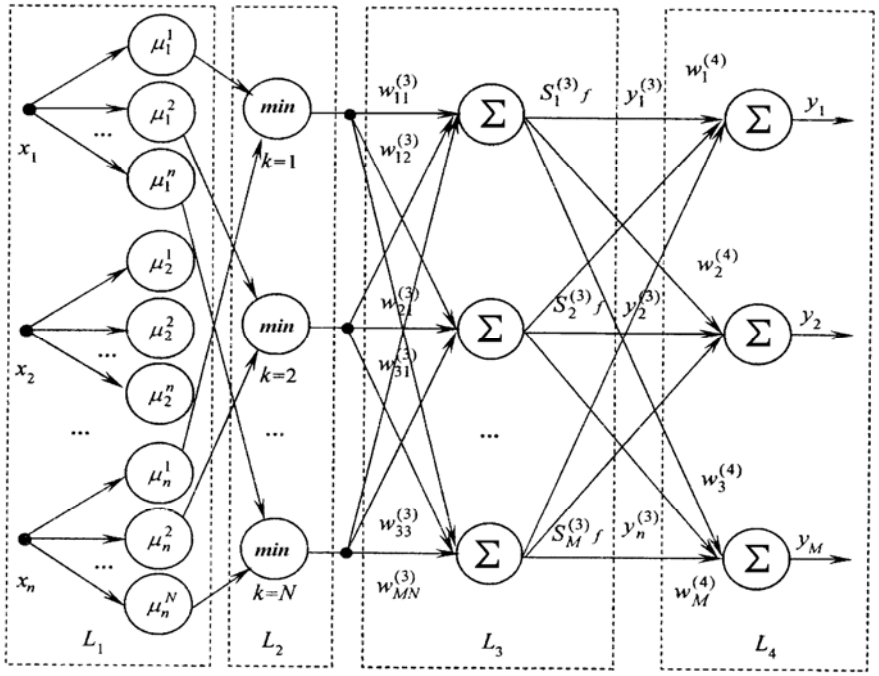


Figure 2. Fuzzy-neuro system for classification of parts and assembly units

In this fuzzy neural classifier each rule is associated with one class. The input values constitute the input vector $X^* = (x_1^*, x_2^*, ..., x_n^*)^T$. The output values represent degrees of rule activation, described as follows:

$$\tau_k = \prod_{i=1}^{n} \mu_{A_i^k}(x_i^*) \tag{2}$$

where $\mu_{A_i^k}(x_i)$ is the Gaussian membership function and

characterized by the center and width parameters, x_i^{*k} and σ_i^k :

$$\mu_{A_i^k}(x_i^*) = exp\left[-\left(\frac{x_i - x_i^{*k}}{\sigma_i^k}\right)^2\right] \tag{3}$$

The neuro-fuzzy system performs a classification task based on the values of τ_k , for k=1,2,...,N. Each input vector $X^* = (x_1^*, x_2^*, ..., x_n^*)^T$ is classified to the class B_l, for *l=1,2,...,m*, which is associated with the maximal degree of rule activation.

In the examples below the presented algorithm have been used the method of geometrical description of the units of machine engines described in [Ed. by Knosala, 2002].

Performance of neuro-fuzzy system has been tested on the following input data. The neuro-fuzzy system has the following features:

- Each neuron represents one fuzzy IF-THEN rule.
- The number of neurons equals to the number of rules in the rule base.
- Weights of the neurons have an interpretation concerning parameters of the

membership functions of the corresponding neuro-fuzzy system.

- It is easy to modify the network architecture when a rule is added or removed (by addition or removal, respectively, the neuron that represents this rule).

Thus, in contrast to classical neural networks, the neuro-fuzzy classifier presented in this paper does not work as a "black box". This classifier is a rule-based neural network.

Neuro-fuzzy classifier can contain many neurons, so it is no problem to increase the number of rules in order to achieve better performance of the classifier.

Conclusions

The approach to the aiding of production systems planning based on the modular technology, proposed in this work, is a very promising direction for research on the field of the new production technologies. In the present the base problem at a practical realization of the presented expert system is lack of an access to data and work immensity, necessary to the pre-processing of the input data and to enter them into the knowledge base.

In the paper we have applied Self Organizing Map of Kohonen and basic soft techniques for classification of parts and assembly units.

The hybrid neuro-fuzzy system briefly presented in the paper was successfully applied for designing intelligent decision support system. By using several advanced technologies (combination of fuzzy logic and neural networks) it is possible to handle a broader range of information and solve more complex problems.

Future research in this work will be using the description of properties received as output date in program to the design CATIA.

The research conducted proves that neural networks of Kohonen and neuro-fuzzy systems are a very effective and useful instrument of classification of the elementary assembly modules and can be employed in order to solve direct clustering of parts families.

Bibliography

[Boothroyd, Knight, 1994] Boothroyd G., Dewhurst P., Knight W.: Product design for manufacture and assembly, Marcel Dekker Inc., USA, 1994.

[Grabmeier, 2002] Grabmeier J., A. Rudolph: Techniques of cluster algorithms in data mining, Data Mining and Knowledge Discovery, 2002, nr. 4, 303–360.

[Jang R., Sun C.T., Mizutani E., 1997] Jang S.R., Sun C.T., Mizutani E.: Neurofuzzy and Soft Computing, Prentice-Hall, Upper Saddle River 1997, p. 245.

[Ed. by Knosala, 2002] Ed. by R. Knosala: Zastosowania metod sztucznej inteligencji w inżynierii produkcji, WNT, 2002

[Kohonen,1990] Kohonen T.: Self-organizing Maps, Proc. IEEE, 1990, 78, NR.9, pp. 1464-1480.

[Lin, 1996] Lin, Chin-Teng and Lee, C.S. George. Neural Fuzzy Systems: A neural-fuzzy synergism to intelligent systems, New Jersey, Prentice-Hall, 1996.

[Łunarski,1993] Łunarski J., Szabajkowicz W.: Automatyzacja procesów technologicznych montażu maszyn, WNT, Warszawa,1993r.

[Malave,1992] Malave C.O., Ramachandran S, Lee H: A self-organizing neural network approach for the design of cellular manufacturing systems //Journal of Intelligent Manufacturing, V.3, 1992, pp. 325–333.

[Moon Y.B., Divers C.K., and H.-J. Kim, 1998] Moon Y.B., Divers C.K., and H.-J. Kim: AEWS: An Integrated Knowledge-based System with Neural Network for Reliability Prediction // Computers in Industry, 1998, Vol.35, N2, pp.312-344.

[Nauck, Kruse,1995] D. Nauck and R. Kruse, NEFCLASS: A Neuro-Fuzzy Approach for the Classification of Data, In Proceedings of ACM Symposium on Applied Computing, George K et al (Eds.), Nashville, ACM Press, pp. 461–465, 1995.

[Ramachandran, 1991] Ramachandran S, Malave C.: A neural network-based design of cellular manufacturing systems // Journal of Intelligent Manufacturing, V.2, 1991, pp. 305–314

[Rutkowska, 2002] Rutkowska D.: Neuro-fuzzy Architectures and Hybrid Learning, Springer-Verlag, Heidelberg, New York, 2002.

[Setlak,1999] Setlak G.: Hybrydowy system ekspertowy do projektowania procesów montażowych //Technologia I automatyzacja montażu, Warszawa, Nr.3, 1999, str. 23-27.

[Setlak,2004] G.Setlak: Intelligent Decision Support System, Kiev, LOGOS, 2004, (in Rus.), 252 p.

[Setlak, 2008] G.Setlak: Intelligent System for Computer Aided Assembly Process Planning // International Book Series "Information Science and Computing", Number 4, "Advanced Studies in Software and Knowledge Engineering", Pub.of the Institute of Information Theories and Applications FOI ITHEA and Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, 2008, N4, pp.9-17.

[Szabajkowicz, 1998] Szabajkowicz W.: Modułowe technologie montażu // Technologia i automatyzacja montażu, Warszawa, Nr.4, 1998, str. 9-11.

[Zolghadri Jahromi, Taheri, 2008] M. Zolghadri Jahromi, M. Taheri, A proposed method for learning rule weights in fuzzy rule-based classification systems, Fuzzy Sets and Systems, V. 159 (2008),PP. 449 – 459

Authors' Information

Galina Setlak, *D.Sc, Ph.D., Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science , W. Pola 2 Rzeszow 35-959, Poland, and The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland, e-mail: gsetlak@prz.edu.pl*

Tomasz Kożak *– PhD Student, Institute of Technical Engineering State School Of Higher Vocational Education, Czarniecki Street 16, 37-500 Jaroslaw, Poland; e-mail: tomasz.kozak@pwszjar.edu.pl*

Monika Piróg-Mazur *– PhD Student, Institute of Technical Engineering State School Of Higher Vocational Education, Czarniecki Street 16, 37-500 Jaroslaw, Poland; e-mail: m_pirog@pwszjar.edu.pl*

Wioletta Szajnar *– PhD Student, Institute of Technical Engineering State School Of Higher Vocational Education, Czarniecki Street 16, 37-500 Jaroslaw, Poland; e-mail wiola@pwszjar.edu.pl*

ROBOT CONTROL VIA DIALOGUE

Arkady Yuschenko

Abstract: *The most rational mode of communication between intelligent robot and human-operator is bilateral speech dialogue using a problem-oriented language. The dialogue mode of control raises the problem of compliance between the human and robot perception of external world, logic, behavior planning and decision making. We suppose that linguistic variables and fuzzy logic is the most suitable approach for the mentioned problems. The model of the world formed on the base of natural space-time relations allows formulating the basic robot operations in terms of linguistic variables and to solve the problem of planning of robot behavior in previously undetermined environment. Using the fuzzy neural networks it is possible to train robot to fulfill complicated operations by human-operator. Relationship between fuzzy logic approach and the procedures of speech recognition are also under consideration*

Keywords: *mobile robot, fuzzy logic, human operator, remote control, artificial intelligence, operation planning, robot learning*

Introduction

Robot is a technical system for autonomous work in previously undetermined environment. Mobile robots are normally equipped with manipulators, different sensors, including vision systems and on-board computer. Robot may be treated as an intelligent robotic system (IRS) because it is capable to describe an image of the current situation, to analyze the environment and separate objects, and to plan its own behavior necessary to reach the aim stated by human. So robot is a single technical system capable for active cognitive behavior. Such systems are usually controlled by human operator because his experience and intelligence are necessary to fulfill the hazardous and responsible operations. But the mode of control of intelligent robot is sharp different from the tradition modes. The control signals now are the speech commands using a professional problem-oriented language. Feedback is the observation of the robot behavior (if possible) and speech reply from robot to operator in the cases of indeterminacy or lack of information. So the most rational mode of communication between robot and operator is bilateral speech dialogue using a problem-oriented language close to natural speech. Intelligent robotic system controlled via robot-human dialogue is not more a “master-slave” system, as robot now is an equal participant of the process; IRS may be treated as a “master-assistant system”. The dialogue mode of control raises the problem of compliance between the human and robot perception of external world, logic, behavior planning and decision making. Some of the problems are under consideration below.

Environment representation and problem of understanding

We suppose that linguistic variables and fuzzy logic may be the most adequate means to solve the problem of external world representation in a human-controlled IRS. The “mental” world model of robot is based on the corresponding representation of external world in human mind. Such description includes the description of physical objects in the environment as well as spatial and temporal relations between them.

To describe the spatial relations of the current scene, extensional and intensional types of relations are applied The former are represented by relations that describes location and orientation of objects. For example *a1 is far, to the right, and a little above a2.* The latter include the relations like R_1 *– to be adjacent to;* R_2 *- to be inside of; etc.* Extensional relations are determined by membership functions, which in its turn may be determined

experimentally using the statistically processing estimation of the same spatial relations by human-operators. As for intensional relations – usually we used two unary relations - R_{00} – *to be horizontal* and R_{01} – *to be vertical*, as well as 28 basic spatial binary relations. Other may be determined from the basic ones as a conjunction [Pospelov, 1989].

The space situation is determined as a set of binary frames <*object1-relations-object2*>. Usually the observer, i.e. the robot itself, is one of the objects. For every chain of binary frames may de determined the fussy relation between the initial and the terminal elements of the chain. It make it possible to provide the mobile robot navigation through the fuzzy determined map. Also is possible to change the point of view. For example, to change the point of view of robot (in relative coordinate system) to point of view of operator, working in the inertial coordinate system [Yuschenko,2002]. Since the environment is ever-changing due the motion of the observed objects as well as to the motion of the robot itself, the scene description changes in time respectively. This circumstance requires that we take into account not only spatial but also temporal relations in the external world, such as *to be simultaneous with, to be prior to, to follow* et.c. The temporal relations evidently need the memory of the preceding situations.

The names of specified objects of the current scene and space-temporal relations between them form the thesaurus of the situational problem-oriented language (SPOL) for human-robot communication. The situation now may be represented as a set of binary relations. Scene description allows a formal semiotic representation that uses the spatial-temporal relations logic. So, a complex relation *a_1 is on the surface S far and to the right* can be written as $(a_1 R_8 S)\&(a_0 d_5 f_7 a_1)$, where a_0 – is the observer, with respect to whom the distance and orientation relations are formulated, R_8 - *to be on the surface,* d_5 is *to be far,* and f_7 is *to be to the right.*.

The world model is sufficient for robot navigation but does not sufficient for manipulation with the objects of the world. The model of world is incomplete without sensation. Robotic system is equipped with tactile sensors, force-torques sensors and other necessary types of sensors. In make it possible to introduce to the world model such terms as heavy or light, warm or cold, smooth or rough [Roy,2003]. All the terms are fussy and may be represented in the knowledge base of robot by corresponding membership function. As show our results the linguistic representation of feeling is possible but doesn't usually effective in robotics. The same as for human the reflective robot motion is a synergetic complex perception-action. That is why for robot is more effective to form the same complexes by previous training using teachable fuzzy (or hybrid) neural networks [Vechkanov 2002]. So the robot model of world seems to be of hybrid type: linguistic relations are complemented by sensory stereotypes.

As far as robotic system can present the image of the scene constructed from the sensors data in linguistic terms, it may describe the situation for the human by text in SPOL or by speech. And vice versa if operator represents the situation by speech, robot can present the graphic image of situation with before known objects. The cognitive activity of robot makes it possible to transform one form of presentation into another which may be determined as "understanding" the situation by robot.

Operations description and commands

External world description allows formalizing the robot's operations within it [Yuschenko,2002]. We assume that complex operations performed by robot can be represented as a sequence of a relatively few types of elementary operations. These are define in advance and are stored in the IRS knowledge base as frames of typical operations. A frame of this kind contains linguistic variables based description of the aims of an operation, the initial stage scene, and the preconditions for the feasibility of the operation. The latter may depend on the specific situation, the capabilities of the robot in question, and the properties of the object of the operation. Thus, the structure of a typical operation frame is as follows: <*operation name*> <*operation object*> <*initial situation*

(modifier of place)> <*target situation*> <*operation feasibility conditions (preconditions)*>.<*additional details*>. For example: <*move*> <*object A*> <*object A on B*> <*object A on C*> <*object A is free*> <*install object A shock-free*>. While performing technological operations this frame should sometimes have an extra slot <operation performance method (modifier of manner)>.

Preconditions are one peculiarity of the discussed operations description approach. Generally, all preconditions can belong to one of at least three types: a) situational, e.g. the condition *object A is free* means that *there are no other objects on object A*; b) preconditions stipulated by the robot's capabilities: *the robot is equipped with the gripper suitable for type and size of the object*; and c) preconditions connected with the peculiarities of the object: *the object is a rigid body and can withstand the force developed by the gripper without any damage*. The operator can control a robotic system directly by giving the names and aims of the typical operations in the problem-oriented language, e.g., forming the commands as ‹*move object A to plane C*› ‹*insert shaft A into orifice O*›. Such commands call the full frame as before. Some slots of the operation frame may be empty. To fulfill them robot may address queries to the operator with corresponding questions. Another way of the command complication is the cognitive activity of robot, which can change the point of observation or investigate the manipulation object using the sensors. Preconditions description may not always be complete in the sense that some of them may not be defined. For example, it may not be known whether there is free space on plane C, on which object A is to be put. Then a query to the cognitive operations base is formed, and an operation is selected for examining plane C that is supposed to provide for filling in the empty slot. The system can also formulate address queries to the operator, if cognitive actions yield no results or uncertainty persists.

The description of typical operations expands the situation description thesaurus introduced above. Another problem for complicated sentences is the syntax of the SPOL. As may be seen above the problem is solved by the rigid structure of the operation frame. It is not the best way for operator, but it allows to solve the problem of separation of words in continuous speech. The commands perceived by robot may be corrected via dialogue. Possibility of IRS to transform the information from one form into another mentioned above allow to present the goal situation in graphic form and to use means of computer aim-pointing. Last time other ways to command by robots are under consideration such as arm movements and gesture [Chernakova, 2009].

Complex operations planning

A distinctive feature of planning procedure in robotics, as compared to numerous methods of AI- planning, is the possibility of continuous comparison of the real situation observations and the conditions defined during the planning stage. The contradiction between the real and prescribed situation generate the plan to resolve contradiction between the reality and it desirable image [Magazov, 2007]. In the case when preconditions of prescribed operation are not satisfied the operation can not be classified as elementary ones. For example the object is to be transported from initial place to another but the precondition of situation type is not satisfied (object is not free). The contradiction emerged from the prescribed and real conditions result to find another operation to solve the contradiction (to make the object free). But preconditions of new operation may be also not satisfied. At last a chain of consented elementary operations will be constructed in such way that result of n-th operation meets the preconditions for the (n+1)-th operation and the last operation in the chain achieve the aim. This procedure can be represented as directed graph, with its root being the target situation [Yuschenko, 2005].

During the realization of the plan we propose that comparison of the real and planned situation is fulfilled after every stage of the procedure. If after the actual completion of n-th operation the prescribed conditions are not achieved, the planning process is repeated, with the current situation being assumed as the initial.

The conflict resolution approach is rather similar to human cognitive activity while planning actions, which is also based on comparing the operative image of the situation and the target image. The conflict resolution principle

application requires a further extension of IRS control language. Besides the typical operations we now need a thesaurus for situational conflicts resolution by means of performing typical operations. If spatial relations are intentional then each type of conflict induces its own typical operation to resolve it. For example: if the aim is: *(a_1 R_8 S),* i.e. object a_1 is on the surface S, while in fact *(a_1 ⌐R_8 S),* then the conflict induces a typical operation *move a_1 to S .* If the aim is defined as *(a_1 R_2 C),* i.e. *shaft a_1 is inside orifice C*, while observation results show *(a_1 ⌐R_2 C),* then a typical operation is induced: *insert a_1 into C.* If the condition *a1 is free* is necessary for further operations, while in fact we have: *(a_2 R_8 a_1),* i.e. *a2 is on a1*, then a typical operation *remove a_2 from a_1* is induced. One can easily proceed with this list of action that resolve intentional type conflicts.

If the relations are extensional there is no need for a special vocabulary for matching the situation with the required typical operation. Conflict can be resolved by performing a typical operation aimed at the relation specified as its precondition. If a mobile robot R is expected to in position *(R d_1 f_1 N)* with respect to observer *N*, while in fact a different conditions holds true: *(R d_2, f_2 N),* then the required operation will be defined in the form of: *move robot R from position (R d_2 f_2 N) to position (R d_1 f_1 N).*

Robot learning by neural fuzzy network

A disadvantage of the existing approach is that the operator has to define the rules for different situations beforehand, hence the situations should also be known in advance. The rules of behavior cannot be formalized in advance when the human control robots using his personally sensory skill. For example a skillful human operator can successfully control the mobile robot movements in environment with complicated obstacles but can not describe the rules he used. Another example is the assembly of the “shaft-hole” type.

For such cases operator guides the robot through typical situations after which information is processed in, e.g. teachable fuzzy (hybrid) neural networks. The network ANFIS (adaptive neuro-fuzzy intelligent system) type was proposed for mobile robot control in the environment with obstacles. The system able to teach itself using the telemetry data received for skillful human operator remote robot control. The robot under consideration had three pairs of caterpillars and the problem to control all of them in real time scale was complicated for human. The task was to teach the mobile robot to get over the typical obstacles by itself. In the experiment the robot was controlled by experienced operator. The input variables were the current angles of the robot platform orientation and the torques of drivers. As the output variables the controlled angles of four robot caterpillars were considered. The training of the net was realized by error back propagation method. The data obtained during experiment by telemetry had been processed by cluster analysis method. It proved to determine both the typical situations and the corresponding control signals formed by operator. After the training procedure the fuzzy network showed results of control close to the same of human control [Vechkanov 2002]. As a matter of fact the hybrid network constructed could by itself formalize the fuzzy rules which human operator could not determine. The fuzzy mode of control allows the robot to get over the obstacles of the prescribed type for wide range of their parameters. But as new situation critically differs from the previous ones the training process has to be repeated again. All the previous tunings may be remembered in the system memory, so the robot “experience” is expanding after all new training cycles.

The same learning control systems may be applied in other types of robotic system to form the “skill” of the artificial system. A good example of such system is the industrial manipulation robot for assembly operations of “shaft-hole” type using a 6 d.o.f. forces-torques sensor. The number of possible situations during such operations is not large and all of them may be described beforehand. So the task of the fuzzy system is to recognize the situation and form the fuzzy rules similar to those for experienced operator. The experiments showed that for situation recognition in this case may be reliably applied an artificial neural network with one hidden layer.

As a matter of fact the reflective behavior of robots use mainly the sensor information so it effectiveness may be improved using other sources of information. The vision system may recognize the type of the obstacle and to adjust the fuzzy controller previously. To recognize the type of the obstacle the robot control system may be taught using the simulation of the environment. The type of the obstacle also may be determined in dialogue with operator.

Another problem under consideration nowadays is to teach robot by itself (i.e. without the teacher), because there are many situations when operator has not necessary information of the robot position and state. For such situation the training without teacher may be realized on the base of trainable neural networks [Zhdanov, 2008]. Author treated the basic feature of such systems as robot emotion, which help the system to recognize the positive or negative result of its behavior in memorized situation. Most perspective direction seems to combine the previously trained system with possibility of self – teaching.

Speech interface

Speech interface consists of recognition and linguistic blocks. The recognition block is a device for transforming speech signals as well as interpreting them as separate words or phrases. The linguistic block performs the interpretation of statements into SPOL, as well as the representation of these statements in a semiotic form.

At present there are two most widely used methods of speech recognition: Dynamic Time Warping (DTW), or template matching, and hypothesis probability estimation using Hidden Markov Models (HMM). For continuous speech recognition one can employ template phrases construction, using the information on the grammar of the SPOL. The HMM method using hypothesis probability estimation with Viterbi algorithm (beam-search) allows to recognize continuous speech almost independent of the speaker. However this method requires a high quality and expensive teaching speech database. N-grams method may be also successfully applied in robotics

Operator/s statements for IRS control can be formulated in the SPOL mentioned above. In the case of successful recognition we obtain, that (1): command is split into semantic units (*action, object, modifiers*), each of the part of the statement-command is assigned permissible and relevant values from the program data-base, feasibility of the operation is assesses, based on the current working scene status. The linguistic analyzer performs the syntactic and semantic decomposition of the statement which is supposed to result in filling the slots of the frame that describes operations. The slots of the frame are filled in with relevant subordinate parts of the command-sentence. The linguistic recognition stage output is a set of encapsulated frames that can be uniquely interpreted over the further stages of command-sentence completion. It was shown above that the sentences represented by linguistic frames can be expressed in the inner semiotic language as a sequence of symbols. Now the formal logic operation and verification of the commands are available.

A part of speech interface is the dialogue control system. There are some possible forms of dialogue scenarios. For example, dialogue for correction of the command syntax, for correction of the slots content, call for operator consent to fulfill the operation and so on. The dialogue control system form the call from robot to human to point the necessary information.

The most complicated problem with speech interface is to provide the continuous operator speech processing. As for speaker dependence the problem is not crucial as there are few operators for a robot and the speech interface may de adjusted individually. The problem will be much more complicate in the case when robot has to support the dialogue with everybody.

Conclusion

The preliminary research has shown that the implementation of speech dialogue type control mode for a robotic system by way of formulating separate commands is inefficient. It is necessary to develop a speech interface meant focused on the use of the situational problem-oriented language similar in its structure to natural language. This allows a substantial simplification of the task of robot control, as it no longer requires any special skills from the operator. There are however a number of tasks in this field that are yet to be solved. In particular, the speech interface application for teaching the robot, rather than merely controlling it. We also attribute crucial importance to the psychological aspects of interaction between a human and an intelligent robotic system, connected with "mutual" ideas about the situations and reasonable behavior. The final aim of the robot control via dialogue investigation seems the including robots to humanitarian socium where everybody can communicate with robots using practically natural language.

Bibliography

[Chernakova, 2009] S.E.Chernakova. Simulation of the intelligent mechatronic systems and robots training with movement demonstration. In: Mechatronika, Automation , Control, № 2, 2009, p.p.53-56

[Magazov, 2007]. S.S.Magazov. Cognitive processes and models. URSS, Moscow, 2007.

[Pospelov, 1989] E.Ju.Kandrashina, L.V.Litvinceva, D.A.Pospelov. Representation of knowledge of time and space in intelligent systems. – Nauka - Phys.matt.lith., Moscow, 1989 .

[Roy,2003] K.-Y.Xsiao, N.Mavridis, D.Roy. Coupling perception and simulation. Steps towards conversational robotics. In: Proc.IEEE/RSJ Int. Conf on Intelligent robotics and systems, 2003.

[Yuschenko,2002] A.S.Yuschenko. Robot distance control using fuzzy concepts, In: Iskusstvennyi intellect, Vol. 4, NAS Ukraine, 2002, pp. 388-396

[[Yuschenko,2005] A.S.Yuschenko. Intelligent planning in robot operation. In: Mekhatronika, Vol.3, 2005, pp. 5-18

[Vechkanov 2002] V.V. Vechkanov, D.V. Kiselev, A.S.Yuschenko. Adaptive system for mobile robot fuzzy control. In: Mekhatronika, Vol. 1, 2002, pp. 20-26

[Zhdanov, 2008] A. A..Zhdanov . Autonomous artificial intelligence. BINOM, Moscow, 2008.

Authors' Information

Arkady Yuschenko *– Professor, head of Department "Robotic Systems", Bauman Moscow State Technical University, 105005, Moscow, 2-nd Baumanskaya str., 5, Russia: e-mail: robot@bmstu.ru*

Business Intelligence Systems

APPLICATION OF DISCRETE OPTIMIZATION IN SOLVING A PROBLEM OF MULTI-ITEM CAPACITATED LOT-SIZING WITH ECONOMIC OBJECTIVES

Igor Kononenko, Igor Protasov

Abstract: *In this paper we study the problem of multi-item capacitated lot-sizing from the point of commercial enterprises. We consider profit as the main criteria. This dynamic problem belongs to the class of discrete optimization and contains boolean variables, algorithmic objective function, where various types of constraints such as analytical functions, algorithmic and simulation models can be used. We present model and direct search algorithm that consists of an intelligent iterative search and upper bound set construction, and allows finding exact solution in reasonable time. We carry out computational investigation and solve a real task with using developed computational tool to show the efficiency and practical application of the proposed model and algorithm.*

Keywords: *direct search; discrete optimization; upper bound; production planning; multi-item capacitated lot-sizing; profit maximization*

ACM Classification Keywords: *G.1.6 Optimization; J.0 Computer Applications*

Introduction

The task of determining the maximum-profit production plan for lot-sizing problem is a very important issue for commercial enterpreises. The quality of the enterprise strategy highly depends on the successful solution of this problem.

Traditionally, the main criteria for the problem of lot-sizing was a criteria of minimizing the production costs to all phases of the product's life cycle [Absi, 2008], [Pochet, Wolsey, 2006], [Walser, 1999].

The paper [Beresnev, Gimadi, Dementyev, 1978] reviews various models of product lot-sizing optimization. An algorithm based on the branch-and-bound method was also proposed. The proposed mathematical models do not take into account the possibility of using the algorithmic objective function and constraints that help the model to reflect the object properties and behaviour more accurately.

Several types of stochastic models, that describe the problem of product lot-sizing optimization, were proposed in the paper [Antipenko, Katz, Petrushov, 1990]. This approach allows to find a local extremum of multiextremal objective function. Considered models take into account the costs of product customization according to specific consumer needs.

The paper [Kononenko, 1990] proposes models of the dynamic product lot-sizing optimization. The models contain single objective function and different types of constraints: analytical, algorithmic, and simulation models. The proposed algorithms allow to find an exact solution.

In the paper [Kononenko, Rogovoi, 2000] a mathematical model and non-Markov approach to multi-objective optimization for a dynamic product lot-sizing were developed. The proposed model minimizes costs in all phases

of the product life cycle, and takes into account specific needs of the consumers. The model's constraints can be of various types such as analytical expressions, algorithms or simulation models. The method is based on minimax and direct search approaches.

The paper [Kononenko, Derevjanchenko, 1999] reviews an approach and a fuzzy constraint based model of dynamic product lot-sizing optimization. To enshure high adequacy of mathematical models the optimization-simulation approach [Tsvirkun, Akinfiev, Filippov, 1985] was proposed for mathematical-programming problems containing algorithmic objective function and constraints.

Use of profit as the main criteria for solving the lot-sizing problem can be beneficial for the commercial enterprises and make them more competitve in changing market conditions [Kononenko, Protasov, 2005],[Kononenko, Protasov, Protasova, 2005].

The purpose of this research is to develop a mathematical model and an approach to the solution of dynamic multi-item capacitated lot-sizing problem with using profit as the main criteria. The constraints of the model can be of various types: analytical, algorithmic or simulation models.

Description of the proposed mathematical model

Let's assume that product *i*, $i = \overline{1,m}$ is used to meet needs of consumer *j*, $j = \overline{1,n}$. *Pij* is the production quantity of product *i* that is required to fullfill the demand of consumer *j*. *T* is a period of production planning.

The mathematical model of multi-item capacitated lot-sizing problem uses profit maximization as the main criteria. This model is formulated below.

$$\sum_{t=1}^{T}\sum_{i=1}^{m}\left[Ц_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)\right]\times\left(\sum_{j\in\{j:\rfloor j/\chi\lceil=t\}} p_{ij}x_{ij}\right)\alpha_t - \sum_{i=1}^{m}\left[w_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)\right]\alpha_{t_i'} -$$

$$\sum_{t=1}^{T}\sum_{i=1}^{m}\left[v_i\left(\sum_{j=1}^{n} p_{ij}x_{ij} + p_{i,npeo}\right)\right]\times\left(\sum_{j\in\{j:\rfloor j/\chi\lceil=t\}} p_{ij}x_{ij}\right)\alpha_t - \sum_{j=1}^{n}\sum_{i=1}^{m} z_{ij}\alpha_{t_j}x_{ij} \to \max_{x_{ij}} \quad (1)$$

$$S_t = S_{t-1}\alpha_t / \alpha_{t-1} + K_t\alpha_t - \sum_{i\in\{i:t_i'=i\}} w_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)\alpha_{t_i'}, \quad (2)$$

$$S_t \geq 0,\ t = \overline{1,T},$$

$$b_t^{(h)} = \varphi\left(b_{t-1}^{(h)}, x_{ij}\right), i = \overline{1,m}, j \in \{j : \rfloor j/\chi\lceil = t\}; \quad (3)$$

$$d_t^{(h)} = \sum_{j\in\{j:\rfloor j/\chi\lceil=t\}} p_{hj}x_{hj}$$

$$b_t^{(h)} \geq d_t^{(h)}$$

$$\alpha_t^{(q)} = f\left(\alpha_{t-1}^{(q)}, x_{ij}\right), i = \overline{1,m}, j \in \{j : \rfloor j/\chi\lceil = t\}; \quad (4)$$

$$\alpha_t^{(q)} = \left\{\begin{matrix}\leq\\ \geq\end{matrix}\right\} e_t^{(q)} \ \forall q \in Q, t = \overline{1,T};$$

$$x_{ij} \in \{0,1\}, i = \overline{1,m}, j = \overline{1,n} \quad (5)$$

$$\sum_{i=1}^{m} x_{ij} = 1,\ \ j \in U$$

$U_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$ function is used as the price of product *i* and depends on the production volume.

$w_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$ function defines the preproduction costs that include costs for scientific research, experimental development, production tooling, etc.

$v_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$ function defines the production costs per unit of product *i* and depends on the production volume.

z_{ij} defines the costs for transportation, adjustments, and possible setup costs that are required for the prouct *i* to meet the needs of the consumer *j*.

Φ is a set of products that were previously developed and do not need preproduction.

$t_i^{'}$ is the year when product *i* was developed.

$t_i^{'} = \min_{j=\overline{1,n}} T_i$, если $i \notin \Phi;\ T_i = \{t_j : t_j =]j / \chi[, x_{ij} = 1,\ j = \overline{1,n}\};$

T_i is the number of years of product *i* life cycle.

To accommodate the change in prices over the life cycle the discount α_t is introduced.

$a_k = (1 + E_H)^{t_p - k}$, E_H - is a costs norm at different periods., t_p is a due year, $a_{t_i^{'}} = 0$ when $i \in \Phi$

$p_{i,\,prev}$ is the quantity of product *i* that was previously developed;

S_0 – funds for the preparation of all types of products production available due to the plannig period;

$b_t^{(h)}$ is the productive capacity's value for product *h* in *t* year;

$a_t^{(q)}$ is a parameter that is estimated by using the analytic function, algorithm or simulation model. $e_t^{(q)}$ is the $a_t^{(q)}$ parameter's claim in *t*-year.

$p_{ij} = 0,\ z_{ij} = 0$, if the product *i* can not fullfill the demand of the consumer *j*.

The various types of constraints can be used in the model and can be defined as analytical funtions, algorithmic or simulation models.

This dynamic problem belongs to the class of discrete optimization and contains boolean variables, algorithmic objective function, where various types of constraints such as analytical functions, algorithmic and simulation models can be used.

The search for optimal solutions constitutes the most important problem in the scope of discrete optimization. Improving the search efficiency is of considerable importance since exhaustive search is often impracticable.

The proposed direct search method consists of an intelligent iterative search that is based on upper bounds definition and avoids visiting those potential decision subsets, which are known not to contain an appropriate solution. The given method finds an exact optimal solution in reasonable time.

Description of the proposed direct search method

1. Define the profit upper bound when servicing all the consumers starting from *(j+1)* to *n*, $j=\overline{1,n}$.

The following equation defines only a profit upper bound for a single *j* consumer, $j=\overline{2,n}$.

$$\Pi_j^{\max}=\max_i\{[U_i(1)-v_i(\infty)]\,p_{ij}\}_{i=1}^{m}$$

Now, we need to find upper bounds that are required for computation of partial decisions vector, which consists of *j* coordinates, $j=\overline{1,n-1}$.

$$\Delta\Pi_j=\sum_{k=j+1}^{n}\Pi_k^{\max}\,\alpha_{]k/\chi[},$$

If $j=n,$ then $\Delta\Pi_j=0$.

Before we go to the next step, it is necessary to set values of the following variables $Д_0=0, V_0:=0, W_0:=0, Z_0:=0$. If $x_{jk}^0:=1$, then set $y_{j:}:=k \quad \forall j=\overline{\chi(1-g)+1,0}$, where $y_{j:}$- is a coordinate of the current decision vector. This decision vector also includes prehistory decisions,

$$Y=(y_{\chi(1-g)+1}, y_{\chi(1-g)+2}, \ \cdots\ , y_0, y_1, \cdots, y_n)^T$$

Set the initial record value for the objective function $\Pi_{record}=-\infty$ and start servicing the first consumer request, so set *j*=1.

2. Try not to use any products for servicing the consumer *j*. This means setting the current decision vector's coordinate y_j and current product *i* as follow $y_j=i=0$.

3. If $j\in U$, then set *i*:=*i*+1 and go to step 4. If $j\notin U$ and *i*=0, go to step 5.

4. Check fulfillment of the constraints.

4.1 If $p_{ij}=0$, then set $y_j=0,\ \ x_{ij}=0$ and go to step 8.

4.2 Set $y_j=i,\ \ x_{ij}=1$. Check fulfillment of the (2) constraint. If $j<n$, then check the constraints for the following year $]j/\chi[=t$. At the same time set $w_i\left(\sum_{j=1}^{n}p_{ij}x_{ij}\right)=w_i(p_i)$, if the product *i* has not yet been designed. If *j*=n , then check fulfillment of the (2) constraint for all years $t=\overline{1,T}$. If any one of the constraints is not met, set $y_j=0,\ \ x_{ij}=0$ and go to step 8.

4.3 Checking fulfillment of restrictions (3) in the year *t* by using the algorithm production capacity or through a simulation model. If any one of them is not met, set $y_j=0,\ \ x_{ij}=0$ and go to step 8.

4.4 Checking fulfillment of the algorithmic constraints (4) in year *t*. If any one of them is not met, set $y_j=0,\ \ x_{ij}=0$ and go to step 8.

5. Determine appropriateness of the decision.

5.1 Calculate the costs $Z_{j,}\, V_j, W_j$ and the value of income $Д_j$.

5.2 Calculate $\Pi_j=Д_j-W_j-V_j-Z_j$ and find values for the profits upper bound $\overline{\Pi}=\Pi_j+\Delta\Pi_j$.

5.3 If $\overline{\Pi}\le\Pi_{record}$, this means that using the product *i* for the servicing the consumer *j* does not provide a better profit value than the record profit value, that has been determined before.

Then, set $y_j = 0,\ x_{ij} = 0$ and go to step 8.

6. If $j < n,$ then consider the next consumer request. Set *j:=j+1* and go to step 2.

7. Store the obtained profit record value $\Pi_{record} := \overline{\Pi}$ and its corresponding decision vector $R(j) = y_j \ \forall j = \overline{1,n}$

8. If i<m, then consider the next product i. Set *i:=i+1* and go to step 4.

9. If j>1, then consider the previous consumer request. Set *j:=j-1, i=*y_j and go to step 8.

When *j*=1, then if $\Pi_{record} > -\infty$ this indicates that solution has been found, otherwise the problem has no solution.

Proof of the valid upper bound definition

When calculating the upper bound while servicing the consumer request j, the maximum possible revenue value $U_i(1)p_{ij}$ and the lowest possible costs $v_i(\infty)p_{ij}$ are used. It can be argued that the mutual influence of servicing the consumers by various products can not give a profit value higher than the obtained upper bound. This is a true statement, because product price depends on its production volume, and the price of product produced in any number of units can not be higher than the price of product produced in single quantity $U_i(1)$. The production cost of product produced in any number of units can not be less than the production cost for product produced in a maximum possible quantity $v_i(\infty)$.

A computational investigation of the given method

A computational investigation of the method was carried out to show the method's efficiency.

Let's assume that we have to find an exact solution to a problem with 60x15 size. Using an exhaustive search to find an exact solution to this problem takes $1{,}17*10^{54}$ years, if the time for checking each possible solution is 10^{-9} sec. The results of the given method's computational investigation are shown below and prove the method's search efficiency. The computational investigation was done using a computational tool that will be further described in this paper. A desktop computer with the following hardware configuration was used: AMD Athlon 64 1.8 Ghz, 1GB RAM.

Table 1. The results of the computational investigation

Average solution search time, sec.	Problem size			Number of problems that have been solved
	n	m	nxm	
2	10	6	60	5
4	20	6	120	5
9	30	6	180	5
17	40	6	240	5
32	50	6	300	5
48	60	6	360	5
4	10	10	100	5
6	20	10	200	5
15	30	10	300	5
29	40	10	400	5
56	50	10	500	5
116	60	10	600	5

4	10	15	150	5
9	20	15	300	5
23	30	15	450	5
46	40	15	600	5
105	50	15	750	5
237	60	15	900	5

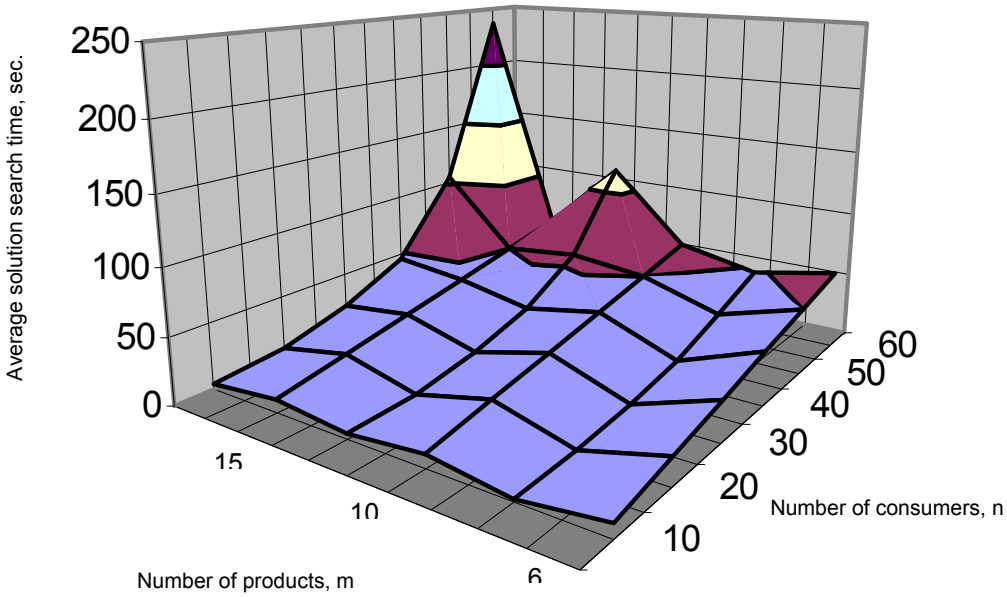


Figure 1. Problem size and its solution search time

Main stages in problem solving process and modular design application

Stage 1. This stage requires carrying out market research and calculating of forecasted demands for products that will meet the needs of the market.

Stage 2. Modules definition

A product can be assembled from the set of various predefined modules. For instance, in computer production the following components can be considered as modules: CPU, RAM, HDD, etc.

Each module has various features and specification figures such as performance, reliability index, demand parameter, economic figures, competitive figures, etc.

At this stage a set of possible modules are defined that will be further used for products' design.

Stage 3. Selection of product clusters

Selection of product clusters is based on forecasted demands, and products' features and specification figures. For example, in computer manufacturing products can be grouped into the following clusters: computers for scientific research (multiple CPU, a lot of RAM), servers (multiple CPU, a lot of RAM, high capacity HDD, RAID), play stations (high performance CPU, high capacity HDD, a lot of video memory and RAM), office computers (high hardware requirements are not needed).

Stage 4. Finding out features and specification figures of products that will meet the needs of selected product cluster.

Stage 5. Design various product items based on the predefined modules.

At this stage by combining different predefined modules we determine specific products that will meet the needs of product clusters.

Stage 6. Solving the multi-item capacitated lot-sizing problem in order to compute a maximum-profit production plan such that all customer orders are met in time.

Using a computational tool

Within scope of the research, a computational tool was developed. The tool was used for the computational investigation of the given method. As it was described above, the detailed market research and informatiom preparation are required for getting accurate input data. Let's assume that we have all the required input data. The computational tool has a number of forms for setting the input data prior to the problem's solution search. The main input parameters are described below. First, we set the matrix of forecasted demands that shows products' production quantity that is required to fulfill the demands of consumers.

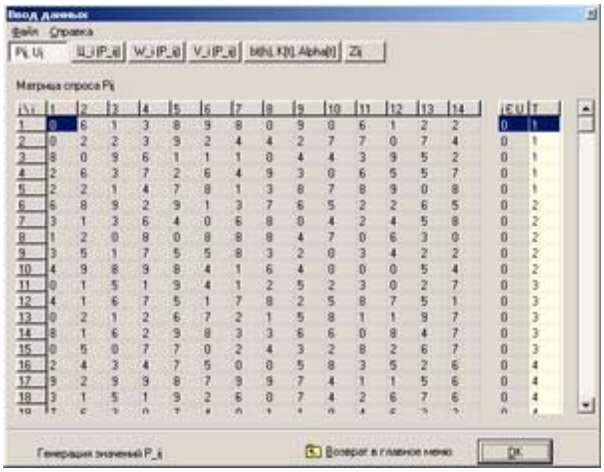


Figure 2. The matrix of forecasted demands

The next step is setting the matrixes of products price $Ц_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$, the preproduction costs matrix $w_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$, and the the production costs $v_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$. All these matrixes are represented as functional dependecies and depend on the production volume.

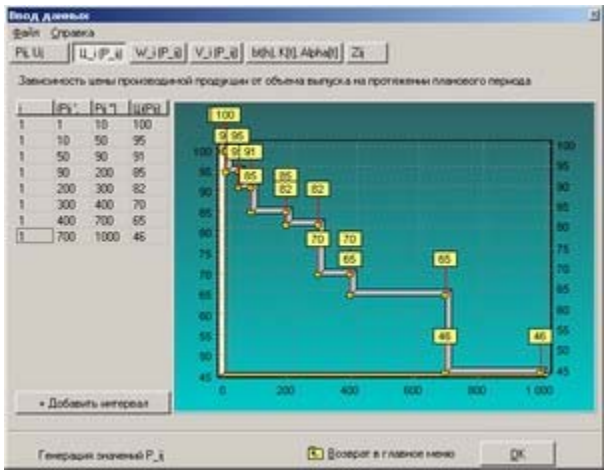


Figure 3. The products price matrix

When the process of inputing the data is accomplished, we can start the solution search process.

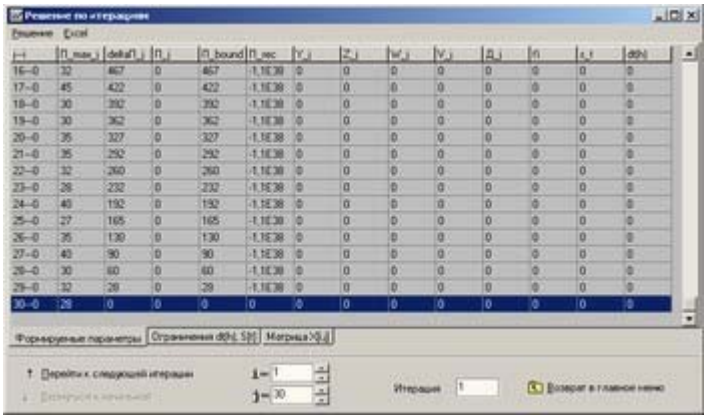


Figure 4. Solution search window

7. A practical application of the given approach

A company that is specialized in computer manufacturing is planning production of 4 basic types of products.

Type 1 - Servers (market value - $5,000.00)

Type 2 - Laptops (market value - $ 1,500.00)

Type 3 - Game stations (market value - $1,000.00)

Type 4 - Office computers (market value - $500.00)

Based on the marketing research data, the matrix of forecasted demands p_{ij} was determined. The production planning period is 3 years.

According to the marketing analysis, it is anticipated that the maximum number of customer orders per one year χ will be 100 units. Thus, the dimension of the matrix of forecasted demands p_{ij} (m x n) is 4 x 300.

There are $500,000.00 of existing funds S_0 prior to the planing period that can be used to cover the preproduction and any other further costs.

Additional capital investments in the first, second and third year will be $100,000.00.

The α_t discount value is 15%. The production planning period $T_i = 3$ years.

The functions that represents products' prices $Ц_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$, the preproduction costs $w_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$, and the the production costs $v_i\left(\sum_{j=1}^{n} p_{ij}x_{ij}\right)$ are defined in the form of nonlinear dependence.

Table 1Functional dependencies of the product price, preproduction costs, production costs and the production volume of the product "Servers"

Pi interval		Vi	$Ц_i$	Wi
1	4	3610	5198,4	24000
5	10	3590	5169,6	24200
11	20	3550	5112	24400
21	30	3500	5040	40000
31	40	3470	4996,8	40000
41	50	3450	4968	52000

Table 2 Functional dependencies of the product price, preproduction costs, production costs and the production volume of the product "Laptops"

Pi interval		Vi	Цi	Wi
1	4	1100	1584	24000
5	10	1080	1555,2	24200
11	20	1060	1526,4	24400
21	30	1042	1500,48	40000
31	40	1010	1454,4	40000
41	50	990	1425,6	52000

Table 3 Functional dependencies of the product price, preproduction costs, production costs and the production volume of the product "Game stations"

Pi interval		Vi	Цi	Wi
1	4	735	1058,4	24000
5	10	725	1044	24200
11	20	715	1029,6	24400
21	30	695	1000,8	40000
31	40	670	964,8	40000
41	50	640	921,6	52000

Table 4 Functional of between the product price, preproduction costs, production costs and the production volume of the product "Office computers"

Pi interval		Vi	Цi	Wi
1	4	397	571,68	24000
5	10	380	547,2	24200
11	20	365	525,6	24400
21	30	350	504	40000
31	40	320	460,8	40000
41	50	297	427,68	52000

Table 5 Productive capacity constraints

t \ i	1 (Servers)	2 (Laptops)	3 (Game st.)	4 (Office comp.)
1	150	300	400	600
2	150	300	400	600
3	150	300	400	600

Zij, the costs of transportation, adjustments, and possible setup costs, is $ 20.00 per unit of the product.

The company also has contractual obligations with some customers for computers supply $j \in U$, where U is a contractual obligations matrix.

As a result, the maximum-profit production plan was found. The maximum possible profit Π_{record} is $ 562, 042.

The obtained production plan is shown below.

According to solution of the problem it follows that during the first year the company needs to fulfill 28 customer orders by supplying 23 units of type-1, 250 units of type -2, 60 units of type-3 and 200 units of type-4.

During the second year it is going to fulfill 30 customer orders and needs to produce 42 units of type-1, 245 units of type-2, 50 units of type-3 and 110 units of type-4.

In the third year it is going to fulfill 19 customer orders and it is necessary to produce 44 units of type-1, 150 units of type-2, 30 units of type-3 and 465 units of type-4.

Conclusion

This research paper reviews the problem of dynamic multi-item capacitated lot-sizing problem from the point of commercial enterprises. The model and the direct search method were proposed. The given method finds an exact solution in reasonable time. The main stages in the problem solving process and modular design application were described. A computational tool was developed and used for the computational investigation of the given method. The results of the research show method's search efficiency and its practical application value for commercial enterprises as it is able to determine a maximum-profit production plan.

Bibliography

[Absi, 2008] N. Absi, S. Kedad-Sidhoum. The multi-item capacitated lot-sizing problem with setup times and shortage costs, European Journal of Operational Research 185 (3) (2008), pp. 1351-1374

[Pochet, Wolsey, 2006] Y. Pochet, L.A. Wolsey. Production Planning by Mixed Integer Programming, Springer, 2006

[Walser, 1999] J. P. Walser, Integer Optimization by Local Search: A Domain-Independent Approach, Springer, 1999

[Beresnev, Gimadi, Dementyev, 1978] V.L. Beresnev, E.K. Gimadi, V.T.Dementyev, Ekstremalnie zadachi standartizacii (Extremal problems of standardization), Nauka, Novosibirsk, 1978, p. 334.

[Antipenko, Katz, Petrushov, 1990] V.S. Antipenko, G.B. Katz, V.A. Petrushov. Modeli i metodi optimizacii parametricheskih ryadov (Models and methods of the parametric series optimization),Mashinostroenie, Moscow, 1990, p. 176.

[Kononenko, 1990] I.V. Kononenko, Optimizacia dinamicheskogo tiporazmernogo ryada oborudobania (Dynamic lot-sizing optimization), Vestnik NTU "KhPI" 10 (1990), pp. 48-51

[Kononenko, Rogovoi, 2000] I.V. Kononenko, A.I. Rogovoi. Vektornaja optimizacia dinamicheskogo tiporazmernogo ryada produkcii (Vector optimization for a dynamic product lot-sizing), Kibernetika i Sistemnyi Analiz 2 (2000), pp. 157-163.

[Kononenko, Derevjanchenko, 1999] I.V. Kononenko, B.I. Derevjanchenko. Optimizacija tipazha produkcii, prednaznachennoj dlja posledovatel'nogo obsluzhivanija zajavok, pri nechetkoj ishodnoj informacii (Product lot-sizing optimization under the fuzzy input data conditions), Vestnik NTU "KhPI" 73 (1999), pp. 84-88.

[Tsvirkun, Akinfiev, Filippov, 1985] A. D. Tsvirkun, , V. K. Akinfiev, V. A. Filippov. Imitatsionnoe modelirovanie v zadachakh sinteza struktury slozhnykh sistem (Simulation in the Problems of Structural Design of Complex Systems), Nauka, Moscow, 1985.

[Kononenko, Protasov, 2005] I.V. Kononenko, I.V. Protasov. Maksimizacija pribyli pri formirovanii tipazha perspektivnoj produkcii (Profit maximization for the perspective products lot-sizing), Vestnik NTU "KhPI" 41 (2005), pp. 63-66.

[Kononenko, Protasov, Protasova, 2005] I.V. Kononenko, I.V. Protasov, L.A. Protasova. Planirovanie proizvodstva perspektivnoj produkcii kommercheskih predprijatij v uslovijah rynochnoj ekonomiki (Production planning of perspective products for commercial enterprises in market economy conditions), 3-d International scientific and technical conference "Modern information technologies in economy and management of enterprises, programs and projects", National Aerospace University "KhAI", 2005, pp.53-54.

Authors' Information

Igor Kononenko *– Prof., Head of Strategic Management Department; National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine ; e-mail: kiv@kpi.kharkov.ua*

Igor Protasov *– Co-author, Strategic Management Department; National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine ; e-mail: igor.protasov@gmail.com*

IMPLEMENTATION OF DATA WAREHOUSE SAP BW IN THE PRODUCTION COMPANY

Maria Kowal, Galina Setlak

Abstract: *in this paper the implementation of Data Warehouse SAP BW in the Production Company is presented. A data warehouse is designed for efficient query processing. The technical environment and data structures are optimized for answering business questions. Data warehouse is stored persistently by time over a particular time period. Users can use comprehensive analysis tools to access data. We will try to answer the question how to implement the SAP Business Warehouse in a company*

Keywords*: information system, data warehouse, business intelligence, data mining*

ACM Classification Keywords*: H. information system*

Introduction

Continuous innovation in data processing possibilities is a reason why more information is stored in a more detailed form. There is the need to both reduce and structure this data so it can be analyzed meaningfully.

A data warehouse can help to organize the data here because it brings together all operative data sources with different degrees of detail in order to provide this data in a scalable form to the whole organization. A data warehouse is designed for efficient query processing. The technical environment and data structures are optimized for answering business questions. Data warehouse is stored persistently by time over a particular time period. Users can use comprehensive analysis tools to access data. These tools offer a user-friendly interface which simplifies query creation. End users have read-only access, meaning that the data is primarily loaded into the data warehouse via the Extraction, Transformation and Loading (ELT) process. A data warehouse is a copy of transaction data, specially restructured for analyses.

In this paper we will try to answer the question how to implement the SAP Business Warehouse in a company.

SAP Business Information Warehouse Structure and Components.

The data model in SAP Business Warehouse (Fig. 1) describes the data flow of the relevant business information from the source to the analyzable object in SAP BW. In modeling you can create and edit the objects and rules for the Administrator Workbench that are required for data transfer, updating and analysis. SAP BW provides a web-enabled, integrated view of information [BW305, 2005].

DataSource

Source Systems are sources that extract data for the SAP Business Information Warehouse The key element of the SAP BW information is the **DataSource** that contains a number of logically related fields arranged in a flat extract structure used to transfer data into the data Warehouse.

It is transferred into the BW according to the type of source system. If the source system is a SAP system, the metadata is simply copied from the source system. There where the source system is a flat file, the metadata is defined directly in the BW.

On the request data is transferred from the source system into the SAP BW in the selection of DataSource fields that eventually contain information relevant to decision-making about a business process- so called transfer structure.

The data entering the SAP BW from source systems is stored in a transparent, relational database table in the SAP BW. This physical store is the Persistent Staging Area table (PSA table). It is created for every source system DataSource. The request data stored in the PSA table have the same format as the corresponding DataSource transfer structure with extra fields of Data Request ID, the data package number, and the data record number.

When you load data using the PSA table the data records are transferred directly to the transfer structure. Data can be written to data targets from the PSA table by using transformations.

There is a possibility to check or change incorrect data in the PSA table either manually or using a program.

This gives the opportunity to correct update errors or validate data before it is processed in BW InfoSource.

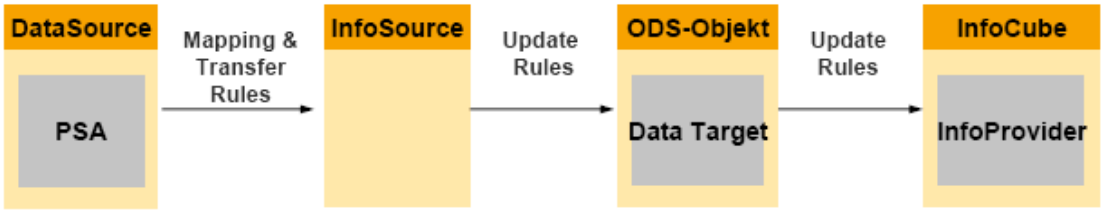


Figure 1. Data Model In SAP BW [BW305, 2005]

Mapping and Transformation

InfoSource is a quantity of logically related InfoObjects summarized into one single unit. It contain either transaction data or master data with attributes, texts and hierarchies. Individual fields in the DataSource are mapped to the corresponding InfoObjects in the InfoSource by the transfer rules. These rules can be applied easily using formulas. In the maintenance to the transaction rules is also stored the information how the data from the DataSource is to be transferred into the InfoSource [Karine Allard, 2006].

Update rules

Update rules specify how data is updated from InfoSources into data targets. With InfoCube it must be specified for each key figure and the corresponding characteristics, for an ODS object, it must be specified for the data and key fields, and for an InfoObject it must be specified for the attribute and key fields. The update type allows a control whether a key figure or data field is updated in the InfoProvider.

There are tree types of update: no update, addition, minimum or maximum or overwriting. Various options of calculation methods are used to control whether and how a characteristic, key figure or a data field should be updated in the data target.

Creating a start routine executed for each data package at the beginning of the update is another option for more complex transformations, if the available calculation methods are not sufficient.

InfoObjects

InfoObjects are the smallest units in BW and divided into characteristics, key figures, units and time characteristics.

Characteristics are groups like, for example product, customer group or region. They determine the level of detail in which the key figures are managed in the data target.

Key figures deliver the values analyzed in a query. They can be quantities, amounts, numbers.

Units give the values of the key figures meaning like units for measure.

Time characteristics like date or year.

Master data is a characteristic that have attributes, texts, or hierarchies at their disposal and remains unchanged over a long period of time. Master data contains information's that are always needed in the same way. A hierarchy is always created for a characteristic and it gives characteristics a structure. Properties of key figure influence on the load process and query display. The basic of them are determination of the aggregation and exception aggregation for non-cumulative key figures and the number of decimal places in the query.

Data Targets

Data targets are the physical objects that stay unchanged during data modeling and when loading the data. Data targets are Basic InfoCubes, ODS Objects, InfoObjects. There are two types of data targets:

- Data targets for which queries can be defined-so called InfoProviders.
- Pure data targets for which no queries can be defined, can only be used as a data source for another InfoCube, ODS object or InfoObject.

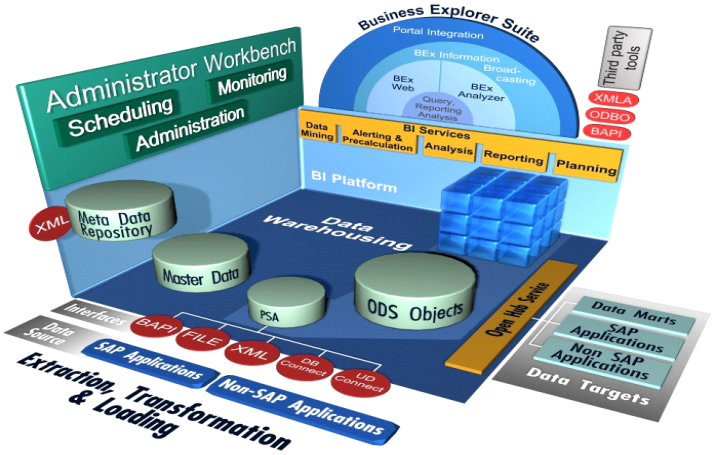


Figure 2. SAP BW Structure Overview [BW305, 2005]

InfoCube

InfoCubes are the main objects of the multi-dimensional model in SAP BW. Theirs structure is optimized for reporting. On those objects queries and analyses are defined and executed [Karine Allard, 2006].

There are following InfoCube types in *SAP BW*:

- BasicCube
- Virtual Cube:
- RemoteCube
- SAP RemoteCube
- Virtual InfoCube with Services:

InfoCubes group a set of InfoObjects (characteristics and key figures) together. A structure of a star schema optimized for reporting consist a number of relational tables and a fact table in the center. The fact table and the dimensions are linked to each other using identifying, abstract numbers.

Multi-dimensional data model are used for the creation of data warehousing or analytical applications OLAP applications. The classic star schema is the most frequently used multi-dimensional model for relational

databases. It classifies two groups of data: **facts** and **dimension attributes.** Logically related dimension attributes are stored as a hierarchy within the dimension table. The dimension tables are linked relationally with fact table by foreign or primary key relationships. The dimensional attribute with the finest level of detail of the corresponding dimension table is a foreign key in the fact table. All data records in the fact table can be identified uniquely.

ODS Objects

An ODS object acts as a storage location for consolidated data. The data in ODS objects is stored in transparent, flat database tables and can be updated with a delta update into InfoCubes or other ODS objects.

InfoProvider

An InfoProvider is an object for which queries can be created or executed in the Business Explorer.

InfoProviders are the objects or views that are relevant for reporting.

However, objects that physically contain data also fall under the term InfoProvider. These are also

called data targets:

- InfoCubes
- ODS Objects
- InfoObjects (characteristics with attributes, texts or hierarchies)

SAP BW Source Systems

The SAP has a wide functionalities to support extract of data at file or database level. Sources that can be used for transferring data to SAP BW (is show in figure 3).

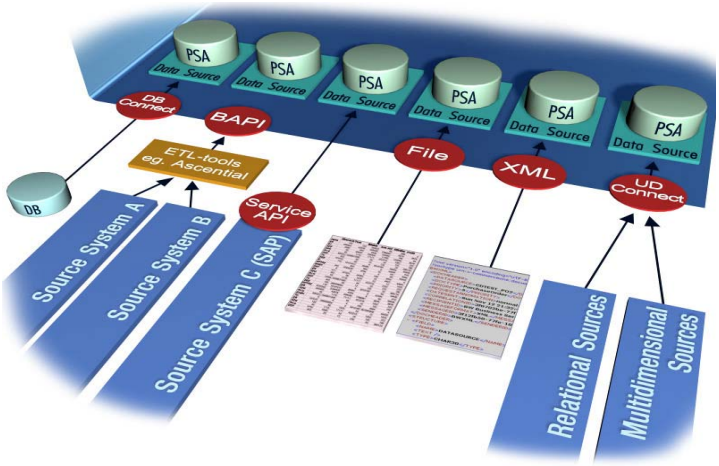


Figure 3. Openness of SAP BW in relation to sources supported for data transfer [BW305, 2005]

- **Transferring Data from SAP Source Systems**

SAP application-level components like SAP R/3, mySAP CRM or mySAP Supply Chain Management integrate with SAP BW using the BW Service API (SAPI) technology package.

- **Transferring Data Between Data Targets Within a SAP BW Data Mart Interface**

The data mart interface is a functionality that makes it possible to update data from one data target into another. It's possible to exchange the data between several BW Systems, between BW systems and additional SAP systems – for example APO systems, BW system (myself connection).

Data marts can be used to save a subset of the data for a data warehouse in another database, better manageability and maintenance or intentionally redundant segments of the global system.

- **Transferring Data from Flat Files**

SAP BW offer very easy transfer of data from flat files, ASCII, CSV format (Comma Separated Value).

Flat file data may be transferred to the SAP BW unchanged from a application server using a file interface. Defining and updating metadata for flat files must be done manually and have to correspond to data target in SAP BW.

Flexible updating from flat files data is possible because of maintaining the transfer structure. Preview able you to check, whether the metadata definition in BW and the structure file mach to each other.

- **Transferring Data Based on the Simple Object Access Protocol (SOAP)**

Data transfer in SAP BW uses a data request sent from SAP BW to the source system. Sending data through external controls is called a data push and take place using transfer mechanisms. The transfer can: use the SOAP service of the SAP Web AS, A Web Service or SAP XI. [BW310, 2005] .

Inbound queue occurs data push in SAP BW. For DataSource, so called DataSource with SOAP connection, with interface to supply the delta queue, system generates an RFC-compatible function module

- **Transferring Data from a System Using Third-Party ETL Tools . Staging BAPIs**

To enable the extraction of data SAP BW offers open interfaces, the staging BAPIs at application level. BAPIs are the programming interfaces that enable external access to the business processes and data.

Third-party tool loads the data from the external system and transforms them to SAP BW format.

- **Transferring Data from Database Management System Tables / Views - DB Connect**

DB Connect is used to datatabase connection as to source systems.

Functionality Multiconnect enable to open database connections in addition to the SAP default connection and use these connections to access external databases. DB Connect allows you to load data by linking a database to the BW generate a DataSource.

- **Transferring Data with UD Connect**

UD Connect (Universal Data Connect) give the opportunity of access both non-SAP data

sources and SAP data sources using the SAP Web AS J2EE connectivity. Using UD Connect, allows to

connect a lot of relational and multi-dimensional data sources to SAP BW. UD Connects transfers the data as flat data.

- **SAP BW as Source System**

The Open Hub Service is available for making opportunity for BW system to act as a source system for additional BW systems.

Performance-Optimized Data Target Modeling

There are a number of functionalities in modeling data targets to support you in improving the load and query performance of your BW system. Optimizing Performance for InfoCubes [BW310, 2005] .

- **Partitioning**

Partitioning is used to split up the whole dataset into couple small and redundancy-free units. That's cause increasing performance.

- **Database and Aggregate Table Indexes**

Checking and repairing the table's indexes improves load and query performance. Indexes created in the fact table for each dimension improves search process.

- **Database Statistics**

The query performances optimized by the database statistics. Update the statistics should happen every time when more than a million new records appears in the InfoCube.

- **Compressing InfoCubes**

Summarize a requests and insert them at the same time also increase the performance.

Implementation of SAP BW in a Company

Every company wants to structure their Business Warehouse according to the analysis of reporting requests and the data model. Task for the administrators is to examine the Business Content to see which standard Content objects can be used for the data warehouse as well as which additional, customer-specific settings have to be implemented.

The Administrator Workbench is the central point in data warehouse SAP BW responsible for managing processes [SAP BIW, 2005]. Data Warehouse Management in the SAP BW includes the maintenance of authorizations, the metadata repository, transporting, tools for activating SAP Business Content, technical content, an analysis and repair environment, and tools required for process management.

Authorizations

An authorization allows a user to carry out a certain activity in the Business Information Warehouse. Each authorization refers to an authorization object and defines each field that is contained in the authorization object. Authorization profiles are made by system administration by summing of a individual authorizations.

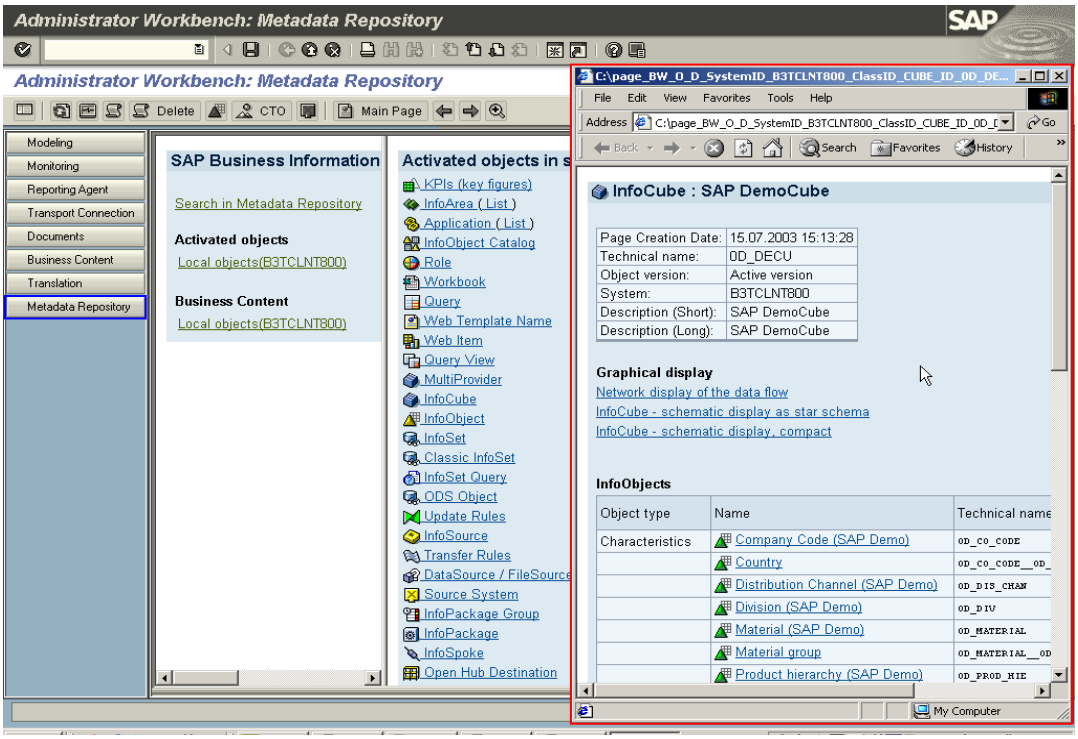


Figure 4. Areas of the Administrator Workbench metadata repository and information on the SAP DemoCube InfoCube [BW310, 2005]

For administrators special authorizations is needed in the BW and in the source system, BW roles (is show in Fig. 4).

- BW Administrator (development system): maintain the source system, upload Metadata, execute queries for the statistics InfoCube and maintain aggregates.
- BW Administrator (production system): maintain the connection to the source system and execute queries for the statistics InfoCube.
- Modeler (development system):design the data targets and InfoProvider, InfoObjects, InfoSources and the data flow, define communication structures and transfer and update rules.
- Operator (production system): upload data from the source, system and monitor the results.
- Reporting Developer (development system): design the queries for the reports
- Reporting User: execute queries using the BEx Analyzer or in the Web.

Metadata Repository

Metadata repository allows to display information on the metadata objects. This metadata includes object properties and their relationships to other objects. The available functions in the Metadata Repository are:

- Displaying information about active objects in the BW system or on the Business Content objects
- Various formats of displaying the information about metadata objects graphically.
- Search in the metadata repository.

Business Content

Business Content gives a lot of opportunities for saving costs and time. Technical and content-prerequisites shorten the implementation process and InfoCubes are already optimized for data storage and analyses. Predefined analyses are the examples of content in projects. User who use prototyping help can correspondingly specify which information is still missing for the project configuration.

Business Content collects information and prepares it for next use. It's based on an data model that's preconfigured and based on tasks stemming from consistent metadata.

Business Content collect's SAP and non-SAP extractors, DataSources (extraction structures), InfoObjects, InfoSources, InfoProviders (for example, InfoCubes and ODS objects), Queries, KPI's and Templates.

Business Content can be used without any adjustments. It can also can be adjusted by means of enhancements and served as a template for customer-specific objects.

Technical implementation in SAP BW

After activating a master data-carrying characteristic, master data tables (attributes, text, hierarchies) are generated in the characteristic maintenance. Before creating and subsequently activating the characteristic InfoObject. a text table is created when the *With texts* checkbox is flagged.

To store the hierarchical relationships between characteristic values the hierarchy table (H table) is used SID Tables shows the connection between master data-carrying characteristics, navigation attributes and external hierarchies.

To use an InfoSource with Direct Update to Load Attribute Data from an SAP R/3 System there's four steps to do:

Set the Global Transfer Routine, create an InfoSource with Direct Updating, create a Generic DataSource and create and Activate Transfer Rules.

To load transaction data from an SAP R/3 System ther's a schedule do to:

- **Create Update Rules**

DataSource need to be preconfigured to load transaction data from the SAP source system. A lot of combination is available in Business Content. Connection between InfoSource and InfoCubeis possible because of set of update rules.

- **Create InfoPackage and Load Data**

Extraction of data from the source system to InfoCube must be precedent by create an InfoPackage containing the parameters for the data extraction.

After activating a master data-carrying characteristic, master data tables (attributes, text, hierarchies) are generated in the characteristic maintenance.

Before creating and subsequently activating the characteristic InfoObject. a text table is created when the *With texts* checkbox is flagged.

To store the hierarchical relationships between characteristic values the hierarchy table (H table) is used SID Tables shows the connection between master data-carrying characteristics, navigation attributes and external hierarchies

Conclusion

Reporting, analysis and interpretation of business data is a central focus for companies that wants to guarantee competitiveness, optimize processes and to be able to react quickly and in line with the market. **SAP Business Information Warehouse (SAP BW)** as a core component of SAP NetWeaver data warehousing functionality, provides both a business intelligence platform and a suite of business intelligence tools.

By using SAP BW companies can easily and quickly get the tool that will meet all theirs requirements. Business Content which contains a great variability of an Info Objects works great in a most of modules of company like Sales, Purchasing, Logistic and Production Planning.

Bibliography

[BW305, 2005] BW305 Business Information Warehouse Reporting & Analysis my SAP Business Intelligence Kurs SAP Course Version: 2005 Q1

[BW310, 2005] BW310 Data Warehousing SAP NetWeaver Course Version: 2005 Q1

[Karine Allard, 2006] Karine Allard: Basic BW Navigation, 2006

[SAP BIW, 2005] SAP Business Information Warehouse Functions in Detail Opis funkcji stosowanych w zapytaniach.

[Inc Paolo Alto,2007] SAP BW Reporting Made Easy SAP Labs, Inc Paolo Alto,California, 2007

Authors' Information

Maria Kowal, *PhD Student, M.Eng., WSK PZL Rzeszów, marysia.kowal@gmail.com*

Galina Setlak, *Ph.D., D.Sc, Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science , W. Pola 2 Rzeszow 35-959, Poland, Phone: (48-17)- 86-51-433, gsetlak@prz.edu.pl*

INFORMATION MODELS OF CUSTOMERS IN CRM AND E-CRM SYSTEMS

Justyna Stasieńko

***Abstract**: The following article deals with customer personality in the CRM systems. Its innovatory approach consists in the method of segmentation performed according to one of psychological factors - personality. The analysis is based on the assumption that the class of customers can be divided into separate types. This is possible due to the preservation of a large internal homogeneity within every type (homogeneous groups) and a considerable differentiation among types. The article argues that the classification of customers should be based on the criterion which has an essential influence on their behavior. Such a criterion is the customer personality. Accordingly, this division can make up a good basis for the CRM strategy: 1) the elaboration of several diverse scripts of the customers service suitably to distinguished groups; 2) the individualization of the service. It can be concluded that such a modeling will allow to supplement the CRM programmes with a module of an individual and institutional customers practical service.*

***Keywords:** CRM, customer's personality, information model.*

***ACM Classification Keywords**: K.6 Management of Computing and Information Systems - K.6.0 General Economics*

Introduction The role of a customer at organization

Till the 1990s , the core interest of the organization in endeavour to the achievement of the profit, was a product [Tiwana, 2001a]. It has been recognized that the product is insufficient to compete on the market. Organizations began to perceive the customer's role. Recently, the customer has become the most valuable commodity, organizations strive for. It was noticed that the success of the organization depends on the individual approach to the customer, the building of a solid bond with them and the effective process of mutual communication [Lotko, 2003].

One of the main objectives facing the organization is building a bond with the customer. The relation with the customer makes up a process consisting of three phases [Mazur, 2001], [Tiwana, 2001b], [Kralewski, 2001] all of them aiming at the exchange of supplies between the organization and the customer.

The initial phase is winning over new customers and reaching new groups of customers applying suitable relation strategies. The next phase is improving and deepening relations with customers, working out the profit growth strategy and maintenance of the relation. In the final phase the signals foretelling breaking the relation are recorded and analyzed and preventive actions are taken to maintain the customers.

Long-term relation with customers depends on the rightness of management addressed at them as well as the proper selection of workers and tools. Such a management covers the analysis of the customers' real needs, collecting the information about them, disclosure and the circulation of the information about their needs and preferences. Management should result in proper functioning of the things, organization and people in accordance with the aim of the manager. It requires solving problems and decision taking regarding matters of various degree of complexity and diverse time horizons.

Managing customers relations means steering contacts that is planning, building, coordinating and controlling them. In other words, management is connected with meeting customers' expectations through getting to know and understanding their needs, using the knowledge, improvement of communication, offering the customer the product others do not have, building a bond between the company and the customer.

Gathering data from various sources aims at finding out clients' needs. Thanks to those data, a complex knowledge of the client is being created. The customer profile is formed on the basis of that knowledge. It is necessary to keep the profile's integrity through systematic collection and updating the data. Understanding customer needs is not easy and requires analysis. On a mass scale, it is usually done with the help of of Bussines Intelligence tools, segmentation and behaviour patterns. In the personalized view, the analysis of the profile is used and the observation of the most valuable customers. Contacts with the customer, questionnaires etc. contribute to the analysis of their preferences. All individual data about the customer are the foundation for starting personalization processes, and consequent actions aiming at tightening the customer's relation with the company. Improvement of the relation management follows together with the increasing number of essential data in the profile.

The individualization of the mass customer service, the minimization of costs connected with the customers service, the assurance of the customers loyalty are the main objectives in the customer relation management [Dejnaka, 2002].

Long-term maintenance of clients is, at present, the main objective. To reach it, works on creation of the suitable system began. As a result, certain customer relation management systems emerged.

The systems steered toward a customer

Customer relationship management (CRM) is a customer-focused business strategy that dynamically integrates sales, marketing and customer care service in order to create and add value for the company and its customers.

CRM consists of the processes a company uses to track and organize its contacts with its current and prospective customers. CRM software is used to support these processes; information about customers and customer interactions can be entered, stored and accessed by employees in different company departments. Typical CRM goals are to improve services provided for customers, and to use customer contact information for targeted marketing.

Knowledge about customers is consistent with this CRM. One of the most important features of CRM is satisfying the customer delivering him proper contents, products and services. Such an approach shows that the customer is an individual recepient who should feel as someone exceptional.

Accordingly, building knowledge about customers and propagating it at the company is the long-lasting and arduous process. However, it can bring quick effects with small costs. It requires the continuous process of learning. It is an important feature of correctly designed and realized approach to CRM in a company. Thanks to continuous winning knowledge about customers, the organization can continually outdistance the competition. It also allows to almost individual treatment of every one. It lets fit offers and ways of service to needs and the customers preference, and also conceal offers and ways of the working from the competition.

The existing systems of the CRM have the customer „flat model". The necessity of the development of the personalization of contacts (offers) is connected with the extension of the multidimensional profile of customers types. In particular, these models characterized customers according to real decission behaviours with omission of causes of such decisions. So well-founded and leaning on scientific bases of assembling criteria were lacked. Because of that there is the necessity of the investigations of possibility of wide-ranging characterizing the customer model.

The informative model of a customer

Researching and analysing the customer behaviours, ecomists in the modules of the systems of CRM class describing a customer, took into consideration general information (Id), preferences (Pr) and a value of a customer for the company (W). After some time, it turned out that not only these factors influenced the customer decision relating to a purchase. The essential meaning of psychological factors was perceived. Personality, the style of life, motives of a purchase, a susceptibility to risks, a creativity. further called („psychological factors" was redefined and replaced with the term) „personality".

Accordingly, the accomplished analysis [Stasieńko, 2007] of the individual and institutional customer allows to accept the thesis, that the personality favours him. The acquaintance of the customer personality plays an important part in the case of winning of the customers, building a relationship with them. However, an influence whether a customer buys a new product, uses new service, or not, depends on different psychological features, such as a susceptibility to risks and a creativity.

The personality influences information which are received from outside. For the reason

not only psychologists are interested in this but also marketing experts, programmers or designers of informative systems.

In the result of investigations [Stasieńko, 2007] the customer model was widen with a customer personality and it takes shape of:

$$\mathbf{IMOK := < Id, W, Pr, Os >} \quad (1)$$

where Id - general information,

W - the value of a customer for a company,

Pr - the preferences,

Os - the personality.

The model of the customer personality was constructed to reveal the profile of the customers personality, both individual and institutional and it has to serve in improvement of the service within a system of the relationship management with customers at the organization. They consider an individual customer (MOKI) and an institutional(MOKIS) separately.

The model of individual customer personality (MOKI) which is described with a relation, (2) is defined as an arrangement of four features: temperament, susceptibility to risk, creativity and personal culture.

$$\mathbf{MOKI := [T, R, In, Kos]} \quad (2)$$

where T - Temperament,

R - Susceptibility to risk,

In - creativity,

Kos - the Personal Culture.

The model of the institutional customer (MOKIS) *personality* also takes into account four variables: the dynamism of the organization, its susceptibility to risks, creativity and organizational culture:

$$\mathbf{MOKIS := [D, R, Ins, Ko]} \quad (3)$$

where D - dynamism,

R - susceptibility to risk,

Ins - creativity,

Ko - the organizational culture.

Comparing models MOKI and MOKIS one can perceive certain analogue and the general model of the customer (OMOK) personality can be emerged which can be defined by the dependence (4):

$$\mathbf{OMOK := [Dn, R, In, K]} \quad (4)$$

where

Dn - the Dynamics

Dn = (T, D), T - Temperament,

D - Dynamism

R - Susceptibility to the risk

In – Creativity

In = (Ind, Ins), Ind - the individual customer creativity

Ins – the institutional customer creativity

K - culture

K = (kos, Ko), Kos - personal culture,

Ko - the organizational culture.

The model of Opposite (lub Opposing) Values (MoWaP)

The models mentioned above suggest the necessity of finding a method of the division of complex community of the customers on homogeneous subsets with identifying opposite values. In the model of the opposite values (MoWaP) [Stasieńko, 2007].

In case of the two-dimensional model the solution consists in indication of two features describing the considered problematic area. In considerations over the customer personality two features were distinguished together with their extreme values They are: the interest in the surroundings and the way of the receiving of information.

In the studied space the model of the customer personality was divided into four quarters (Fig. 1). One of the axis - the axis X - the direction of perception and energy steered towards the world, the second - the axis Y - the way of receiving the information.

The axis X on the one pole assembles the values of extraversion features - emphasizing susceptibility to elasticity, risk and extraversion, dynamism (E); on the second - stability, caution, introversion (I). It means that there are customers who do not fear changes, adapt to them quickly and have large cognitive needs (E). From the other side there are careful customers in making a decision relating to new products and they are stable in their choices (I).

The axis Y includes values which describe customer communicativeness, their openness on advice, different prompts in making the decision, on supporting his arguments without a regard on circumstances and on low (U) or high cognitive needs (F).

MoWaP became a tool of customers segmentation. The extreme values of distinguished two features of the personality divided customers community into four definite types (fig. 1) and gathering of customer personality, which was not successful in qualifying. Every part received own name describing the most characteristic features of a customer. They were gathered from the literature relating to behaviours strategies [1]. Five types of customers were distinguished: UI – an OBSERVER, FI - CONSTANT, FE - EVOLUTIONARY, UE - DYNAMIC and a NONDEFINED type.

The left bottom quarter of the drawing (UI - an OBSERVER) distinguishes a careful customer, observing and copying developmental customers, slow in making a decision, a litlle bit active, acting without hurry, methodological, having small cognitive needs, although stable and loyal.

The left upper quarter (FI - CONSTANT) assembles careful customers, recognizing own rights, with large self-control, sedate and loyal.

The right upper quarter (FE - EVOLUTIONARY) describes the elastic customers who do not fear new challenges likes risk and novelty, with optimistic attitude towards the external world, making decisions quickly, with large cognitive needs, adjustment, independent, with the strong personality. However, they tend to be abrasive and they change their minds with obstinacy.

The right bottom quarter of the drawing (UE - DYNAMIC) presents open customers on new challenges, elastic in relation to susceptibility to risk, with optimistic attitude towards the external world , havings however small cognitive needs, they are educated, believing in their possibilities and skills, they are self confident. These customers, because of the large dose of energy, can appear less loyal.

The centre (NONDEFINED) accumulates the customers who were not successfully unambiguously classified

because of various reasons, for example because of the divergence of variables value (THE TEMPERAMENT - melancholic, interest in the external world - low (introvert), high susceptibility to risk and creativity). these are the subject of farther investigation of personality.

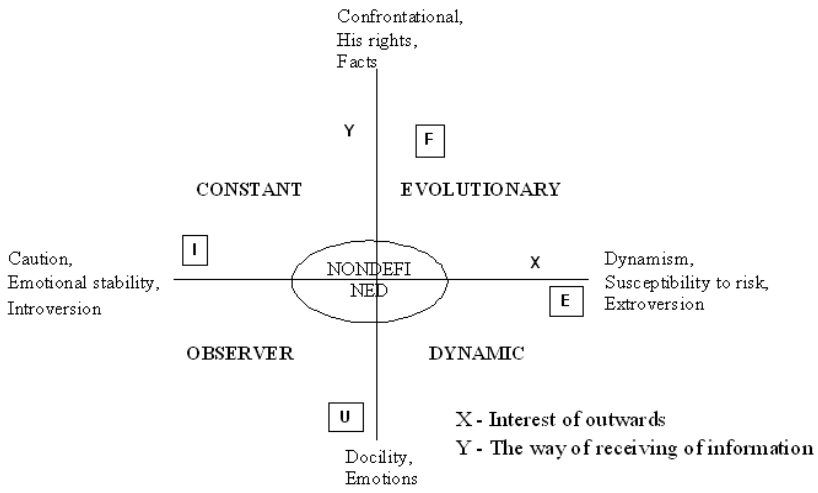


Figure 1. Model of customers division

MoWaP makes the division of customers community to distinguish homogeneous groups. It allows to build the common strategy of customers service within every profile and the differentiation of the strategy of the service in various groups.

The STRAT-SE module

The theoretical model of segmentation customers became the basis to build the STRAT-SE module supplementing CRM systems with the auxiliary tool in the range of recognizing individual and institutional customers, their psychological figure of customers and building on thie basis of the rational strategy of their service .

The structure of the module

To illustrate STRAT-SE - an expert system simulating the human reasoning process. was built on the basis of the OMOK and MoWaP models. The basic function of the module is to support of the CRM system in the range of the analysis of individual and institutional customers and to build the schedules of the service.

Basic tasks of SE are as follows:

- cognition of the customers personality and attributing each of them to the definite class in relation to the personality,
- the choice of the worked out strategy of the customer service.

STRAT- SE is built from two process units. First defines the customer changing personalities, the second - makes his segmentation, that is it classifies one of the customer five personality types.

The structure of STRAT-SE is placed in standard SE which contains [Stefanowicz, 2002]: the dialogue module (user interface) - the user services, the “knowledgebase”, the knowledge engineering.

The function of the user interface is to present questions and information to the user and supply the user's responses to the inference engine. It allows for co-operation and communicating the customer with the system. After preliminary announcements, linking the contact with the customer, the module represents the pattern of the description in the shape of a questionnaire which contains the series of questions relating to customers personalities. The questions were drawn up in the form of closed statements. The answers are given by the customer are the facts consisting in the entrance vector, which for STRAT-SE, makes up the basis to the beginning the process of the analysis of the customer psychological figure. After answering all questions from the

pattern, the system gives the values of variables and the type of the customer personality.

The knowledge gathered about the customer is written down in the figure of rules. “knowledgebase” is divided into two subsets: the base of facts and the base of rules.

The base of facts contains the information in the shape of entry vectors on the subject of the customer personality, gathered on the basis of the pattern of the customer description. Every such vector consists of data which are values of all changing enterings which are included in the OMOK model, necessary to the qualification of the customer personality. It is concluded that, in case of lack of an answer from a customer, the entrance vector is replenished by the organization staff on the basis of the observation of the customer behaviors.

The base of rules contains rules letting to qualify the customer personality on the basis of facts accumulated in base of facts and to execute classifying every customer to one of five formed groups in MoWaP in the consequence. The whole base of rules is divided in two gatherings: the subsets of preliminary rules and the subset of stratification rules.

The experimental module was built on the platform of the skeletal expert system EXSYS CORVID of the firm EXSYS Inc.

The working of the system STRAT_SE leaned on the system called EXSYS CORVID consists in the creation of the automatic and interactive session which copies consultations with an expert – a psychologist in the opinion of the personality and helps the process of the classification .

STRAT-SE describes the process of solving the problem in the shape of rules. The inference engine uses these rules to the automatic qualification which information (facts) is necessary and what will be the result of individual events, and then represents the solution. The system manages the rules which make possible the qualification of customer (his variables) personality and his division according to the principles of the MoWaP model.

The system gives next questions from the frame. On the basis of answers given by the customer, the system generates an answer with the profile of the personality. On the screen changing personalities are shown (though they can be they skipped) and the type of the customer who was settled by the system on the basis of segmentation rules.

The verification of the STRAT-SE module

To review the trial of pilotage introducing of STRAT-SE was undertaken. The main aim was examining the legitimacy of the model.

1. 118 persons were the subject of studies who filled the questionnaire. The questions and the rules of inference in the base of knowledge were worked out according to psychologists hints.

The larger part of studied (98 people - 83%) filled the questionnaire directly in STRAT-SE. Remaining 17% (20 persons) – filled it in the paper version. The gathered data was introduced to the STRAT-SE system SE.

The system qualified the changing personalities of every one from studied, it qualified to what type of the personality the given person can be subordinated to. The results - the type of the personality – the gathered knowledge was passed to these studied people. All studied people could take an attitude to the result passed by the system: accept it or question its legitimacy. 80% persons agreed with the result settled by STRAT SE. Remaining 20% had total or partial (they agreed with some changing values passed by the system) restrictions.

It should be noticed however, that classifying some customers to the different type of the personality, than they judge about themselves, does not disqualify the STRAT-SE system. It can be resulted from the different knowledge on the subject of one's personality.

2. The result in the figure 41% customers described as UNDEFINED should be treated as a signal underlining the necessity of the supplement of the base in knowledge about additional classifying rules and to continuing researches towards the direction of greater depth of the analysis of customers personality by inclusion to the model of more features or enlargement of the number of questions defining THE TEMPERAMENT (in this case there were the most undefined values - the temperament UNDEFINED). This shows the fact that 13 questions

are insufficient to qualify the customer personality. The investigations of the temperament or personalities led by psychologists contain tens questions. The number of questions studying the temperament should be enlarged. THE TEMPERAMENT - UNDEFINED shows the most often on the customer about the UNDEFINED type. Because of that there is so high (41%) percentage.

Figure 2. illustrates customers studied according to their personality settled by STRAT-SE.

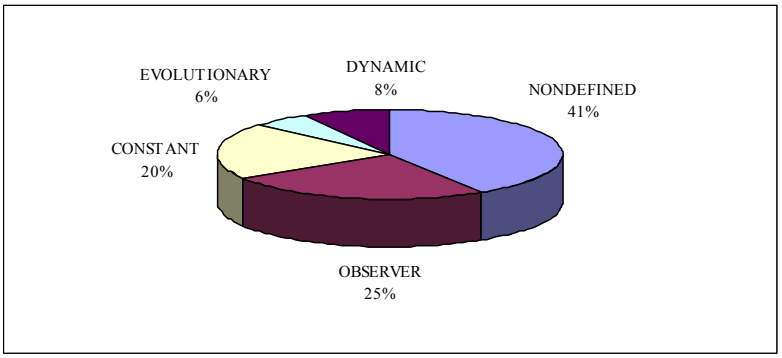


Figure 2. According to the types of the personality in the studied sample customers schedule.

The conclusions were formulated towards the development of the advanced system of the customer service. These conclusions (introduced below) focus particularly on the part of the STRAT-SE module.

1.Conducted experiments on the STRAT-SE module showed that applied in it models made up the good basis to customers segmentation according to the types of their personality.

2. Using of the STRA-SE module showed two main problems:

Difficulty with gathering information about customers: they are not always willing to devote several minutes on the fulfillment of the questionnaire. It is necessary to prepare workers having contact with the customer to their observation and building on this basis the profile of the customer personality.

Existence of numerous groups of customers which have UNDEFINED profile of the personality. It is probably connected with a small number of questions relating to one of changing personalities - THE TEMPERAMENT. Experts formulate tens questions in psychological tests. In the case of STRAT-SE system there are 8.

The application of CeReMOs

After the pilotage results of the tests of the STRAT-SE module, CeReMos application was built

CeReMOs is leaning on the OMOK model which except typical data processed in existing CRM systems appeals to the customer changing personalities.

It is underlining that workers and the owners of a company were satisfied with the improvement of customer service. It confirms one of the main thesis of the work / the investigation, that the theoretical OMOK model can be a helpful instrument, enriching the functions of the systems of customer relationships management.

Three main informative structures were distinguished within CeReMOs: general data, data connected with the personality, the schedules of the service.

The application CeReMOs came into being in the environment of Lotus Notebook. It can function in the conditions of the traditional and electronic economy in relationship with the generality of this software. The possibility of the existence on many equipment platforms and operating systems is the essential advantage of this software.

The conducted introducing works of the application CeReMOs leaning on the model IMOK show that CeReMOs can make up the part of the integrated CRM system of the customer service. It is connected with already existing system which has data about a customer, history of his transaction, orders, accounts, strategy of the behavior etc. CeReMOs can make up an independent module in less advanced systems, in which it takes over the whole management with customers .

Conclusion

To sum up, the aim of every organization playing the essential part on the competitive market, is preservation of financial fluency, winning loyal customers, the assurance of the dynamic development of the company, consolidation of the a good mark, enlargement of the profits and the assurance of the satisfaction among workers. The proposed application CeReMOs goes towards the realization of the majority of these aims. The method of the division of customers (The Model Opposite Value) seems to be obvious, what makes it much more convincing.

Bibliography

[Cameron, 2003] K.S. Cameron, R.E. Quinn. Kultura organizacyjna – diagnoza i zmiana. Model wartości konkurujących. Oficyna Ekonomiczna, Kraków 2003.

[Dejnaka, 2002] A. Dejnaka. CRM Zarządzanie kontaktami z klientami. Helion, Gliwice 2002.

[Kralewski, 2001]] D. Kralewski: Customer Relationship Management: integracja procesów w celu budowy relacji z klientem. Human-Computer Interaction w reorganizacji procesów gospodarczych i tworzeniu zintegrowanych systemów informatycznych. 457-463, Gdańsk 2001.

[Lotko, 2003] A. Lotko. Zarządzanie relacjami z klientem. Politechnika Radomska, Radom 2003.

[Mazur, 2001] A. Mazur, K. Jaworska, D. Mazur. CRM. Zarządzanie kontaktami z klientami. Wydawnictwo MADAR, Zabrze 2001.

[Reynolds, 2002] J. Reynolds. A Practical Guide to CRM. CMP Books, New York 2002.

[Stasieńko, 2004a] J. Stasieńko Customer's profile in the process of Customer Relations Management, Proceedings of the International Conference Mass Customization and Personalization, Rzeszów 2004.

[Stasieńko, 2004b] J. Stasieńko. Aspekty modelowania profilu klienta w CRM. Multimedialne i Sieciowe Systemy Informacyjne, Materiały IV Krajowej Konferencji MISSI 2004, Oficyna Wydawnicza Politechniki Wrocławskiej, s. 385 - 394, Szklarska Poręba 2004.

[Stasieńko, 2007] J. Stasieńko. Informacyjny model klienta w systemach klasy CRM i e-CRM. Doctoral thesis Warsaw School of Economics, Warszawa 2007.

[Stefanowicz, 2002] B. Stefanowicz. Sztuczna Inteligencja i Systemy Eksperckie. Oficyna Wydawnicza SGH, Warszawa 2002.

[Tiwana, 2001a] A. Tiwana. The Essential Guide to Knowledge Management. Prentice Hall PTR 2001a.

[Tiwana, 2001b] A. Tiwana. Przewodnik po zarządzaniu wiedzą e-biznes i zastosowania CRM. Wydawnictwo „Placet", Warszawa 2001b.

Authors' Information

Justyna Stasieńko *- The Bronisław Markiewicz State School of Higher Vocational Education, Czarnieckiego Street 16, 37-500 Jarosław, Poland; e-mail: nieyoul@sgh.waw.pl*

ОЦЕНКА КРЕДИТНЫХ БАНКОВСКИХ РИСКОВ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКОЙ ЛОГИКИ

Юрий Зайченко

Abstract*: The problem of credit risk evaluation is considered in this paper. The classical method of credit risks evaluation based on borrowers credit rating determination is presented. A novel method of credits rating determination based on fuzzy logic systems is suggested {Mamdani algorithm)/Fuzzy rules for credit rating determination and general class of borrowers are formulated. The applications of the suggested method fuzzy logic for practical problems of credit risk evaluation are presented and discussed in this paper.*

Keywords*: credit risk evaluation, credit rating, fuzzy logic, credit management*

Введение

Задача анализа кредитоспособности заемщиков банков и оценки кредитных рисков представляет значительный интерес в связи тем, что в последние годы в Украине наблюдается бум в сфере кредитования. Поэтому для того чтобы соответствовать требованиям Национального Банка Украины и рекомендациям Базельского комитета по банковскому надзору (попросту говоря, не разориться самим и е разорить своих клиентов), банкам необходимо иметь тщательно разработанную политику по управлению рисками. Одной из важнейших составляющих данной политики является набор профессионально подобранных, с учетом экономического состояния страны и специфики деятельности банка, статистических и (или) математических методов оценки ожидаемых и неожидаемых потерь.

Целью данной работы является рассмотрение и сравнительный анализ методов оценки кредитоспособности заемщика:

- классический анализ кредитоспособности (четкий метод)
- анализ с помощью нечеткой логики, используя алгоритм Мамдани.

Элементы теории рисков

Деятельность любого предприятия неразрывно связана с понятием «риск»: банк, в котором вы держите свои денежные средства, может обанкротиться, деловой партнер, с которым заключена сделка, - оказаться недобросовестным, а сотрудник, принятый на работу, - некомпетентным. Не стоит забывать и о стихийных бедствиях, компьютерных вирусах, экономических кризисах и других явлениях, способных нанести урон компании. Вместе с тем рисками можно управлять так же, как процессами производства или закупки материалов.

Рассмотрим общее понятие риска.

Несмотря на то, что вся жизнь и деятельность людей происходит в условиях неопределенности, которая порождает риск, среди исследователей до сих пор не существует единственного мнения, касательно определения риска. Приведем простейшие из них.

Риск – возможная неудача.

Риск – неуверенность в возможном результате.

Риск – угроза неблагоприятного результата.

Риск – способ действия в условиях в непонятной или неопределенной обстановке.

Риск – событие, которое может произойти или не произойти.

Итак, Риск – объективно-субъективная категория, которая связана с преодолением неопределенности, случайности, конфликтности в ситуации неизбежного выбора и отображает степень достижения ожидаемого результата.

В зависимости от сферы деятельности, деловой среды, стратегии развития и других факторов компания может сталкиваться с различными видами рисков. Тем не менее, существуют общие цели, достижению которых должен способствовать эффективно организованный процесс управления рисками.

Кредитный риск – это возможность потерь вследствие неспособности или нежелания контрагента выполнить свои контрактные обязательства.

Данный риск присутствует во всех видах деятельности банка, где результат зависит от действий контрагента, эмитента или заемщика. Он возникает каждый раз, когда банк предоставляет (берет обязательство о предоставлении), инвестирует средства или другим образом рискует ими согласно условиям договоров независимо от того, где отображается операция – в балансе или вне его.

Для кредитора последствия невыполнения обязательств измеряются потерей основной суммы задолженности и невыплаченных процентов за вычетом суммы восстановленных денежных средств.

Целесообразно кредитный риск разделить на *индивидуальный* и *портфельный*.

Источником индивидуального кредитного риска является отдельный контрагент банка. Соответственно, оценка индивидуального кредитного риска предусматривает оценку кредитоспособности отдельного контрагента, т.е. его индивидуальную способность своевременно и в полном объеме рассчитаться по принятым обязательствам.

Портфельный кредитный риск проявляется в уменьшении стоимости активов банка. Источник портфельного кредитного риска – совокупная задолженность перед банком по операциям, которым свойственен кредитный риск (кредитный портфель, портфель ценных бумаг, портфель дебиторской задолженности и т.п.). Оценка портфельного кредитного риска предусматривает оценку концентрации и иверсификации активов банка.

Источником индивидуального кредитного риска является отдельный контрагент банка. Соответственно, оценка индивидуального кредитного риска предусматривает оценку кредитоспособности отдельного контрагента, т.е. его индивидуальную способность своевременно и в полном объеме рассчитаться по принятым обязательствам.

Портфельный кредитный риск проявляется в уменьшении стоимости активов банка. Источник портфельного кредитного риска – совокупная задолженность перед банком по операциям, которым свойственен кредитный риск (кредитный портфель, портфель ценных бумаг, портфель дебиторской задолженности и т.п.). Оценка портфельного кредитного риска предусматривает оценку концентрации и диверсификации активов банка.

Одной из основных и нетривиальных задач в оценке риска – это задача вычисления вероятности дефолта заемщика (PD). Распространены два подхода к вычислению PD. Первый основан на качественной и количественной оценке рейтинга заемщика по его внутренним финансовым показателям и особым бизнес-факторам. Второй основан на капитализации заемщика на фондовом рынке и уровне его долгов перед кредиторами. К сожалению, второй подход, хоть и является наиболее объективным, применим лишь к очень малому числу украинских открытых компаний. Рассмотрим некоторые методы первого подхода, основанного на качественной и количественной оценке рейтинга заемщика по его внутренним финансовым показателям и особым бизнес-факторам.

Классический метод анализа риска банкротства

Рассмотрим распространенный классический подход к анализу кредитоспособности заемщика, основной целью которого является классификация потенциальных заемщиков по степени риска неплатежеспособности. Количество групп риска выбирается произвольно с учетом уровня градации, до которого необходимо осуществить распределение кредитов (например, 10, 8, 6 групп и т. п.) При этом нежелательно выбирать большое количество групп, поскольку грани между ними стираются, и возникают трудности при классификации ссуд.

Согласно этому подходу, процесс оценивания кредитного риска заемщика заключается в оценке:

- Кредитоспособности заемщика;
- Риска кредитного продукта.

При **анализе кредитоспособности заемщика** можно выделить несколько этапов:

1. Первый этап (самый важный)

Комплексный анализ финансового состояния заемщика, который включает:

1) анализ структуры активов и пассивов (заемные средства предприятия группируются по степени их напряженности, а активы — по степени их ликвидности. Осуществляется оценка изменений структуры активов и пассивов в динамике),

2) анализ денежных потоков,

3) анализ финансовой устойчивости предприятия,

4) оценка эффективности деятельности предприятия.

Для оценки финансового состояния предприятия используются следующие основные аналитические коэффициенты:

1. *Коэффициент текущей ликвидности или коэффициент покрытия* =

$$\frac{\text{Оборотные средства}}{\text{Краткосрочная кредиторская задолженность}}$$

(характеризует платежеспособность предприятия в период, равный средней продолжительности одного оборота краткосрочной дебиторской задолженности)

2. *Коэффициент быстрой ликвидности* =

$$\frac{\text{Денежные средства + Дебиторская задолженность + Прочие ликвидные активы}}{\text{Краткосрочная кредиторская задолженность}}$$

(показывает возможности предприятия по погашению текущих обязательств за счет наиболее ликвидных активов).

3. *Коэффициент напряженности обязательств* =

$$\frac{\text{Краткосрочная кредиторская задолженность}}{\text{Дневная выручка от реализации}}$$

(указывает на долю текущих обязательств в дневном обороте заемщика).

4. *Коэффициент финансовой независимости* =

$$\frac{\text{Собственные средства}}{\text{Заемные средства}}$$

(характеризует уровень собственных средств предприятия).

5. *Коэффициент обеспеченности покрытия* =

$$\frac{\text{Собственные оборотные средства}}{\text{Оборотные средства}}$$

(показывает долю собственных оборотных средств в оборотном капитале)

6. *Коэффициент рентабельности продаж* =

$$\frac{\text{Прибыль от реализации}}{\text{Выручка от реализации}}$$

(показывает эффективность основной деятельности предприятия за данный период времени).

7. *Коэффициент оборачиваемости* =

$$\frac{\text{Всего активов}}{\text{Выручка от реализации}}$$

(показывает скорость оборота активов предприятия за определенный период времени).

Рассчитанные значения коэффициентов для потенциального заемщика сравниваются со среднеотраслевыми или нормативными значениями. По результатам сравнения делается вывод о текущем финансовом положении заемщика.

2. второй этап

Анализ деятельности высшего управленческого звена предприятия-заемщика посредством экспертной оценки уровня его профессионализма, компетентности, используемых стратегий и достигнутых результатов.

3. третий этап

Анализ отрасли, к которой принадлежит данное предприятие, а также позиции предприятия на рынке по отношению к конкурентам по отрасли.

4. четвертый этап

На завершающем этапе оценивается так называемый «страновой риск», т.е. проводится анализ текущей ситуации и перспектив развития в экономической и политической сферах, возможных изменений политики регулирующих органов и т. д.

На основании проведенного комплексного анализа заемщик относится к одной из групп риска, другими словами заемщику присваивается **кредитный рейтинг**.

Кредитный рейтинг представляет собой интегральную оценку финансовой устойчивости и платежеспособности страны, заемщика или отдельного кредитного продукта. Рейтинг выражает мнение агентства относительно будущей способности и намерения заемщика осуществлять выплаты кредиторам в погашение основной суммы задолженности и процентов по ней своевременно и в полном объеме.

Кредитные рейтинги отражают объективную оценку вероятности дефолта и используются для определения характера инвестиций.

После этого оценивается **группа риска кредитного продукта**. Следующие факторы влияют на группу риска кредитного продукта:

1. срок кредитного продукта

Чем меньше срок до погашения кредитного продукта, тем ниже риск, и наоборот.

2. ставка процента

3. условия предоставления кредита или ограничения, наложенные назаемщика при предоставлении кредита

4. Первостепенное влияние на группу риска кредитного продукта оказывает обеспечение по кредиту (гарантирует возврат, как правило, части денежных средств).

Оценка кредитных рисков с применением нечеткой логики

Рассмотрим метод оценки кредитных банковских рисков с использованием аппарата нечеткой логики, а именно алгоритма нечеткого вывода Мамдани. Для решения поставленной задачи введем лингвистические переменные, описывающие кредитоспособность заемщиков банковских кредитов.

кредитная история = (безупречная, плохая, нет данных)

Входная переменная: ответ на вопрос «Были ли просроченности у контрагента по оплате основного долга или процентов за ранее полученными кредитами, в т. ч. в других банках? »

Варианты: (не было, были, нет данных)

Соответствия: «не было - безупречно», «плохо- были», «нет данных - нет данных»

Способ получения данных: – ответ человека через диалоговое окно

Вид переменной: в данном случае - четкая.

Финансово-экономическое состояние = (стабильное, негативные тенденции, позитивные тенденции)

Вообще, для оценки финансово-экономического состояния используется три показателя – ”коэффициент независимости”, “моментальная ликвидность”, “период оборотности запасов”.

Для того, чтобы определить значение лингвистической переменной «**Финансово-экономическое состояние**», необходимо сравнить значения выбранных коэффициентов за текущий период и за прошедший период, т.е необходимо вычислить дельту для каждого коэффициента, причем считать рациональнее относительные изменения. Далее для простоты относительные изменения будут называться просто «дельты».

Следует обратить внимание, что при уменьшении показателей «коэффициент независимости» и «период оборотности запасов» рейтинг возрастает, а при уменьшении «моментальной ликвидности» рейтинг падает. Это учтено при составлении функций принадлежности для термов «негативные тенденции» и «позитивные тенденции».

Будем рассматривать „дельту” для каждого показателя как нечеткую переменную. Для нее необходимо ввести функцию принадлежности.

Рассмотрим дельту для каждого коэффициента в отдельности:

1) «коэффициент независимости »:

- сам коэффициент изменяется от 0 до 1 - дельта изменяется от -1 до 1

- функция принадлежности (ФП) дельты к терму «стабильный» (треугольного вида, ось абсцисс (дельта) пересекает в т. -0.05 и 0.05, а ось ординат в т. 1, (Рис. 1)

$$\mu(d)=\begin{cases}1+20d & ,если \quad -0{,}05 \le d < 0\\ 1 & ,если \quad d=0\\ 1-20d & ,если \quad 0 < d \le 0.05\\ 0 & ,если \quad d \le -0.05, d \ge 0.05\end{cases}$$

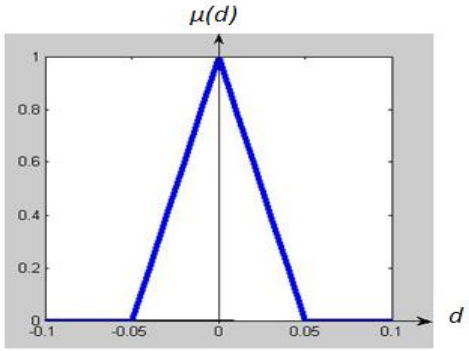


Рис.1

- ФП дельты к терму «негативные тенденции» (рис.2)

$$\mu(d)=\begin{cases}-20d & ,если \quad -0{,}05 \le d < 0\\ 1 & ,если \quad -1 \le d < -0.05\\ 0 & ,если \quad 0 \le d\end{cases}$$

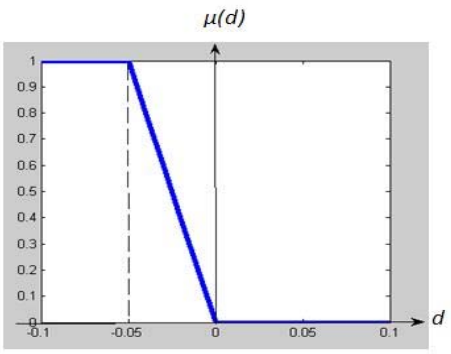


Рис.2

-ФП дельты к терму «позитивные тенденции» (рис.3)

$$\mu(d)=\begin{cases}20d & ,если \quad 0<d\le 0.05\\ 1 & ,если \quad 0.05<d\le 1\\ 0 & ,если \quad d\le 0\end{cases}$$

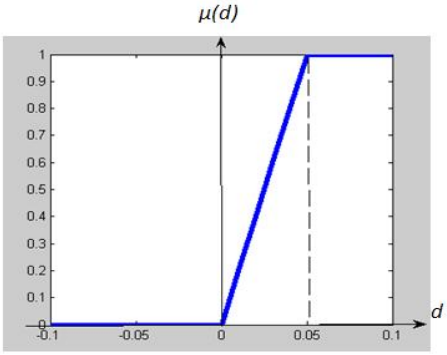


Рис.3

2) коэффициент «моментальной ликвидности»

- сам коэффициент изменяется от 0 до 1

- дельта изменяется от -1 до 1

-ФП дельты к терму «стабильный» (если нарисовать, получим треугольник, ось дельта пересекает в т. -0.1 и 0.1, а ось ординат в т. 1.)

$$\mu(d)=\begin{cases}1+10d & ,если \quad -0{,}1\le d<0\\ 1 & ,если \quad d=0\\ 1-10d & ,если \quad 0<d\le 0.1\\ 0 & ,если \quad d\le -0.1, d\ge 0.1\end{cases}$$

- ФП дельты к терму «негативные тенденции»

$$\mu(d)=\begin{cases}-10d & ,если \quad -0{,}1\le d<0\\ 1 & ,если \quad -1\le d<-0.1\\ 0 & ,если \quad 0\le d\end{cases}$$

-ФП дельты к терму «позитивные тенденции»

$$\mu(d)=\begin{cases}10d & ,если \quad 0<d\le 0.1\\ 1 & ,если \quad 0.1<d\le 1\\ 0 & ,если \quad d\le 0\end{cases}$$

3) «период оборотности запасов».

- сам коэффициент изменяется от 0 до 180

- дельта изменяется от -1 до 1

-ФП дельты к терму «стабильный» (если нарисовать, получим треугольник, ось дельта пересекает в т. -0.1 и 0.1, а ось ординат в т. 1.)

$$\mu(d)=\begin{cases}1+10d & ,если \quad -0{,}1\le d<0\\ 1 & ,если \quad d=0\\ 1-10d & ,если \quad 0<d\le 0.1\\ 0 & ,если \quad d\le -0.1, d\ge 0.1\end{cases}$$

-ФП дельты к терму «позитивные тенденции»

$$\mu(d)=\begin{cases}-10d & ,если \quad -0{,}1\le d<0\\ 1 & ,если \quad -1\le d<-0.1\\ 0 & ,если \quad 0\le d\end{cases}$$

-ФП дельты к терму «негативные тенденции»

$$\mu(d)=\begin{cases}10d & ,если \quad 0<d\le 0.1\\ 1 & ,если \quad 0.1<d\le 1\\ 0 & ,если \quad d\le 0\end{cases}$$

– **вероятность несвоевременного погашения** = (низкая, средняя, высокая, очень высокая)

Входная переменная: дает ответ на вопрос «вероятность несвоевременного погашения»

Варинты: (любая вероятность от 0 до 1)

Рассматриваем эту переменную как нечеткую и введем функцию принадлежности

- вероятность изменяется от 0 до 1

-ФП дельты к терму «низкая»

$$\mu(p)=\begin{cases}1-20p & ,если \quad 0\le p<0.05\\ 1 & ,если \quad p=0\\ 0 & ,если \quad 0.05\le p\end{cases}$$

-ФП к терму «средняя»

$$\mu(p)=\begin{cases}3.33-33.33p & ,если \quad 0.07\le p<0.1\\ 1 & ,если \quad 0.05<p<0.07\\ 0 & ,если \quad p\ge 0.1, p\le 0.05\end{cases}$$

-ФП к терму «высокая»

$$\mu(p)=\begin{cases}33.33p-2.33 & ,если \quad 0.07<p\le 0.1\\ -5p+2.5 & ,если \quad 0.3<p\le 0.5\\ 1 & ,если \quad 0.1<p\le 0.3\\ 0 & ,если \quad p\le 0.07, p\ge 0.5\end{cases}$$

-ФП к терму «очень высокая»

$$\mu(p)=\begin{cases}5p-1.5 & ,если \quad 0.3<p<0.5\\ 1 & ,если \quad 0.5\le p\\ 0 & ,если \quad p\ge 0.3\end{cases}$$

– **риск убытков** = (высокий, средний, низкий)

Входная переменная:

Варинты: (высокий, средний, низкий)

Способ получения данных: – ответ через диалоговое окно

Вид переменной: в данном случае - четкая. Т.е. Функция принадлежности принимает значение либо 0, либо 1.

5. состояние документации = (плохое, нормальное, хорошее)

Входная переменная: ответ на вопрос

«Состояние документации определено как: »

Варинты: (плохое, нормальное, хорошее)

Способ получения данных: –ответ через диалоговое окно

Вид переменной: в данном случае- четкая.

6. банкрот= (да, нет)

Входная переменная: ответ на вопрос «Признано предприятие банкротом»

Способ получения данных: – ответ через диалоговое окно

Вид переменной: в данном случае - четкая.

7. обеспечение = (первоклассное, ликвидное, неликвидное, отсутствует, возможность проблем)

Входная переменная: ответ на вопрос «Обеспечение»

Способ получения данных: –ответ через диалоговое окно

Вид переменной: в данном случае- четкая. **8. платёжеспособность** =(высокая, средняя, низкая)

Входная переменная: ответ на вопрос «Платёжеспособность контрагента»

Варианты: (высокая, средняя, низкая)

Вид переменной: в данном случае- нечеткая.

9. Функция принадлежности для выходной переменной

Приведем ФП для выходной переменной - рейтинг заемщика

-ФП к терму «класс Д»
$$\mu(r)=\begin{cases}5r+0.25 & ,если \quad -0.5<r<1.5\\ 0 & ,если \quad r\le -0.5, r\ge 1.5\\ 1 & ,если \quad r=1.5\end{cases}$$

-ФП дельты к терму «класс Г»
$$\mu(r)=\begin{cases}5r-0.75 & ,если \quad 1.5<r<3.5\\ 0 & ,если \quad r\le 1.5, r\ge 3.5\\ 1 & ,если \quad r=3.5\end{cases}$$

- ФП дельты к терму «класс В»
$$\mu(r)=\begin{cases}5r-1.75 & ,если \quad 3.5<r<5.5\\ 0 & ,если \quad r\le 3.5, r\ge 5.5\\ 1 & ,если \quad r=5.5\end{cases}$$

-ФП дельты к терму «класс Б»
$$\mu(r)=\begin{cases}5r-2.75 & ,если \quad 5.5<r<7.5\\ 0 & ,если \quad r\le 5.5, r\ge 7.5\\ 1 & ,если \quad r=7.5\end{cases}$$

-ФП дельты к терму «класс А»
$$\mu(r)=\begin{cases}5r-3.75 & ,если \quad 7.5<r<9.5\\ 0 & ,если \quad r\le 7.5, r\ge 9.5\\ 1 & ,если \quad r=9.5\end{cases}$$

Соответствующий вид ФП для различных классов рейтинга приводяться на рис.4.

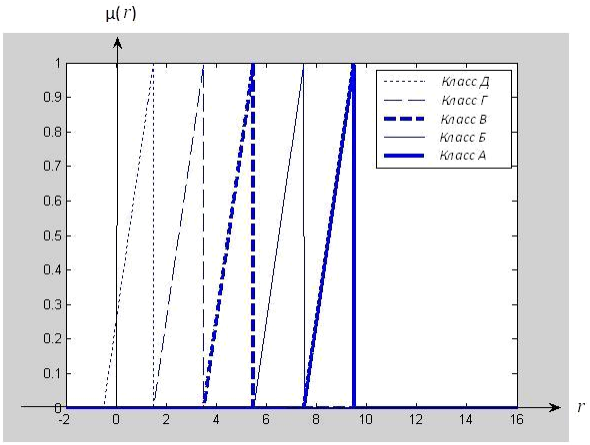


Рис.4

Правила для оценки рейтинга. Текстовая формулировка экспертных выводов

Для оценки класса заемщика на основе рейтинга используются следующие правила вывода

1.1 Если -0.50 ≤рейтинг<1.50, то рейтинг-класс = Д

1.2 Если 1.50 ≤рейтинг<3.50, то рейтинг-класс = Г

1.3 Если 3.50 ≤рейтинг<5.50, то рейтинг-класс = В

1.4 Если 5.50 ≤рейтинг<7.50, то рейтинг-класс = Б

1.5 Если 7.50 ≤рейтинг≤9.50, то рейтинг-класс = А

2.1. Для окончательного вывода относительно класса заемщика используются следующие правила вывода
Если рейтинг-класс – **А**,
репутация руководства контрагента – **безупречная**,
кредитная история - **безупречная**,
несвоевременное погашение основной части долга- **почти исключено**
несвоевременное погашение % - **почти исключено**, то класс А

2.2. Если рейтинг-класс – **Б**,
кредитная история - **безупречная**,
несвоевременное погашение основной части долга - **низкая**
несвоевременность погашения % - **почти исключено**
финансово-экономическое состояние контрагента – **негативне тенденции в деятельности**
недочёты в деятельности контрагента - **потенциальные**. то класс Б

2.3. Если рейтинг-класс – В,
несвоевременное погашение основной части долга и в сроки платежеспособность контрагента - средняя,
финансово-экономическое состояние контрагента–позитивные тенденции, то класс В

2.4 Если рейтинг-класс – Г
финансово-экономическое состояние контрагента - негативные тенденции в деятельности, то класс Г.

2.5 Если рейтинг-класс – Г
Риск значительных убытков - высокий;
Вероятность полного погашения кредитной задолженности и процентов/ комиссий – низкая, то класс Г.

2.6 Если рейтинг-класс – Г
Вероятность несвоевременного погашения кредитной задолженности и процентов/ комиссий – высокая, то класс Г.

2.7 Если рейтинг-класс – Г
Состояние необходимой документации, которая подтверждает наличие и
ликвидность залога – плохо, то класс Г.

2.8. Если рейтинг-класс – Д
Признанный банкротом контрагент Банка – да, то класс Д

2.5. Если рейтинг-класс – Д
вероятность несвоевременного погашения контрагентом – очень высокая, то класс Д .

3.1. Если класс – В,
Обеспечение кредитной операции - ликвидное, то класс В.

3.2 Если рейтинг-класс – Г
Обеспечение по кредитным операциям – вероятность проблем, то класс Г.

3.3. Если класс – Д
Обеспечение кредитной операции - неликвидное, то класс Д.

3.4 Если класс – Д

Обеспечение кредитной операции - отсутствует, то класс Д.

3.5. Если класс – Г
Обеспечение кредитной операции - вероятность проблем, то класс Д.

3.6. Если класс – В,
Обеспечение кредитной операции - вероятность проблем, то класс Г.

3.7. Если класс – Г
Обеспечение кредитной операции – не ликвидное, то класс Д.

3.8. Если класс – Г
Обеспечение кредитной операции – отсутствует, то класс Д.

3.9. Если класс – В,
Обеспечение кредитной операции - первоклассное, класс Б.

3.10. Если класс – Г
Обеспечение кредитной операции - первоклассное, то класс – В.

3.11. Если класс – Д
Обеспечение кредитной операции - первоклассное то класс = Г.

Применение методики определения кредитоспособности на основе нечеткой логики

- Рассмотрим применение вышеизложенной методики оценки кредитоспособности заемщика с использованием аппарата нечетко логики на следующих примерах
- **Оценка кредитоспособности ЗАО «АвтоЧасть»**
- Исходные данные:
- Данное предприятие занимается оптовой поставкой автозапчастей ведущих производителей, таких как Universal Industries, Fenox, Jestic , Прамо и др.

  1. В данном Банке обслуживается не полностью, более двух месяцев.
  2. Опыт работы в основной сфере деятельности - 10 лет
- Данные составлены на конец 10.2005

  3. обеспечение - не вызывает сомнений
  4. коэффициент независимости (КН) - 0.36
  5. Маневренность собственных средств (КМ) 0.75
  6. Сумма среднемесячной задолженности (тыс. грн.) 342
  7. Доходы (тыс. грн.) 514
  8. Поступление денежных средств на счета (количество раз) (Спдс) 4
  9. Среднемесячные поступления денежных средств (тыс. грн.) 100
  10. Среднемесячные поступления чистых денежных средств (тыс. грн.) 80
  11. Рентабельность продаж (РП) 0,14
  12. Рентабельность активов (РА) 0,067
  14. Коэффициент мгновенной ликвидности (КЛ1) 0,14
  15. Коэффициент текущей ликвидности (КЛ2) 1,6
  16. Оборачиваемость запасов 40
  17. Оборачиваемость краткосрочной дебиторской задолженности (ОДЗ) 50
  18. Оборачиваемость краткосрочной кредиторской задолженности (Окз) 50
  19. Текущая дебиторская задолженность (тыс. грн.) 300
  20. Краткосрочная кредиторская задолженность (тыс. грн.) 401
  21. Собственные активы (тыс. грн.) (СА) 1144
  22. Оборотные активы (тыс. грн.) (ОА) 655

23. Темп роста валовой прибыли (%)110
24. Темп роста объема реализации (%) 112
25. Темп роста суммы активов 114 (%)
26. Значение рейтинга за предыдущий период 6
27. Срок кредитной операции (мес.) 36
28. Просрочек по оплате нет
29. Вероятность несвоевременного погашения основной суммы 0.05
30. Вероятность несвоевременного погашения % 0.05

- Основные финансово-экономические показатели предприятия, а также кредитный рейтинг и обеспеченность кредитной операции и правило вывода итогового рейтинга предприятия приводятся в табл.1.

Таблица 1. Результаты работы программы для оценки кредитоспособности ЗАО «АвтоЧасть»

Параметр	Значение (рейтинг)	Параметр	Значение (рейтинг)	Итоговый параметр	Значение итогового параметра
СА	1144 (10)	Одз	50 (9)	Итоговая рейтинговая оценка	5,26
КН	0,36 (9)	Чд	45,801		
КМ	0,75 (8)	Окз	50 (9)	Окончательный класс платежеспособности	В
КЛ	0,7 (3)	Кдк	0,748		
КЛ1	0,14 (8)	Зп	0 (0)	К.п. с учетом кредитного обеспечения	В
КЛ2	1,6 (10)	Сп.н/р	0,19 (2)		
Рп	0,14 (9)	К	0,23 (0)	Лимит безопасного кредитования на 24 мес.	352,92 тыс. грн
Ра	0,067 (8)	Спдс	4 (0)		
Класс В, поскольку рейтинг-клас = В и кредитная история – безупречная и вероятность несвоевременного погашения кредита и вероятность несвоевременного погашения процента по кредиту – низкие.					
Итоговый рейтинг-нечеткий класс = В					

Заключение

В докладе рассмотрены проблемы оценки кредитных рисков заемщиков банков. Изложен классический метод оценки кредитных рисков на основе определения кредитного рейтинга заемщиков, а также оценки кредитного обеспечения. Далее изложен метод определения кредитоспособности заемщиков на основе использования систем с нечеткой логикой (алгоритм нечеткого вывода Мамдани). Сформулированы правила нечеткого вывода для оценки кредитного рейтинга и общего класса заемщиков. Приводятся примеры применения изложенного нечеткого метода.

Литература

[Зайченко, 2008] Ю.П. Зайченко. Нечеткие модели в интеллектуальных системах-К.: Издательский дом «Слово», 2008.-344 с.

[Зайченко, 2008] Зайченко Ю.П. Анализ кредитных рисков с использованием систем с нечеткой логикой и нечетких нейронных сетей.- Матеріали X1 Міжнародної науково-технічної конференції .-«Системний аналіз та інформаційні технології_ САІТ-2009». 26-30 травня 2009 р. Київ, Україна. -с.305

Информация об авторе

Зайченко Юрий Петрович, *профессор, д.т.н., кафедра «Институт прикладного системного анализа». Киев, НТУУ «КПИ», ул. Политехническая 14. тел: +8(044)241-86-93, e-mail: zaych@i.com.ua*

Medical and Diagnostic Systems

SYNTHESIS OF STRUCTURED MODELS OF COMPUTER SYSTEMS IN MEDICAL DIAGNOSIS

Anatoliy Povoroznyuk, Nataliya Bilous, Olga Kozina

Abstract: *Method for structural identification of objects of diagnosis in design of computer systems to support decision-making in medicine at all stages of information transfer is developed. This takes into account not only the structural and functional bases, but also the uncertainty of the model parameters.*

Keywords: *medical diagnostics, computer system, structural identification, the model, decision rule.*

ACM Classification Keywords*: I.5. Pattern recognition and J.3. Llife and medical sciences*

Introduction

The introduction of new information technologies in medicine gives rise to transition from traditional medical information retrieval systems to intelligent computer systems to support decision-making in medicine (ICSSDM) with advanced mathematical apparatus and elements of expert systems that are capable of monitoring of diagnostic criteria, build an advanced computer diagnoses, implement and management the treatment process.

Developers of ICSSDM denote following actual optimization problems that require urgent solutions:

- distinguishing of diagnostically significant structural elements in the background of artifacts and interferences (structural identification) in the analysis of biological signals, and medical imaging;
- formalization, the evaluation of informativeness of diagnostic criteria and formation of the minimum set of informative features;
- synthesis of decision rules according to which process of diagnosis (classification) of patients inside a given set of diagnosed conditions has carried out.

Application of traditional mathematical methods (deterministic methods, probabilistic, logical-and-linguistic, based on pattern recognition theory) for solving optimization problems in medicine, which was noted before, does not entirely correct [Ахутин 2002, Поворознюк 2006] because the complexity and multiple-level interactions between diagnostic objects (subsystems of the organism), the heterogeneity of diagnostic criteria (numeric, ranking or dichotomous data) and the heterogeneity of approaches of their obtaining, for example, by questioning, examination, clinical studies, the analysis of biological signals and medical imaging are not accounted for these methods.

In traditional methods diagnostic criteria are discussed in the form of a linear vector, and algorithms for automatic distinguishing of structural elements of biological signals are designed for each type of signal, and have heuristic nature. In addition, such methods are critical to volume of training set and have limitation on the dimension of objects of diagnostics in their implementation. In this paper, methods for constructing ICSSDM which based on unified formal approach of structural identification of objects of diagnostics at all stages of information transformation are proposed.

Principles of structural identification of objects of diagnostic

Proposed method of structural identification is based on conception of evolutionary identification of structured models which are effectively used for solving the problems of forecasting, recognition, management, etc. [Букатова 1990]. As known, the sense of structured C-models is the transformation $F: X \to Y$ or, in other words, reflection of input vector X to output Y via a set of operators f from the set F ($f \in F$) in accordance with structure S, which determines the sequence of operators f. Thus, the C-model is given by the graph, nodes of which are the functional elements of a given class of models, and information connection of nodes is defined by structure S, i.e. $C = \{f, S\}$. Synthesis of C-models is a recurrent and a stochastic process for consistent improvement of C-models, evaluation of C-model due to some criteria Q, and the selection of locally effective C-model. Classes of C-models, their functional and structural bases, as well as the modes of change are described in [Букатова 1990].

Taking into account the above optimization problem in the construction ICSSDM, heterogeneity of diagnostic criteria, an iterative process of diagnosis and the need to using of expert assessments in the formation of diagnostic output, the following stages of information transfer in ICSSDM according to structural identification have been formalized:

- structural identification of physiological signals and images;
- formalization of the description of heterogeneous diagnostic criteria;
- synthesis of hierarchical structure of diagnoses;
- synthesis of hierarchical diagnostic criteria;
- synthesis of diagnostic rules;
- synthesis of individual diagnostic prognostic models;
- recommendations on choosing the optimal treatment.

At each of considered phases local optimization problem is decided for which its own set of input data, own criterion of optimality and optimization algorithm must be used. That is why expanded notion of C-model (C'-model) and the procedure for reconfiguring the model is proposed to use in this work.

Extended C'-model is defined as follows

$$\mathrm{C}' = \{S, F, \mathrm{E}, \varepsilon_\sigma, \varepsilon_\Delta\}, \quad (1)$$

where – $S = \{P, V\}$ – structure model specified through a set of nodes P and arcs V;

F – functional basis of the model where the functions can be attributed as nodes f_p so and arcs f_v;

E – expert estimations which if necessary can be added to components F;

ε_σ – the uncertainty of parameters of F-basis that is determined by statistical properties of training set;

ε_Δ – the uncertainty of parameters of F-basis depended on accuracy of their determination (step of quantization).

Value ε_σ is used not only as an estimate of the model parameters (traditional approach), but is a parameter of optimization procedures in the synthesis model. Since some algorithms of optimization require quantization of the model parameters, then the task of selection the type of quantization (uniform or nonuniform) and selection of corresponding steps of quantization arises, thus ε_Δ can be considered like a parameter optimization procedures also.

Consider the application of the method of structural identification in the implementation mentioned earlier stages of information transfer in ICSSDM.

Identification of structural elements of biological quasi-periodic signals

Structural identification (allocation of the structural elements of the signal parameters which are diagnostic features) of quasi-biological signals (ECG, rheograms and others) is the most responsible and difficult phase of their processing, as well as errors in identification of structural elements(omission of structural elements or false identification) leads to gross errors in calculating of diagnostic criteria and errors in computer diagnosis (in cases when special means of exposure of these errors don't used, in particular, when the processing of suspected period is refused or confirmation of human-operator about correct structural identification is missed). The difficulty of identification is fact that the amplitude and temporal characteristics of structural elements holds information about the diagnosed subsystems of organism and have wide range of variability their values. In addition, these signals may include artifacts and can be registrated against the background noise. That is why formalized procedure for identifying the structural elements of biological signals on the basis of individual space of parameters for each type of structural element using the methodology and computational procedures of Hough transforms proposed in this work [Поворознюк, 2003].

A classic application of Hough transforms is the approximation of contours points of two-dimensional monochrome images by analysis curves of first or second order. Thus, each i-th point of the image is transferred to the space of parameters (parameters are the coefficients of approximating curve), and it forms a subspace of admissible values of parameters M_i. Intersection area Ω of subspaces M_i for all contour points determines the true value of parameters [Toronto, 2007]. Hough transforms is a basis for implementing various types of recognition and classification algorithms whose main idea is to move into more informative space of parameters for this task. In the structural identification of physiological signals conversion from the original space (set of discrete points of the signal $X(t_i)$) in the minimum-required space of parameters $Y(P)$ is executed. Space of parameters $Y(P)$ is formed at description of structural element pattern by a limited set of basic reference functions (approximation of pattern). So comparison of analyzed signal with pattern and the decision on the presence or absence of a structural element of a given type is performed in the parameters space.

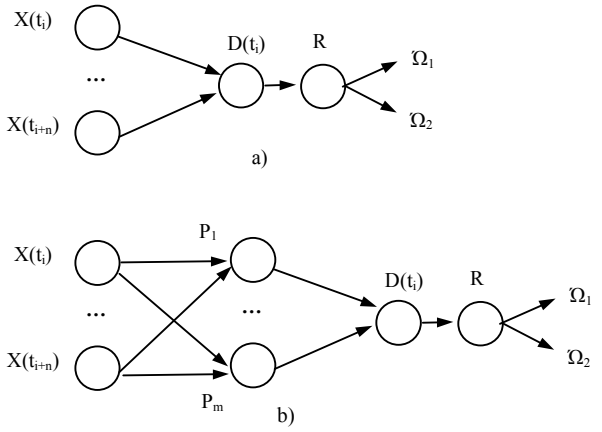


Figure 1. C'-model representation of structural element recognition: a) in contour analysis; b) in proposed method

In considered cases scanning of whole signal and signal part analysis with an aperture n, which is equal to the length of the structural element, are performed. In contour analysis through complex algorithms for each type of structural elements in each type of biological signal function $D(t_i)$ which reflects the distinctive features of given structural elements, is constructed. Then function $D(t_i)$ is compared with boundary R and decision about whether is the part of signal structural element (class Ω_1) or not (class Ω_2) is accepted. In the proposed method, at first, the operator mark out a pattern's structural element of this type in the signal in learning mode.

After that the formalized procedure for approximation of the pattern by graphic primitives is started, and parameters of individual space with dimension *m* are formed by applying Hough transforms. In such case $D(t_i)$ is a function of differentiation of distance, in other words, each its point is value of remoteness from the pattern in parameters space $Y(P)$. Since parameters space $Y(P)$ must be discrete, the value ε_Δ is optimized in such way that center of the cluster Ω_1 has been the level of quantization and its effective diameter has been the step of quantization $_\Delta P_i$.

Adaptation of identification of two-dimensional diagnostic structural elements is result of description of two-dimensional patterns by different types of spatial adaptive developments where transition from original rectangular system of coordinates i, j ($X_{i,j}$ – values of brightness in rectangular system of coordinates) into the space I (X_I –- values of brightness along the line scan) is performed. In contrast to traditional arrangement of structural identification of two-dimensional images (filtering, contouring, the search for structural elements), proposed approach give abilities to define pattern directly to the grayscale image.

Synthesis of hierarchical structures of diagnostic criteria S_X and structures of diagnoses S_D based on their self- informativeness

Synthesis of hierarchical structures of diagnostic criteria S_X and diagnoses S_D based on their self-informativeness is considered like hierarchical clustering based on the analysis of correlations through transformations $F_1 : X \to S_X, F_2 : D \to S_D$. To implement the conversion F_1 the procedure of hierarchical clustering of diagnostic criteria based on representation of clustering in view of a streaming model is proposed. Thus, initial characteristics become nodes of fully connected graph. Arcs of such graph can be equivalents of pair correlation coefficients. Then, task of hierarchical clustering of nodes is reduced to a consistent procedure for finding of the minimum cut of the graph in the streaming task with limitations. To solve streaming task adaptation algorithm of "defect" is proposed.

In streaming tasks with limitations each arc is characterized by the following parameters:

- f_{ij} – flow on arc (i, j);
- L_{ij} – lower bandwidth arc (i, j);
- U_{ij} – upper flow capacity of arc (i, j);
- C_{ij} – the cost of passing the unit flow from node i to node j.

For considered problem of clustering of diagnostic criteria of the algorithm on the basis of algorithm of "defect", value L_{ij} is equal 0, and U_{ij} is equal to criterion of self-informativeness (pair correlation coefficient). If apriori information is absent, then for all arcs C_{ij} can be equal 1 or their values can be received from expert estimations in contrary. Values of streams f_{ij} are determined when the algorithm of "defect" is completed and initially may be equal to 0. To reduce the dimension of the problem it is possible to consider only arcs for which self-informativeness criterion is significant in the sense of Student's criterion.

For implementation of algorithm of "defect" it is necessary to define initial and finish node (source S and sink). Results of its work are the calculation of streams circulation which minimized the total cost of streams on all arcs according to limitations of arcs capacity (L_{ij} and U_{ij}). The correct choice of source and sink is heuristic task, therefore, at first estimated clustering must be carry out by method of galaxy correlation, and as source and sink are chosen the most connected nodes in each cluster. After determining of streams on all arcs the minimum cut of graph R is determined, for which the true relationship:

$$\begin{aligned} f_{ij} &= U_{ij} \quad \forall (i,j) \in (N_c, \overline{N_c}), \\ f_{ij} &= 0 \quad \forall (i,j) \in (\overline{N_c}, N_c). \end{aligned} \quad , \tag{2}$$

The minimum cut divides all nodes of the graph into two disjoint sets ($\overline{N}_C$ include an initial node of S and $\overline{N}_C$ include and finish node T), thus, streams saturate all direct arcs of cut and zero arcs of cut in inverse direction (2). In this case, the sum of streams of all arcs of the cut equals the maximum circulation of stream.

Result of hierarchical clustering is hierarchical structure, on lower level of hierarchy into which separate diagnostic criteria, if they are informative and independent, or clusters of correlated features can be present.

Synthesis of hierarchical system of diagnoses – transform F_2 – is performed on a similar way of submission problem like streaming model with limitations and finding the minimum cut with help of algorithm of "defect". However, there are the following differences:

- top capacity of arc i, j – value U_{ij} – is intercluster distance of diagnoses D_i and D_j in space of criteria;
- dichotomic clustering is performed without compression, so the above criteria for linear clustering and optimality in this procedure is not used.

Reconfiguration of structures of diagnostic criteria S_X

Structures of diagnostic criteria S_X is obtained by the criterion of self-informativeness of criteria and can not be used for creating of additional diagnostic specification directly because the resulting structure S_Z must contain minimum necessary amount of informative diagnostic criteria for given diagnostic problem, i.e. given set of diagnoses. The structure must be submitted with structure S_D and procedure of clusters replacement at all levels of hierarchies S_X should be developed. Therefore, reconfigure of structure S_X is proposed in order to create the structure S_Z. Criteria of diagnostic value of parameter x_i with regard to given set of diagnoses $\{D\}$ can be used for this purpose [Поворознюк 2007, 1]. Diagnostic value $I_p(x_i)$ is amount of information introduced into the system after patient observation to parameter x_i. For all that transformations F_3: $S_X \to S_Y$ and F_4: $S_Y \to S_Z$ are completed.

If diagnoses set are formed whole group of incompatible events – only one diagnosis D_i correspond to one patient in training set, no patients with multiple diagnoses), the uncertainty of diagnoses set is estimated entropy:

$$H(D) = -\sum_{i=1}^{n} P(D_i) \cdot \log_2 P(D_i), \tag{3}$$

where $P(D_i)$ – apriori probability of diagnosis D_i.

Expression (3) show estimation of maximum possible amount of information that can be introduce in set of diagnostic criteria, i.e. for any system of diagnoses *D* and any set of diagnostic criteria $X = \{x_1, x_2, \ldots x_p\}$ inequality is correct:

$$I_D(X) \le H(D), \tag{4}$$

which becomes an equality only for set of deterministic parameters.

Transformation F_3 determine informativeness of original space of criteria X according to the system diagnoses set $\{D\}$. For this, clusters of lower-level hierarchy of structure S_x – subsets of correlated features – are

replaced to most informative and other clusters of remaining level of S_x are replaced to integral features. In addition, in synthesis of S_y set of heterogeneous elements S_x is given to a single scale – every original parameter x_i is represented like disjoint set of diagnostically significant intervals.

Changing range of numerical value $\Delta = \lfloor x_{j\min}, x_{j\max} \rfloor$ for sequence of intervals is in fact the replacement of theoretical law of distribution for histogram. From this point of view to obtain an acceptable accuracy of approximation of distribution law, the number of intervals must be large enough. But it is necessary to increase training set, because at constant sample size and increasing the number of intervals relatively small number of points gets in each interval, which reduces the reliability of statistical estimates. Thus, with limited training set, the task of dividing Δ on intervals is optimization task in which integral error ε_k is minimized [Поворознюк 2007, 2]:

$$\varepsilon_k = \frac{\Delta_k}{\sqrt{6N_k}} \sqrt{\left(\frac{G_{k+1} - G_k}{\Delta_{k+1} + \Delta_k}\right)^2 + \left(\frac{G_k - G_{k-1}}{\Delta_k + \Delta_{k-1}}\right)^2}, \quad (5)$$

where m – number of nonuniform intervals Δ_k;

N_k – number of points from training set which get in interval Δ_k;

G_k – ordinate of the histogram into interval Δ_k.

Calculation of ε_k due to (8) requires apriory dividing Δ by Δ_k and calculation of values inside not only in the current k-th interval – Δ_k and G_k – but also in neighboring: $\Delta_{k-1}, G_{k-1}, \Delta_{k+1}, G_{k+1}$. Minimum of ε_k can be reached by iterative procedure for formation of Δ_k.

The goal of transformation $F_4 : S_Y \to S_Z$ is agreement of topologies S_Z and S_D in order to complete the method of diagnostic specification and to ensure the optimal plan for individual patient diagnostic observations. Thus, each pair of S_D with total parent has agreement with one element of S_Z. Elements of structure S_Z are formed from elements of S_y on the basis of their informative completeness and diagnostic value so that they could perform a differential diagnosis for each level of diagnoses hierarchy.

Set of diagnostic criteria X relative on diagnoses set D is described by coefficient of informative completeness $k_{IC}(X, D)$:

$$k_{IC}(X, D) = \frac{I_D(X)}{H(D)}. \quad (6)$$

Similarly, coefficient of informative completeness for each parameter x_j is determined:

$$k_{IC}(x_j, D) = \frac{I_D(x_j)}{H(D)}. \quad (7)$$

Since process of diagnostic criteria measurement requires to using medical equipment and some resources (time, material, financial, etc.), given the complexity of combined measurement, each parameter is characterized by the coefficient of diagnostic value:

$$k_{DV}(x_j, D) = \frac{k_{IC}(x_j, D)}{r_t(x_j)}, \quad (8)$$

where $r_t(x_j)$ – the total ratio of measuring complexity of parameter x_j.

The formulation of S_Z begins from top-level element, i.e. from element $Z^1_{1,2}$, which should provide a differential diagnosis between states $D^1_{1,1}$, and $D^1_{2,1}$.

Initially entropy diagnoses $D^1_{1,1}$, and $D^1_{2,1}$ is calculated by (3) and required informativeness of $Z^1_{1,2}$ is calculated by (6) for a given ratio of measuring complexity. It should be noted that informativeness of any element of S_Z does not exceed the 1bit and reaches its maximum for equiprobable diagnoses, in accordance with (4).

Then coefficients of diagnostic values of all elements of S_y for pair of diagnoses $D^1_{1,1}$, and $D^1_{2,1}$ is calculate by (8) and elements of S_y is organized in accordance with

$$k_{DV}(y_i) \geq k_{DV}(y_j) \geq \ldots \geq k_{DV}(y_q), \quad (9)$$

For each element of S_y coefficient informative completeness for pair of diagnoses $D^1_{1,1}$, and $D^1_{2,1}$ is defined, and then sequential process of including of elements y_i from ordered sequence (9) into $Z^1_{1,2}$ is repeated until not reached the necessary level of informative completeness of element $Z^1_{1,2}$.

In order to formulate elements of next level of hierarchy $Z^2_{1,2}$ and $Z^2_{2,3}$, elements informativeness of which has been exhausted during formulation of $Z^1_{1,2}$ are excluded from structure S_y.

Thus, every element of S_Z is ordered subset of elements of S_y and uniquely specifies order of their inclusion in the diagnostic procedure of diagnostic specification at each phase of implementation.

Conclusion

Thus, the methods of synthesis of computer systems to support decision-making in medicine based on the structural identification of objects of diagnosis is developed. Conception of extended structural model takes into account not only structural and functional bases, but also the uncertainty of parameters and also allows using expert estimations.

Construction of hierarchical structures from 9 parameters of clinical blood analysis and 10 different diagnoses are completed for 434 patients. The average probability of diagnostics has increased from 88.48% when using a standard cluster analysis to 92.8% when using the developed method. From the total number of negative responses differentiation for 25 negative responses to the preliminary diagnoses of different levels is received while using only 3.61 diagnostic parameters an average for each patient, that in 2,5 times less than the original 9 parameters which was used for standard procedures of diagnostic.

The developed methods allow removing restrictions on dimension of diagnostic criteria space, to improve the reliability of computer diagnosis and to adapt to specific objects of medical diagnosis.

Bibliography

[Ахутин 2002] В.М. Ахутин , Шаповалов В.В., Иоффе М.О. Оценка качества формализованных медицинских документов. Медицинская техника. № 2 (2002), pp. 27-31 .

[Поворознюк 2006,1] А.И. Поворознюк Формализация этапов проектирования интеллектуальных компьютерных систем медицинской диагностики. Электронное моделирование. Т. 28, №1 (2006), pp.85-97.

[Букатова 1990] И.Л. Букатова, Ю.И. Михасев, А.М. Шаров Эвоинформатика: теория и практика эволюционного моделирования.-М.:Наука, (1990), pp.212.

[Поворознюк 2003] А.И. Поворознюк Применение преобразования Хока для структурной идентификации физиологических сигналов. Моделювання та інформаційні технології. Збірник наукових праць інституту проблем моделювання в енергетиці. Вип. 22, (2003), pp.143-149.

[Toronto, 2007] N. Toronto, B.S. Morse, D. Ventura, K. Seppi. The Hough Transform's Implicit Bayesian Foundation. IEEE International Conference on Image Processing, ICIP 2007. Vol. 4, (2007), pp.377-380.

[Поворознюк 2007, 1] А.И. Поворознюк Реконфигурация структуры диагностических признаков при синтезе компьютерного диагноза. Системи обробки інформації. Вип. 2(60), (2007), pp. 135 – 139.

[Поворознюк 2007, 2] А.И. Поворознюк Формирование диагностических интервалов численных признаков при дифференциальной диагностике. Вісник Хмельницького національного університету. №3, Т.1, (2007), pp. 106-109.

Authors' Information

Anatoliy Povoroznyuk *– Ph.D., lecturer of Computers and Programming Department of National Technical University 'KPI', Kharkov, Ukraine; e-mail: koa@kture.kharkov.ua*

Nataliya Bilous *– Ph.D., head of ITLCVS laboratory, professor of Kharkov National University of Radio Electronics, Kharkov, Ukraine; e-mail: belous@kture.kharkov.ua*

Olga Kozina *– Ph.D., lecturer of Computers and Programming Department of National Technical University 'KPI', Kharkov, Ukraine; e-mail: koa@kture.kharkov.ua*

ARTIFICIAL INTELLIGENCE APPROACH TO DIABETES DIAGNOSTICS

**Galina Setlak, Mariusz Dąbrowski, Wioletta Szajnar,
Monika Piróg-Mazur, Tomasz Kożak**

***Abstract**: In this paper a concept of designing and building intelligent system in diabetes diagnostic is introduced. The way of the expert, classifying the input data system (symptoms) for four types of diabetes (classes) has been shown in that article.*

***ACM Classification Keywords:** I. Computing Methodologies, I.2 Artificial Intelligence, 1.2.5.Expert system tools and techniqes*

***Keywords**: artificial intelligence, expert systems, hybrid systems, neural, networks, diabetes mellitus.*

Introduction

The concept of the artificial intelligence was started up together with the beginnings and developing of the computer era. The possibility of adopting one of the basic features distinguishing a man, that is an intellect were concerned. Computers outdid and outdo people in mathematical calculations speed, but they lack the basic element, thanks to which they cannot match the human beings. It is the consciousness.

However, the artificial intelligence, defined as the intelligence hallmarks, is the science sphere that is developing quite fast and in the course of time one can count on new achievements.

The artificial intelligence has been defined in many different ways. According to Minsky, the artificial intelligence is science about the machines realizing the tasks that need the intelligence when they are done by a human being [Kowalczuk, Wiszniewski, 2007], [Kwaśnicka, 2005].

Figenbaum defined intelligence as the computer science field concerning methods and techniques of symbolic deduction by a computer and the symbolic representation of knowledge applied during that deduction [Kowalczuk, Wiszniewski, 2007].

Whereas Turing proposed the following definition: if the unbiased, external observer is not able to differentiate the machine answer from the human answer, we can talk about the intelligence of that machine [Kowalczuk, Wiszniewski, 2007].

Expert systems

An expert system is the information system which as its name suggests performs tasks as the expert in that sphere of the science or knowledge. This system, on the basis of particular knowledge, rules basis, draws conclusions, takes decisions, and its activity is similar to the human activity in that sphere.

Expert systems can be classified from different points of view. It can be the advisory systems, the ones that suggest the direction; systems taking decisions without the help and human interference; criticizing systems the ones which on the basis of a particular problem and the predicted solution by a man, analyze and comment the particular reasoning and action way.

Forming a system based on the knowledge background needs the expert knowledge who often finds a solution on the basis of the information about the problem and on the basis of his own experience. The expert system, having a written expert knowledge from a chosen field, can use it many times in an economic and effective way

because it does not need the presence of the expert. At the same time, it allows the expert not to repeat the analogous reports and take up more creative tasks. The special advantage of such systems is the possibility of solving the particular tasks without the direct expert's participation, and also the possibility of knowledge aggregation in one system of the numerous experts team [Mulawka, 1996].

Neural networks

The concept of neural networks has its own background in the biological nervous system. It is a very complicated structure consisting of neurons (nerve cells) and connections between them. Not delving into strictly biological background, let us look how the neural networks are formed (because our model should be treated like that) [Witkowska, 2002].

Artificial neural networks, as it is easily predicted, consist of artificial neurons. From the technical point of view, is the element of which the features match the chosen features of the biological neuron. The artificial neuron is not a faithful copy of the biological neuron, but the element that should fulfil particular functions in the artificial neural network. Such an artificial neuron is in a sense a transducer with the signal at the entrance, and it is then multiplied by the particular for each transducer, weighting kit and summed up. We receive the new signal at the way out, which defines the neuron activity.

Neural networks can be divided according to the build into:

- feed forward networks:
    - one-layer,
    - multi-layer,
- recurrent networks,
- cellular networks.

The most important feature of the neural networks is their ability to learn that is the ability to independent adjusts the weighting factors. Learning is done in particular cycles, so each task to solve for the neural networks is at the same time a new stimulus, causing the increase of “knowledge” of a particular network. Thanks to such a phenomena, the neural networks represent the sphere of the artificial intelligence [Rutkowska, Piliński, Rutkowski, 1997].

Hybrid systems

A hybrid system is a new category of systems based on the artificial intelligence. They rely on connection of the best features of such systems as: expert systems, learning systems, neural networks, and genetic algorithms. Thanks to that, the particular system is able to solve the most difficult problem, the single system which is the part of the hybrid system could not cope with. It is obvious that implementing such system is connected with the additional difficulties resulting from the necessity of connection of these elements [Białko, 2000].

Hybrid expert systems, as others classical expert systems, are built upon fundamental components:

- a knowledge base,
- an inference engine (interpreting knowledge stored in the knowledge base and making deductions),
- knowledge engineering system,
- automatic knowledge acquisition,
- explanation subsystem,

- user interface – one for accessing the knowledge base through the knowledge acquisition module, and another one for system users accessing the system in the consultation mode or in the explanation (tutor) mode and
- additional component part – the neural network.

In this context the cooperation between systems usually follows by data interchange. Each of the subsystems realizes specifying purposes. It works by autonomous way and transmits results of its activity to the other system. Especially spectacular and also practically useful are results of expert system and neural network integration. We can describe following examples of their cooperation:

- the neural network realizes numeric data processing for the expert system ;
- expert system controls the learning process of neural network;
- the neural network is made for building knowledge base of the expert system;
- the expert system transforms the output neural network data in order to show there suitable for people interpretation.

Expert system outlined in this paper uses PC-Shell 4.2. – domain independent expert system shell, having strong hybrid properties. The PC-Shell has been implemented in Artificial Intelligence Laboratory (AITECH, Katowice). The PC-Shell 4.2 system integrates the expert systems shell using blackboard architecture elements and the simulator of the neural network. It assures the knowledge representation as declarative expressed rules, facts and distributing knowledge in the neural network. The expert knowledge can contain in some knowledge sources. This system enables procedural knowledge representation too.

The knowledge representation language SPHINX is a mean for building intelligent applications. It is the way of integration of particular artificial intelligence systems. We can find a quota on this subject in paper [Bubnicki, 1990], here we will discuss only selected aspects of knowledge representation. The knowledge base structure of the PC-Shell 4.2 system is distributed following:

The block sources and control will be the most important for this paper. They are means of integration of the expert system and the neural network.

The source block: The PC-Shell is a hybrid system with elements of blackboard architecture. This determines that we can use a lot of heterogenic knowledge sources for problem solving. The declaration of ource in sources block consists of symbolic name and account of properties. One of properties is source type, we can declare sources for following types:

♦ kb – expert knowledge base,

♦ neural_net – neural networks,

♦ metaphors and what_is_file – the data base containing explanations.

The source has properties file describing a file name for source, which is, creates in the Neuronix subsystem.

The control block: The PC-Shell 4.2 system makes possible an integration of the declarative knowledge representation

```
knowledge base name
        sources
                sources description
        end;
        facets
                facets description
        end;
        rules
                rules description
        end;
        facts
                facts description
        end;
        control
                program
        end;
end;
```

Figure 1.The knowledge base structure of the PC-Shell 4.2

with procedural knowledge representation language, which enables programming of the system activity. The program in the PC-Shell system consists of instruction set included in the control block. The subset of language instruction enables the integration of neural network and the expert system. The instructions are following:

- initNetwork(X) – induces generating of neural network. X (the parameter) – defines the name of the knowledge source.
- DelNetwork(X) – induces mowing of neural network pointed by parameter X, where X is name of source declared in source block.
- RunNetwork(X,Y) – induces running the neural network, earlier inducted by initNetwork instruction. The parameter X defines the input vector of data and Y defines the output vector. Input and output vectors make possible passing the input data for neural network and also taking of results.

The example the realization of hybrid application in the PC-Shell

The realization of application in the PC-Shell system is following:

- Creating by the NEURONIX subsystem one or some neural application.
- Elaborating knowledge base in form of knowledge sources.
- Integration of the elaborated knowledge sources on the level of knowledge representation language.

Expert system of diabetes diagnostic

Expert systems or generally the artificial intelligence, are useful where the expert's knowledge is used. One of the spheres is medicine. As it is commonly known, in such an important aspect of everyday life there is no place for half measures – a doctor can not be an expert in one field only at the certain level, be not educated enough, can not count that in case of the mistake, something can be turned back. Obviously, such assumptions should be binding in all spheres of life, but let us leave divagation on that aside. Therefore a doctor is the biggest authority concerning the particular problem, he is the expert. He has to use his knowledge, knowledge of the expert each time when he intervenes (and he does it every day-quite often). But it happens very often, that the problem diagnosed by him is quite trivial, and diagnosis identical, so he has to devote lots of time to mechanical deduction. But it does not have to be that way, if instead the expert system -properly construed, can be “employed” instead of that [Bizoń, 2008].

One of such systems can be the system diagnosing the types of diabetes. Let us have a look at the characterization of the expert system operation, diagnosing this disease. Its results can be the following types (classes) [WHO, 1999]:

- Type 1 diabetes (immunologically conditioned or idiopathic),
- Type 2 diabetes (with obesity or without obesity),
- Other defined types of diabetes,
- Diabetes mellitus in pregnancy.

To start diagnosis, one should gather the proper number of data. The source of the data can be the following: a patient himself, patient's record, primary physician, specialist, biochemical lab, specialist tests.

The system gathers data by the tests: subjective and objective, laboratory and additional tests.

a) Subjective tests – history taking

- the system takes data such: name, surname, age, occupation, place of work, life style (diet, addictions (smoking), alcohol), general condition, dietary habits, nutritional status (obesity, emaciation), history of the body mass, family interview (did the family members suffer similar diseases), beginning of the

disease (when, how many years ago) course of disease, ailments from other organs and systems, past diseases and operations, present and previous infections, diabetes education, course of the treatment (in case of previously treated due to the diabetes) medicines used [PTD, cukrzyca info].

b) Objective tests

- height and body mass measurement (BMI), counting the proper body mass and comparing with the real mass, personal development phase evaluation, (sexual - bodily phase, old age phase), arterial pressure evaluation in a lying position and standing position (with the measurement of the orthostatic reaction), ophthalmoscopes tests of fondues (with papillary dilation), thyroid test, heart test, taking pulse and testing all peripheral arteries accessible when fingering and auscultating, feet test, neurology test, teeth and gums test, skin test and mucosa tests [WHO, 1999], [PTD, cukrzyca info].

c) Laboratory tests:

- glycaemia (blood glucose level) test on an empty stomach and the daily glycaemia profile,
- notation of the glycated haemoglobin and fructosamine,
- notation of the lipid profile on an empty stomach: total cholesterol and cholesterol in lipoproteins of high density (HDL – high density lipoproteins), cholesterol in lipoproteins of low density (LDL – low density lipoproteins) and trigliceryde,
- urine test (apart from glucosuria) ketone bodies and protein presence (macro- and micro-albuminuria) and microscope test of the sediment,
- bacteriological test (urine cultures and antibiotic gram),
- euthyroidism test and morphological status of the thyroid test (concentration evaluation T3, T4 and TSH, scintigraphy of the thyroid),
- peripheral arteries tests (potency and rush of blood),
- electrocardiogram, echocardiography, ergo meter test,
- neurological tests, especially the electromyography test,
- ophthalmic review (general test of the organ of sight) [WHO, 1999].

d) Additional tests

- fundus test

Conclusion

After gaining all information about the patient, symptoms and disease and after basic tests, we are at the phase when our expert system draws conclusions on their background. Diagnosing is done on the rule of moving on the decision tree. As it was mentioned before, the system that has been defined recognizes 4 classes of the disease: type I diabetes, type II diabetes, secondary diabetes, diabetes mellitus in pregnancy. In order to present the way of activity there is no need to show the conclusions scheme for each class in details – let us have a look only on one of them, the rest, of course is similar.

Diagnosis Scheme is quite simple and in a way obvious. One direction means the natural course of the disease and the possibility of the pharmacological treatment, the other one show abnormalities or alarming course of the disease and its symptoms. The final result can be the statement that the patient was (is) diabetic of the particular type, and the recommended remedy is just the proper diet and the pharmacological treatment.

That way we are approaching the diagnosis. If we analyzed one of the possible ways in details, we would notice that the particular phases do not differ from the doctor's way of proceedings. Our expert system has a task to achieve the appropriate conclusion, behaving like an expert in that field that is a doctor.

It is important that there is a close cooperation between a system engineer and experts on those interest fields when constructing the appropriate expert system. Nothing can replace the knowledge and doctor's experience.

Bibliography

[Kowalczuk, Wiszniewski, 2007] Z. Kowalczuk, B. Wiszniewski. Inteligentne wydobywanie informacji w celach diagnostycznych, Pomorskie Wydawnictwo Naukowo Techniczne, Gdańsk 2007.

[Kwaśnicka, 2005] H. Kwaśnicka. Sztuczna inteligencja i systemy ekspertowe. Rozwój i perspektywy, Wydawnictwo Wyższej Szkoły Zarządzania i Finansów, Wrocław 2005.

[Bubnicki, 1990] Z. Bubnicki. Wstęp do systemów ekspertowych, PWN, Warszawa 1990.

[Rutkowska, Piliński, Rutkowski, 1997] D. Rutkowska, M. Piliński, L. Rutkowski. Sieci neuronowe, algorytmy genetyczne i systemy rozmyte, PWN, Warszawa 1997.

[Białko, 2000] M. Białko. Podstawowe właściwości sieci neuronowych i hybrydowych systemów ekspertowych, Wydawnictwo Uczelniane Politechniki Koszalińskiej, Koszalin 2000.

[WHO, 1999] World Health Organization (WHO): Definition, diagnosis and classification of diabetes mellitus and its complications. WHO. Geneva 1999.

[PTD, cukrzyca info] http://www.cukrzyca.info.pl – strona PTD poświęcona cukrzycy.

[Witkowska, 2002] D. Witkowska. Sztuczne sieci neuronowe i metody statystyczne, Wydawnictwo C. H. Beck, Warszawa 2002.

[Bizoń, 2008] D. Bizoń. Analiza klasyfikacji danych medycznych za pomocą wybranych metod sztucznej inteligencji, w: Technologie Informatyczne e Medycynie, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2008.

[Mulawka, 1996] J. J. Mulawka. Systemy ekspertowe, WNT, Warszawa 1996.

Authors' Information

Galina Setlak, D.Sc, Ph.D., Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science , W. Pola 2 Rzeszow 35-959, Poland, and The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland, e-mail: gsetlak@prz.edu.pl

Mariusz Dąbrowski *- Dr. of medical science, specialist diabetolog; e-mail: madab@esculap.pl*

Wioletta Szajnar - *M.A. The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland: e- mail:* wiola@pwszjar.edu.pl

Monika Piróg Mazur - *M.A. The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland: e-mail:* m_pirog@pwszjar.edu.pl

Tomasz Kożak - *M.A. The State Professional High School, Czarnieckiego 16, 37-500 Jarosław, Poland: e-mail:* tomasz@pwszjar.edu.pl

Computer Engineering, Mechanical Engineering

NAVIGATION SOFTWARE OF AUTOMATED GUIDED VEHICLE

Magdalena Dobrzańska, Paweł Dobrzański

Abstract: *In the article it has been presented the structure of the control system and measurement data processing of an automated guided vehicle. The basic navigation technique – odometry, which is applied in the automated guided vehicle, has been described, as well as connected with it errors of position tracing. Next the navigation software was shown which enables to design the trajectory of the vehicle movement as well as the registration and reading of the measurement data from the measurement sensors.*

Keywords *:Automated Guided Vehicle, navigation, odometry, odometry errors, Visual C++*

ACM Classification Keywords*: I.2.9 Robotics: Autonomous vehicles, Sensors, J.7 Computers in other systems*

Introduction

Automated Guided Vehicles (AGV) are commonly used in various areas of man's activities. Within two decades one could notice a rapid development of automated guided vehicles which have their applications in totally different areas, for example medicine, industry, transport, defense or agriculture. Some of the vehicles have additional equipment which enables them to perform very sophisticated activities. Presently available vehicles can perform many functions. Depending on the area there are vehicles which can be used in a wide variety of applications, for example in industry, transport (haulage, loading), medicine (automatic nurses, self-propelled wheelchairs), defense (patrol vehicles, vehicles for weapon defusal, vehicles equipped with armament).

The above classification does not reflect in full the applications of the AGV. The aim of their construction is to eliminate dangerous, hard and monotonous work performed by people.

Such a vehicle has been also developed at Rzeszow University of Technology. It is used for the various researches and for the didactics. Some research concerned the issues connected with navigation, especially with the precision of positioning and following the given route.

The aim of the article was to show the basic information connected with analytical navigation, to show already accepted for realization constructional solutions applied in the vehicle and to present the navigation software used to design trajectory, measurement data recording and processing.

Odometry

The base for positioning in the automated guided vehicles is odometry – the analytical navigation. Odometry rely on the determination of the vehicle current position on the base of way covered by the characteristic point K . The analytical navigation uses to determination of the vehicle movement direction angle θ difference of speed of driving wheels the v_L and v_P . The basic idea of this solution was shown in fig.1.

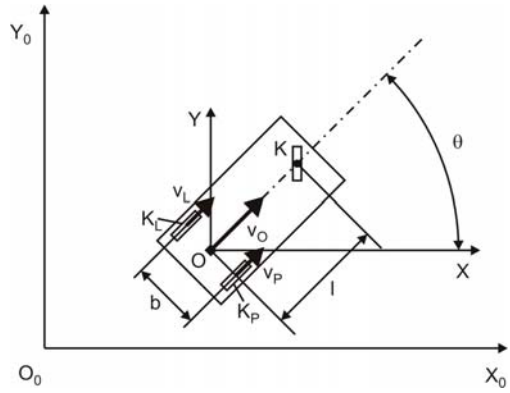


Figure 1. Coordinate system used in analytical navigation

Applied method rely on measurement of the way covered by the driven wheels K_L and K_R and the determination in each iteration of the vehicle movement direction angle θ is used in the vehicles in which two drive wheels independently driven are used to steer the vehicles. The proper differentiation in the rotational speed of these wheels forces the turn of the vehicle around its vertical axis of rotation that goes through the point O and the change in the direction angle θ [Dobrzańska, 2005].

If the position of point O the vehicle in which two drive wheels independently driven K_L and K_P are used to steer the vehicles in the basic reference system $X_0O_0Y_0$ (fig. 1) in *k*th iteration is given by the state vector $(x(k), y(k), \theta(k))$, then position of the vehicle in (*k*+1)th iteration is given by the equation:

$$\begin{bmatrix} x(k+1) \\ y(k+1) \\ \theta(k+1) \end{bmatrix} = \begin{bmatrix} x(k) \\ y(k) \\ \theta(k) \end{bmatrix} + \begin{bmatrix} \Delta t \cdot v_{\mathrm{O}}(k+1) \cdot \cos(\theta(k) + \Delta t \cdot \omega(k+1)) \\ \Delta t \cdot v_{\mathrm{O}}(k+1) \cdot \sin(\theta(k) + \Delta t \cdot \omega(k+1)) \\ \Delta t \cdot \omega(k+1) \end{bmatrix} \tag{1}$$

Velocity $v_{\mathrm{O}}(k+1)$ and $\omega(k+1)$ can be determined from the following relations:

$$v_{\mathrm{O}}(k+1) = (v_P(k+1) + v_L(k+1))/2 \tag{2}$$

$$\omega(k+1) = (v_P(k+1) - v_L(k+1))/b \tag{3}$$

where: $v_P(k+1)$ - speed of the right wheel

$v_L(k+1)$ - speed of the left wheel

b - wheelbase of the driven wheels.

Velocity $v_P(k+1)$ and $v_L(k+1)$ can be determined from the following relations:

$$v_P(k+1) = \omega_P(k+1) \cdot r \tag{4}$$

$$v_L(k+1) = \omega_L(k+1) \cdot r \tag{5}$$

where: *r* - radius of the steered wheels

In the above consideration it was assumed that the wheels are rigid and they roll without spin, the contact between the wheel and the floor is a point contact and the radiuses *r* of the driven wheels are the same.

Presented and accepted for realization method is very simple but it has got some drawbacks connected with errors. We can distinguish several sources of errors that have an impact on the accuracy of positioning. These sources were divided into two categories:

- Systematic errors caused by: unequal wheel diameters, misalignment of wheels, uncertainty about the effective wheelbase (due to non-point wheel contact with the floor), limited encoder resolution, limited encoder sampling rate.
- Non-systematic errors caused by: travel over uneven floor, travel over unexpected objects on the floor, wheel-slippage (due to: slippery floor, over-acceleration, fast turning (skidding), external forces (interaction with external bodies), internal forces (e.g., castor wheels), non-point wheel contact with the floor).

Additional odometry errors can be caused by the odometry equations themselves, since they approximate an arbitrary motion as a series of short rectilinear segments. The precision of this approximation depends on the program step.

Two dominant sources of errors in odometry are:

- unequal wheel diameters - most mobile robots use rubber tires to improve traction. It is very difficult to manufacture wheels with exactly the same diameter. Furthermore, rubber tires compress differently under asymmetric load distribution. Either one of these effects can cause substantial odometry errors. We denote this error as Ed and define it as

$$E_d = \frac{d_R}{d_L} \tag{6}$$

  where: d_R and d_L are the actual wheel diameters.

- uncertainty about the wheelbase - the wheelbase is defined as the distance between the contact points of the two wheels of the moving vehicle and the floor. Uncertainty in the effective wheelbase is caused by the fact that rubber tires contact the floor not in one point, but rather in a contact area. We denote this error as Eb and define it as

$$E_b = \frac{b_a}{b_n} \tag{7}$$

  where b_a is the actual wheelbase of the vehicle,
  b_n is the nominal wheelbase of the vehicle.
  E_d and E_b are dimensionless values.

Description of construction

Automated guided vehicle designed to transport of cargos executed at Rzeszów University of Technology was the object of investigations. Vehicle has two drive wheels independently driven which are used to start and steer the vehicle. The object of investigations has onboard PC, cards of data acquisition, and control-measuring equipment. The vehicle is built on the base of the three-wheeled construction with two drive wheels and one independent rotating wheel. Such solution allows the vehicle to be very maneuverable and, at the same time, to have simple steering and movement direction control through the constant monitoring of the rotational speed of both wheels.

Technological parameters of the vehicle are :

- weight of the robot ready for work including battery is 200 kg,
- it can move at speed 1m/s,
- two direct current motors, each with supply voltage 36 V, were used to drive road wheels, nominal moment 0.55 Nm at speed 3200 rpm, maximum current 27 A, maximum moment 2.8 Nm,

- there are three series connected batteries that supply the robot with voltage 36V, their total capacity 150 Ah,
- the vehicle is able to carry loads up to 100 kg.

Robot supporting structure is created by the steel frame made of the structural sections where all the subassemblies were placed (fig.2). The drive wheels are made of metal and covered with vulcanized rubber rim. The wheel is fixed to the hub placed on the axle going out of the reducer that is driven by the motors through the belt - gear transmission. The gear transmission ratio is 1:2 and the reducer transmission ratio is 1: 40. The self-adjusting wheels are fixed to the robot structure in a way that allows the turn around their own axle (perpendicular to the robot base).

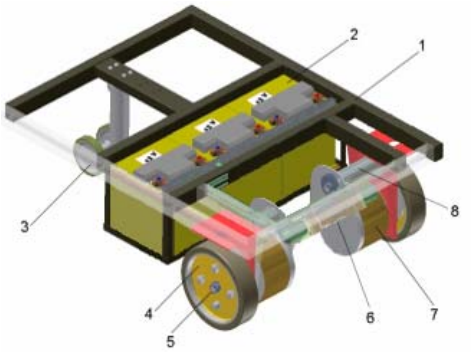


Figure 2. Lay- out of the particular subassemblies: 1 - . robot supporting structure, 2 batteries, 3 - supporting self-adjusting wheel, 4 - drive wheels, 5 - encoder, 6 – cogged pulley, 7 - reducer, 8 - driving motors

The elements used to steer and supply the robot driving motors are the servo-amplifiers SCA-SS-70-10. Automated guided vehicle has two encoders used to measure angular displacements that are to measure an angle and an angular velocity. The base for positioning in this type of vehicle is odometry which introduces some errors. To eliminate these errors the additional sensors allowing to improve the accuracy of vehicle positioning are used in the contemporary vehicle. Such sensors are the laser, sonar and gyro enhanced orientation sensor.

The control system and measurement data processing of an automated guided vehicle

The vehicle movement along the given trajectory is controlled by the onboard guided system. It includes: the onboard PC, cards of data acquisition, measurement sensors and computer software with implemented guided algorithm.

The steering voltage in each measure is generated by the computer on the analog output of the measurement card. Then they are fed on the input terminals of the servo-amplifiers. Inside the servo-amplifier there is the system of the feedback whose aim is to keep the constant rotational speed of the supplied electric motor. This speed depends on the voltage signal from the measurement card of the computer.

Information gathered from the measurement cards is used by the application which works in the Visual C++ environment (fig.3). The application itself takes information, processes it and generates the guided signals. The application was written in a way which enables the users who do not know the software language or the details connected with the vehicle construction and the cooperation of the subsystems to design the trajectory, the vehicle movement on the designed trajectory, reading and analysis of the measurement data obtained from the sensors installed on the vehicle.

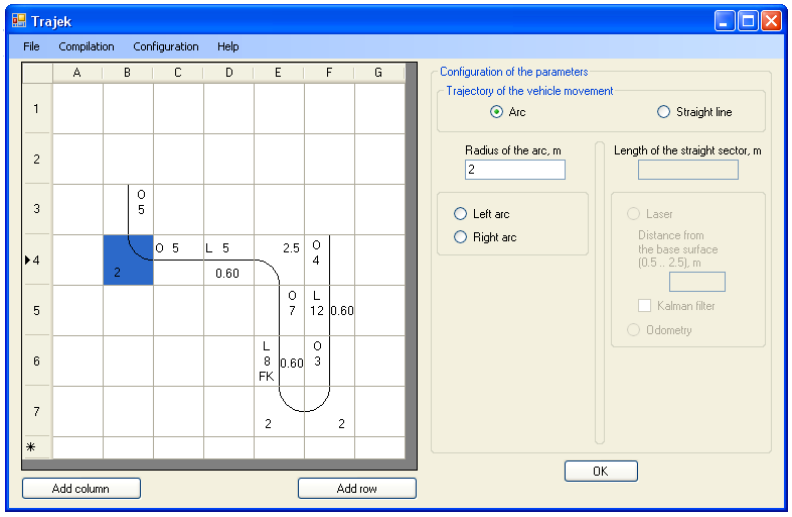


Figure 3. The view of the main window

In the application one can choose some of the basic tasks: trajectory design, edition of the designed trajectory, compilation, reading of the measurement data, analysis of the measurement data.

The application enables for the movement according to the designed earlier trajectory. The trajectory may consist of the elementary segments such as straight line or an arc (fig. 4), for which we can define the right parameters.

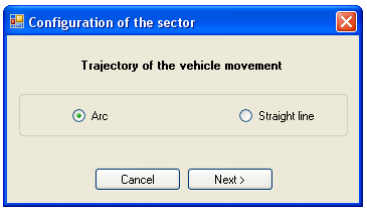


Figure 4. Window of choice of the elementary segment

In case of the straight sector the application enables for the choice of guiding the vehicle: analytical navigation, following along the base surface on the basis of the measurements from the laser sensors. It is also necessary to define the length of the straight sector (fig. 5a).

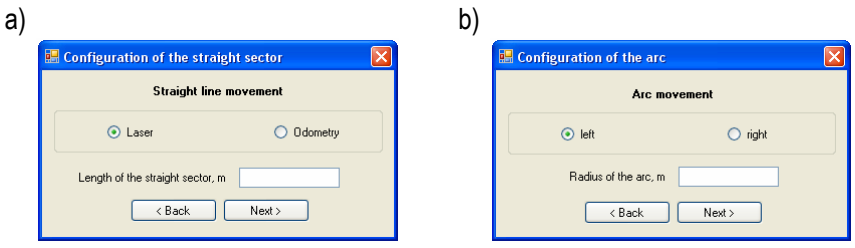


Figure 5. Windows of configuration of the elementary segments

Additionally, when the user chooses the straight sector where the vehicle is to be guided along the base surface on the basis of the reading from the laser sensors, here is a possibility to guide the vehicle in a given distance from the base surface on the basis of the reading from the laser sensors which were not subjected to filtration, he can also choose the filtration option after which the measurements from the laser sensor will be filtered in a real time with the application of the Kalman filter. In case of the following along the base surface it is necessary to quote the distance from the base surface in which the vehicle will be guided. Due to the measure range of the laser sensors the value can change from the 0.5 to 2.5 m. It is also necessary to quote the side where the base

surface will be placed. After choosing the right option the measurement of the distance from the base surface will be done from one of the two sensors installed on the both sides of the vehicle (fig. 6).

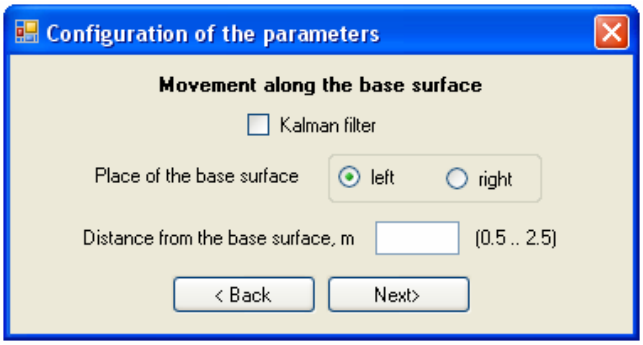


Figure 6. Window of configuration of the straight sector where the vehicle is to be guided along the base surface

After choosing the second option – an arc movement - it is necessary to define whether the vehicle should move in the left or in the right. The radius of the arc should be defined as well (fig.5b).

After choosing the trajectory there is a possibility to write it in. The program enables for the modification of the already designed trajectory by changing the parameters of the chosen elementary sector, change of the sector, and also adding or removing the sector. In order to simplify the analysis of the trajectory, at each element there is a detailed description. It can be obtained by its marking.

This application enables also for the compilation of the designed trajectory. This compilation consists in the check of the correctness of the designed trajectory as for the errors which result from the logical continuity of the trajectory (fig. 7).

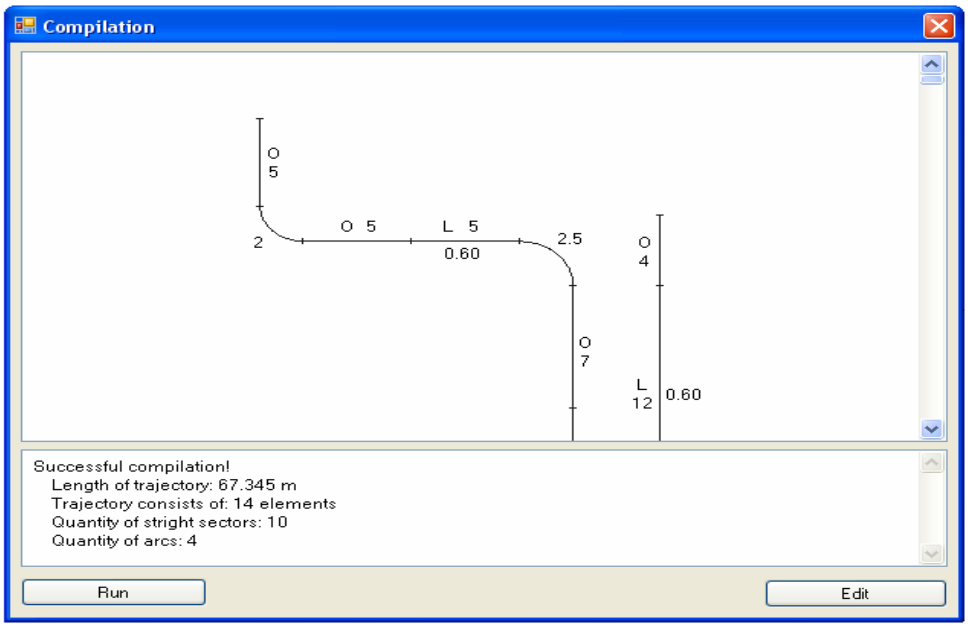


Figure 7. Result of compilation of trajectory

The software enables the reading of the measurement data from the installed in the vehicle sensors. The data is then written in the file. Next it can be analyzed. The application enables then for the edition of the written data, filtering them with use of digital filters, as well as the generating of the chosen graphs (fig. 8b). The user has a possibility to choose what measured quantities will be written in the file (fig. 8a).

a) b)

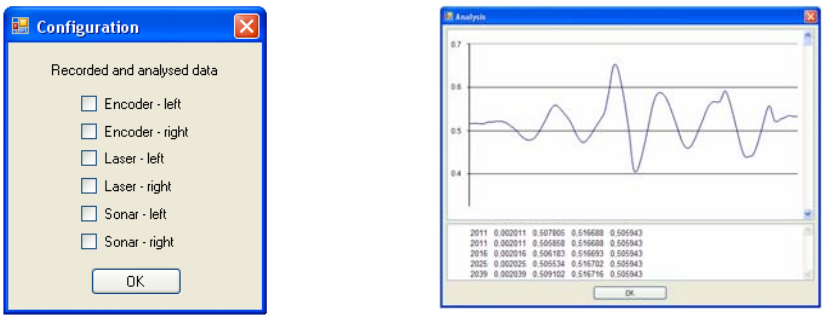


Figure 8. Window of choice of the measurement data designed to record (a) and window of analysis data (b)

To start or to stop the vehicle as well as to give the defined trajectory along which the movement should be performed the second off-board computer is used and it performs the primary function. Two-sided communication of the vehicle with the primary computer is wireless by means of the radio. The application of such a solution was done due to the fact that the creation of the mobile network with the use of the cables is practically impossible. When we apply the cable the vehicle mobility is limited to the length of the transmission cable. In the research for the two-sided communication with the primary computer the stationary wireless network has been used.

Conclusion

The article presents the software written in Visual C++. The software is the part of the system which has been created on need the vehicle. The software will be extended and well-fitting to needs of modified vehicle and carried out on him the investigations.

Bibliography

[Dobrzańska, 2005] M. Dobrzańska. Metody wyznaczania pozycji automatycznie kierowanego pojazdu transportowego. In: VIII Seminarium Naukowe Studentów i Młodych Inżynierów Mechaników, Gdańsk 2005.

Authors' Information

Magdalena Dobrzańska, PhD, Eng *– Rzeszow University of Technology; Department of Computer Science, ul. W. Pola 2 Rzeszow 35-959, Poland; e-mail: md@prz.edu.pl*

Paweł Dobrzański, PhD, Eng *– Rzeszow University of Technology; Department of Computer Engineering in Management, ul. W. Pola 2 Rzeszow 35-959, Poland; e-mail: pd@prz.edu.pl*

THE AUTOMATION OF PARAMETERS IDENTIFICATION PROCESS FOR PROFILES WITH FUNCTIONAL PROPERTIES

Wiesław Graboń

Abstract*: The paper presents a way of software construction for calculation the roughness parameters of surface having stratified functional properties. This software is created for using transition model to describe data becomes from the bearing area curve plotted on normal probability paper and is supported by ISO 13565-3, which describes a way of characterization independent components of two-process surfaces.*

Keywords*: transition model, two-process surfaces, roughness parameters .*

ACM Classification Keywords*: Algorithms, Measurement.*

Introduction

The geometrical structure of surfaces (GSS) which collaborates with machine parts are being influenced by friction, lubrication and wear, therefore lubricated and sliding surfaces are made as surfaces having stratified functional properties which are named two-process surfaces. These surfaces are the most critical and tribological structure of surfaces. The most often, they are created by putting up several technological processes. They are characterized by occurrence deep valleys on precisely machined top surface layer. Top part of surface, usually smooth is bearing surface and its main goal is to reduce wear, but parts of valleys aim at oil storage and is a trap on small pieces of wear. Plateau honed cylinder surface is the typical example. The topography of this surface is an effect of two processes: base honing and plateau honing.

The precise description of manufactured 2-process surface is a vital problem from practical point of view. There are a lot of methods of two-process surface topography description. One of the alternative approach is to model the surface into two parts, one representing the plateau and the other representing the valley. This becomes possible if the bearing area curve is plotted on normal probability paper. The x-axis of the probability plot is a linear scale of standard deviations and y-axis represents profile height in micrometers [Sannareddy, Raja, Chen, 1998]. If the data from a normal distribution is plotted on a probability paper, then all the data point will fall along a straight line (fig. 1a1). Slope of this line is interpreted as Rq value [Sannareddy, Raja, Chen, 1998]. As Fig. 1.a1 shows, the profile illustrates GSS after base honing (high peaks, and deep valleys), its image at Fig 1.a3 is a large slop line which is equal to big value parameter Rq. The second profile (Fig. 1.b1 – permanent line) describes parts of structure remained before last process which image is less slope line, which corresponds to low value of parameter Rq. Only the last profile (Fig. 1.c1) is measured. In this profile only the dippiest valley the base roughness surface occur, which corresponds to the bottom part of more slope line (Fig. 1.c3). The bottom part of origin roughness profile is removed and replaced by less roughness structure of surface (which is called - plateau), which corresponds to the bottom part of less slope line (Fig. 1.c3). The intersection point on normal probability graph of abscissa Rmq defines the separation of plateau and base textures and is an important feature of the model. The proposed plateau roughness Rpq, valley roughness Rvq and Rmq are three parameters characterising plateau honed surface. Those parameters seem to be of great importance because they are connected with honing process parameters. Therefore automation of determination those parameters have important influence on control a process creating those surfaces.

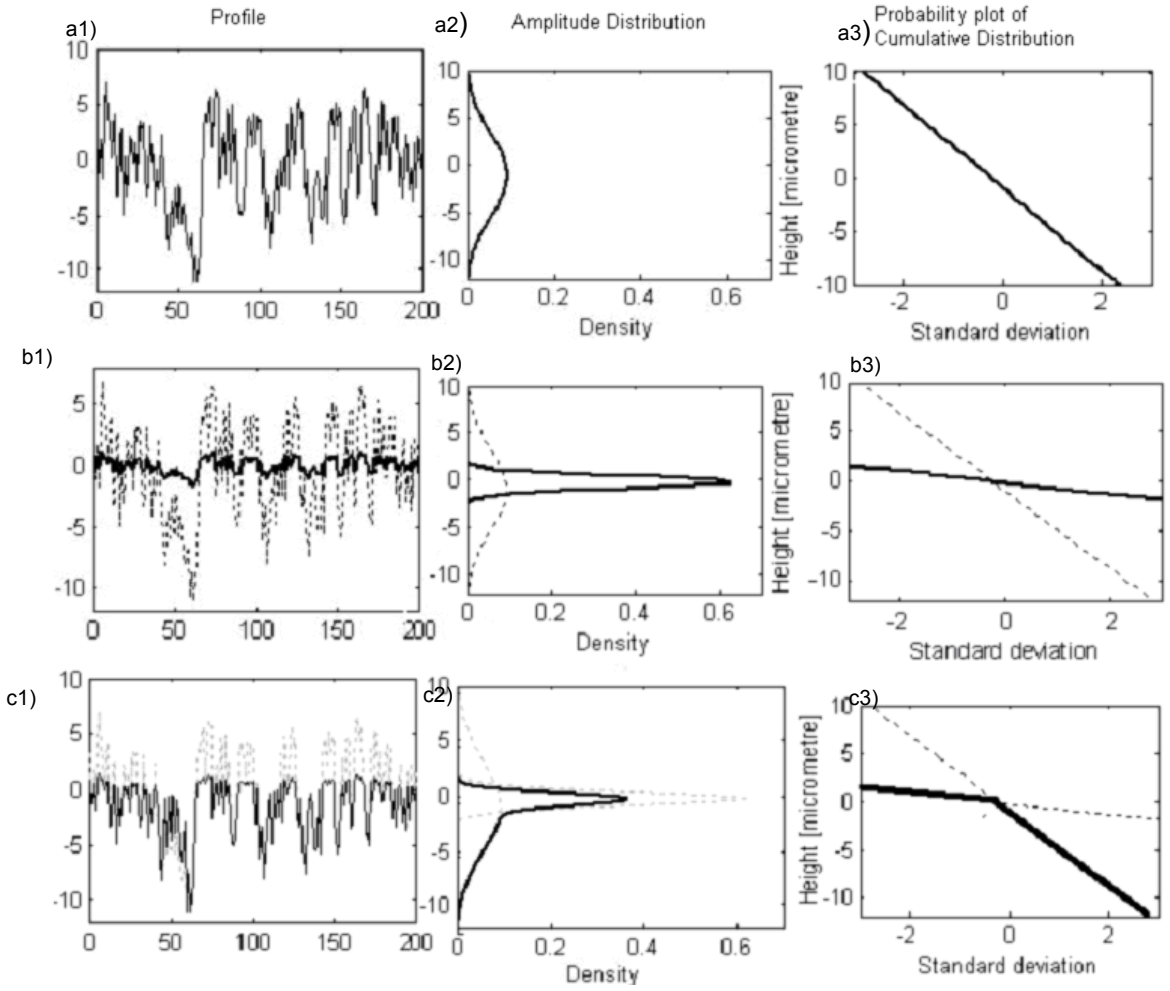


Figure 1. Graphical interpretation of parameters included in ISO 13565-3 standard

Model analysis

In simulation profile the abrupt transition between platau and valley takes place. Finding the transition point is possible by approximation the material probability curve by model 1 suggested by the authors of paper number 2.

$$Y = a_o + a_1(x - x_0) + a_2(x - x_0) \text{ sign } (x - x_0) \quad (1)$$

Where

sign - signum function

Y = sign(X), where each element of Y is:

1 if the corresponding element of X is greater than zero

0 if the corresponding element of X equals zero

-1 if the corresponding element of X is less than zero

The parameter (x_0 , a_o) determine the location of the join point, two straight lines are slopes $(a_1 – a_2)$ and $(a_1 + a_2)$ respectively [Watts, Bacon, 1971]. All parameters (a_o, a_1, a_2, x_0) which appear in this model can by evaluated on the base of date through they approximation by model 1.

Fig. 2 shows real profile roughness of 2-process surfaces and corresponds to graph of material probability curve.

In this profile unstable area of transition (curve) between part of the top area plateau and the area of valley which is caused by mixture of two distribution (marked as 3) can be distinguished.

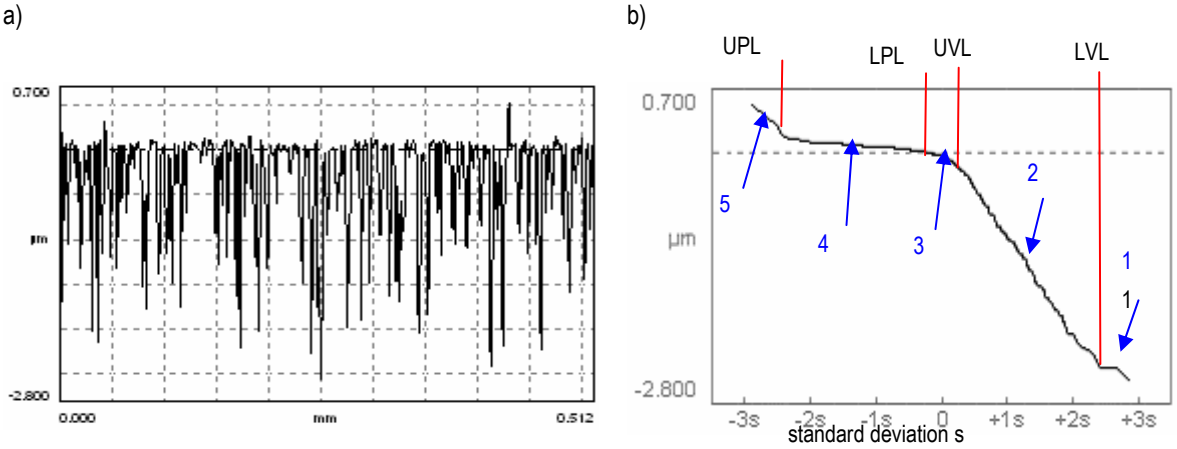


Figure 2. Example of roughness profile and corresponding graph of material probability curve

What is more, two linear areas (2 i 4) on the graph appear, which correspond to component profiles having a normal ordinate distribution. Moreover, this graph has two nonlinear areas, which come from:

- debris or outlying peaks in the data (profile) (labeled 5),

- deep scratches or outlaying valleys in the data (profile) (labeled 1) [ISO 13565-3].

The upper boundary of region (4) – "plateau" describes point Upper Plateau Limit (UPL), lower boundary of region – point Lower Plateau Limit (LPL), separating linear area (2) from nonlinear areas (5 and 3). The upper boundary of the region (2) – valley indicates point Upper Valley Limit (UVL), lower – point Lower Valley Limit (LVL), separates linear area (2) from nonlinear areas (3 and 1). To define value of parameter Rq from the region plateau and the region of valley properly, storage parts correspond to normal distribution should be separated.

Model (1) is only appropriate if it is known that an abrupt transition between plateau and valley takes place. In real profile the nature of the transition point from plateau to valley is not know a priori. Therefore was proposed more general model worked out by the authors paper number 2, which permits a smooth transition from plateau to valley by replacing the sign function sgn(x-x0) in 1 by transition function

trn $\{(x - x_0)/\gamma\}$

There are many transition functions which could be used but for this research used hyperbolic tangent function:

trn $\{(x - x_0)/\gamma\} = \tanh\{(x - x_0)/\gamma\}$

This function satisfy the conditions specified in [Watts, Bacon, 1971].

The genaral model (1) thus becomes

$$Y = a_o + a_1(x - x_0) + a_2(x - x_0)\tanh\{(x - x_0)/\gamma\} \qquad (2)$$

The parameters x_0 and γ determine the location of the join point and the radius of curvature of the model at the join point. In the limit, as γ approaches zero, model (2) devolves to two straight lines of slopes $(a_1 - a_2)$ and $(a_1 + a_2)$ respectively, intersecting at the coordinates (x_0, a_o). For nonzero γ model (2) is asymptotic to these two

lines at values of x distant from the join point x_0 but at x = x_0 the curve passes through the join point (x_0, a_o). The radius of curvature R for the model (2) at the join point x_0 is:

$$R = \frac{\gamma}{2|a_2|}\left(1 + a_1^2\right)^{\frac{3}{2}}$$

Thus γ could be referred to as the radius of curvature parameter [Watts, Bacon, 1971].

The structure of program

To automatize the determination process of parameters Rq in plateau and valley area, computer program was created. This program was partly based on algorithm described in ISO 13565-3 standard. To find preliminary transition point from plateau to valley area, material probability curve graph was approximated (used nonlinear regression - algorithm Levenberg-Marquardt) by model 2. All parameters (a_o, a_1, a_2, x_0, γ) appear in this model can be estimated on the base on data by their approximated used model 2. Next graph point of value ordinate equal x_0 was used as preliminary transition point from valley to the plateau regions. According to methodology which is recommended in ISO 13565-3 standard, nonlinear material probability curve graph regions were eliminated (Fig. 2 – regions 1 and 5) and UPL and LVL points were assigned. Following step - material probability curve graph was normalized between points UPL and LVL towards y-axis coefficient:

$ks = (Y_{UPL} - Y_{LVL}) / (X_{UPL} - X_{LVL})$

and one more approximated in area between UPL and LVL using model (2).

In normalized graph lower boundary of the region plateau (UPL) and upper region valley (UVL) were determined by elimination *n* points which are situated partly right and partly left from point x_0. The number of eliminated points was determinated from the value of radius of curvature *R*. In unnormalized graph linear regression lines between points UPL and LPL and between points UVL and LVL were determinated. Directional coefficients values of these lines were assigned as values Rpq and Rvq. Rmq parameter was assigned as value of abscissa in intersection of regressions lines assigned in plateau and valley region - Fig.3.

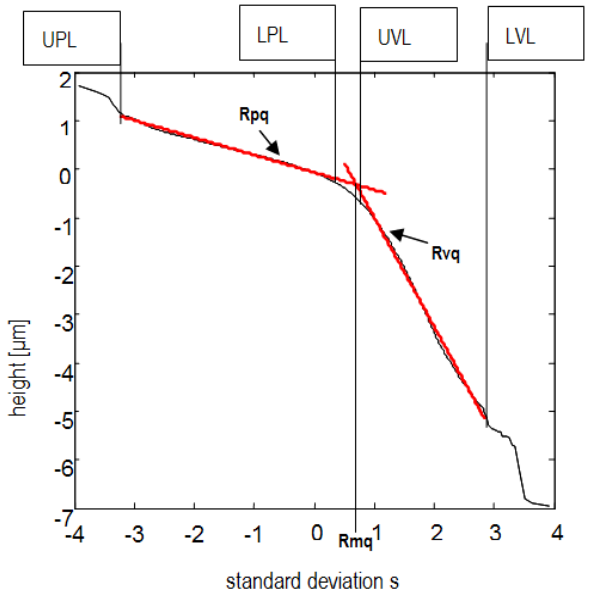


Figure 3. The graph of material probability curve and regression lines situated in plateau and valley regions.

Examples of results

In the aim of checking if prepared software assure proper value of roughness parameters, two-process profiles programmed Rpq, Rvq and Rmq parameters were modeled. The profiles after 2 processes were simulated using procedure presented in reference [Pawlus, 2008]. About quality of estimation these parameters one can find out on the bases of comparison parameters obtained after applying prepared software to programmed in model parameters (Table 1)

Table 1. Example values of parameters modeled profile of two-process surfaces (contains 8000 points) received with using suggested software.

Profile	Output parameters			Suggested software		
	Rmq [%]	Rpq [µm]	Rvq [µm]	Rmq [%]	Rpq [µm]	Rvq [µm]
T1	84.1	0.1	4	88.78	0.15	3.843
T2	15.9	0.1	4	11.81	0.17	4.25
T3	80.1	0.3	2.47	82.70	0.336	2.74
T4	50	0.6	5	53.87	0.56	4.89
T5	97.7	0.1	6	94.85	0.18	5.12
T6	78.5	0.36	2.25	75.33	0.312	1.92
T7	78.8	0.36	4.54	81.98	0.345	4.12
T8	84.1	0.3	4	89.15	0.41	3.703

In case of analizing modeled profiles, it was found that the most proper results one received when values of parameter Rmq were in range from 17% to 82%. In other cases obtained results differed from values come from the given model. Figures 4 shows the examples of the parameters calculation results for sample T4.

a) b)

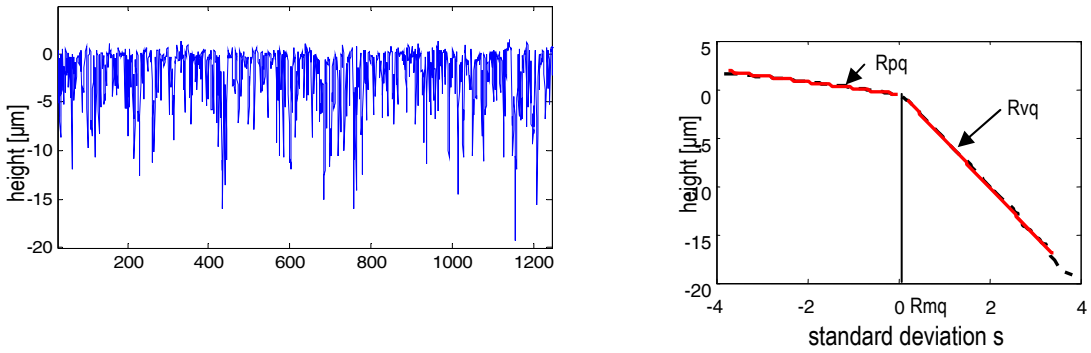


Figure 4. Modelled two-process profile (a) (sample T4), probability plot of height distribution (b) Ppq = 0.56 µm, Pvq = 4.89 µm, Pmq = 53.89%

Real profiles were also tested. While analysing measured profiles one can define accuracy of determinated parameters on the base of approximated graph material probability curve. It was found that the errors of parameters are bigger when the ratio of Pvq to Ppq were very small.

Conclusion

In the aim of better understanding the properties of surfaces having stratified functional properties and their connection with using properties, precise description should be applied. Accurate description of these surfaces using only one parameter is very difficult, because independent components appear which should be characterized very precisely and rather separately.

Independent description of particular components of those profiles enables their better understanding. It can make a contribution for better progress in designing these types of surface structure. Automatic determination of surfaces parameters having stratified functional properties improves their process of control. Proposed software reach out this mentioned expectation.

Bibliography

[Sannareddy, Raja, Chen, 1998] H. Sannareddy, J. Raja, K. Chen. Characterization of surface texture generated by multi-process manufacture. Int.J. Mach. Tools Manufact. Vol. 38, Nos 5-6, pp. 529-536, 1998.

[Watts, Bacon, 1971] D.G. Watts, W. Bacon. Estimating the Transition between Two Interseecting Straight Lines Biometrika Vol. 58, No. 3, 525-534, 1971.

[Pawlus, 2008] P. Pawlus: simulation of stratified surface topography. Wear 264 (5-6), 457-463, 2008.

[ISO 13565-3] ISO 13565-3 Geometrical Product Specifications (GPS) – Surface texture: Profile method; Surfaces having stratified functional properties, Part 3: Height characterization using the material probability curve.

Authors' Information

Wiesław Graboń *– Department of Computer Science, Rzeszow University of Technology; W. Pola 2, 35-959 Rzeszów, Poland ; e-mail: wgrabon@prz.edu.pl*

PREDICTION OF STYLUS TIP RADIUS

Sławomir Górka, Paweł Pawlus

Abstract: *The existing method of stylus tip radius prediction was analysed. Computer generated and measured stochastic surfaces were the objects of the investigations. The effects of various sources of errors on tip radius prediction accuracy were analysed. On-line skid existence doesn't affect this accuracy, but quantisation errors caused some problems. The application of low-pass digital filtering in order to eliminate high-frequency noise was analysed.*

Keywords: *surface topography; roughness; stylus measurement; tip radius*

ACM Classification Keywords: *J.2 Physical Sciences And Engineering; G.1.2 Approximation; G.1.3 Numerical Linear Algebra;*

Introduction

Operational properties, like materials contact, sealing, friction, lubricant retention and wear resistance are related to surface topography. Stylus measurement is the most commonly used for surface topography measurement. The stylus tip has finite dimension (radius of curvature of 2, 5 or 10 micrometers). The mechanical filtering behaviour of tip depends on not penetrating irregularities of wavelengths smaller than its radius. The increase of spherical tip size causes profile distortion (decrease of the Rq parameter and increase of the RSm parameter) [Whitehouse, 1974].

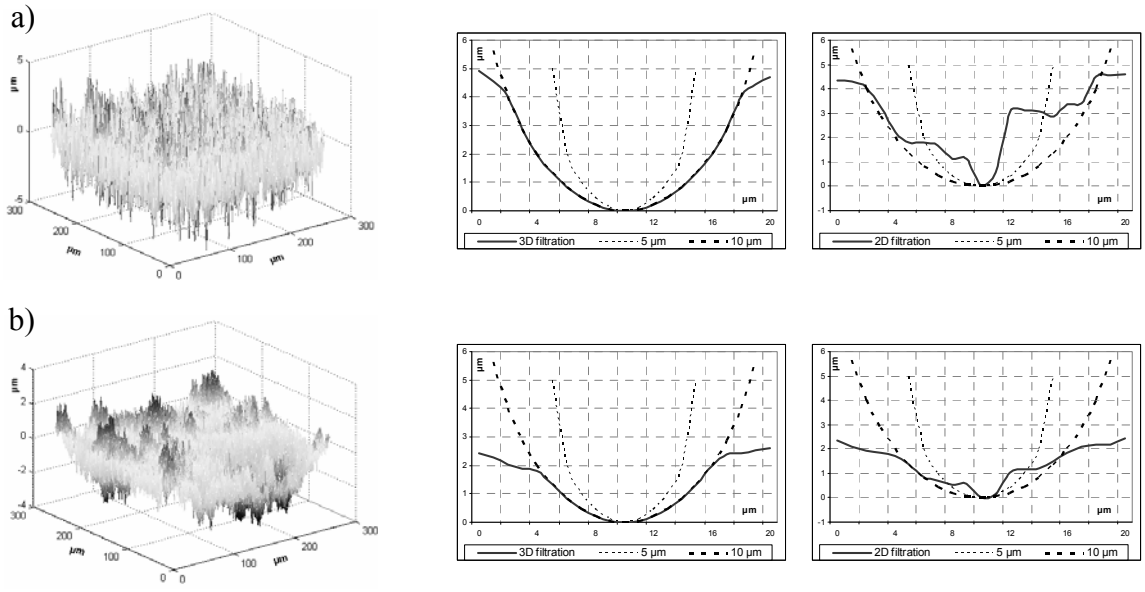


Figure 1. The isometric views (left graphs) of surfaces subjected to mechanical filtration and the shapes of reconstructed stylus tip in measurement direction after using 3D ball (middle graphs) and 2D wheel (right graphs), a) isotropic surface, b) anisotropic surface

Due to large distortion of profile, the scientists try to reconstruct of measured surface. However the surface reconstruction can be done only when the tip geometry is known. Usually the mathematical morphology erosion (lower envelope) operation was applied [Villarubia, 1994; Villarubia, 1997].

However the profile analysis can lead to correct reconstructed tip shape when anisotropic unidirectional surface is measured perpendicularly to main texture direction (across the lay). When measurement direction is different and isotropic surface is measured the errors of tip reconstruction basing only on the profile analysis can be great. Figure 1 presents the views of the anisotropic and isotropic surfaces subjected to mechanical filtration by the 3D spherical stylus and the shapes of the stylus tip in the measurement direction after reconstruction by 3D ball and 2D wheel [Górka, 2006].

The authors of paper [Dongmo, 1996] presented method of the radius of tip estimation. It depends on an erosion (lower envelope) following by a dilation making of image when radius of modelled tip "r" increases. Erosion followed by dilation is called an opening procedure. The difference between the open image and the experimental image can be assessed repeating the opening procedure, for various r values and calculating the differences and allows to fix the upper limit for the effective tip radius (see Figure 2). The special software was implemented by the authors of this paper. The effect of various errors on tip radius reconstruction was studied.

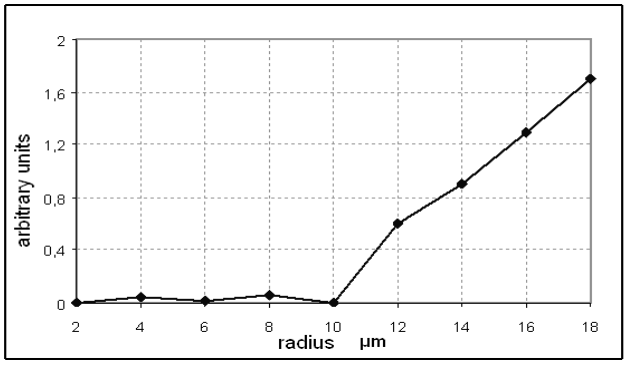


Figure 2. Plots of differences between open and dilated (measured) profiles

Results and discussion

Firstly, the analysis of simulated profiles will be done. The computer generated profiles by the present author were subjected to simulated mechanical filtering by 2D wheels of commonly used radii (2, 5 and 10 µm). In the majority of cases it was possible to obtain correct results. Figure 3 shows example of plot of differences. The radius of simulated stylus tip was 10 µm. The negligible effect of on-line skid usage on profile distortion independently on the skid radius was found.

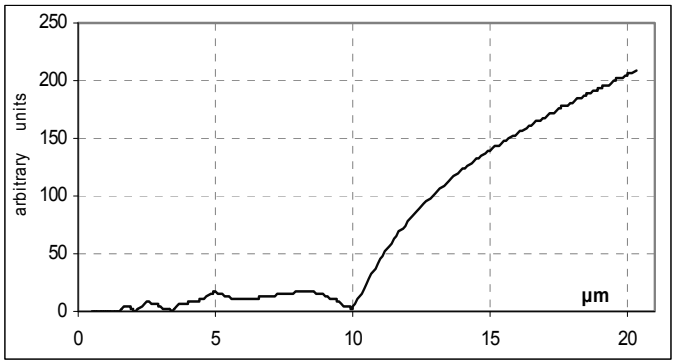


Figure 3. Plots of differences between open and dilated (measured) profile

The influence of quantisation errors on the stylus tip radius reconstruction accuracy was analysed. This error caused lift of the difference graph. Often the small number of amplitude levels was sufficient to correct tip radius estimation (50-100). Figure 4 presents example of the effect of quantisation errors on total difference graph.

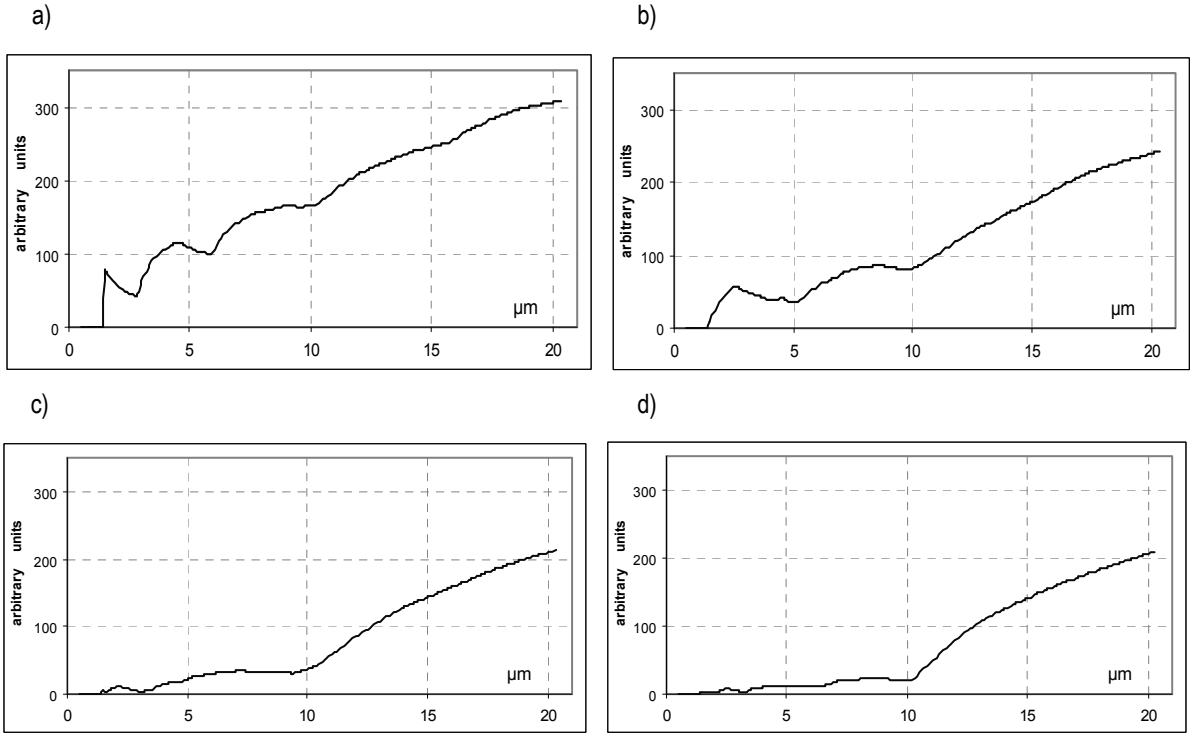


Figure 4. Plots of differences between open and dilated (measured) profiles for 10 (a), 20 (b), 50 (c) and 100 (d) height levels

The computer generated high-frequency noise was added to profiles after modelled mechanical filtration. This procedure also caused the lift of graphs, of character different to the changes caused by quantisation errors. Often the existence of high-frequency noise allowed to correct estimate of the tip radius. Figure 5 presents the effect of high-frequency noise existence on the shape of difference graph. Connection of profiles prior to tip radius estimation can be helpful.

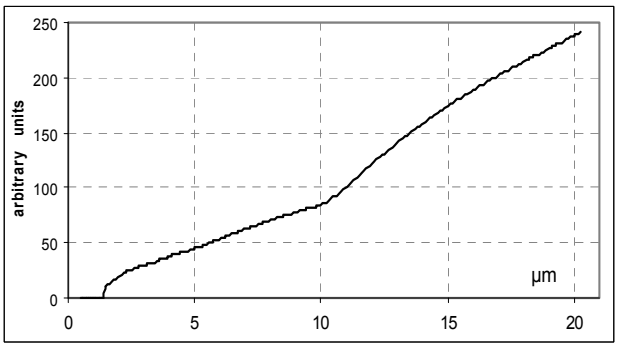


Figure 5. Plot of the difference between open and dilated (measured) profile for high-frequency noise existence

The high-frequency noise effect can be decreased by low-pass digital filtration if the main wavelength of noise is lower than tip radius. Cut-off of filter should be smaller than tip radius. For example when the main wavelength of noise was small and the tip radius was 10 μm after when the cut-off was selected precisely low-pass filtering can improve the accuracy of the tip radius estimation.

The tip radius was predicted in the measurement direction when surface topography was measured across the lay. In order to obtain precisely tip radius one should plot tangents of the two linear parts of the obtained curve searching for the point of their crossing. However often it is difficult to achieve straight lines fragments. Connect profiles is the better possibility. The results improvement is then very possible. Figure 6 presents the example of results obtained using this method.

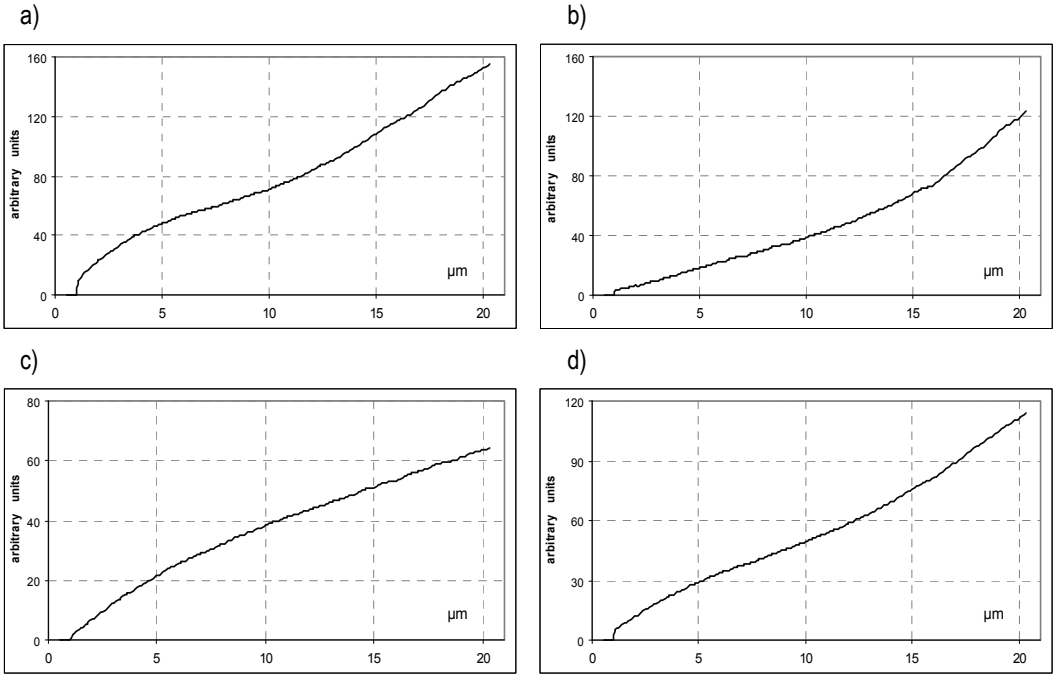


Figure 6. Plots of the differences between open and measured profiles (a, b, c) and connected profile (d). Probe tip of radius was 10 µm

This method can be further improved by using the low-pass short wavelength digital Gaussian filter. After the noise elimination the curve moves down. The results can be improved after profile connections. The results concerning the profiles which difference plots were presented in Figure 6 are shown in Figure 7.

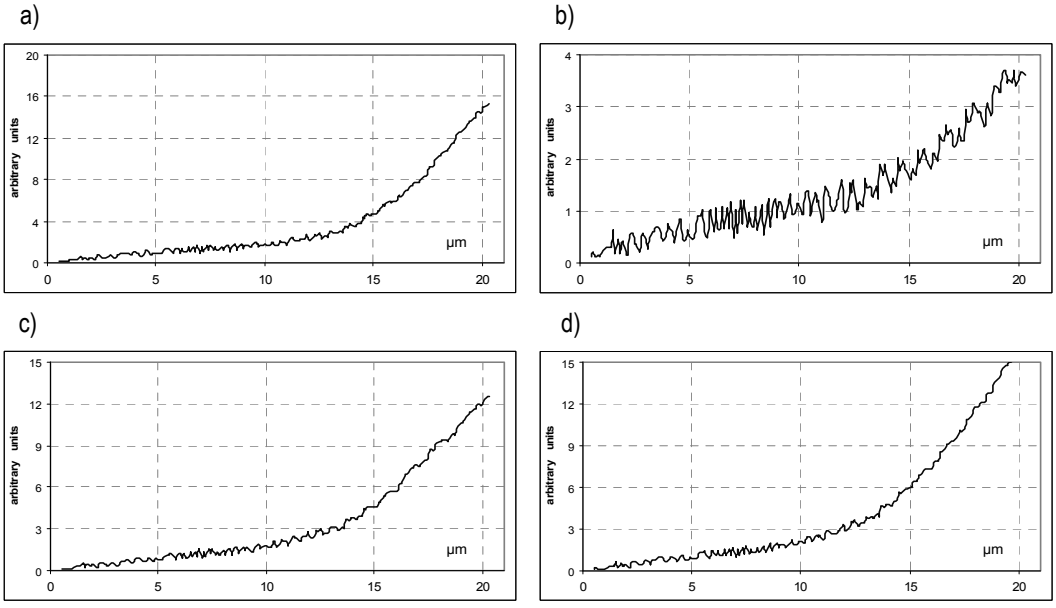


Figure 7. Plots of the differences between open and measured profiles (a, b, c) and connected profile (d) when low-pass digital Gaussian filter was used (cut-off was 10 µm)

Conclusions

The application of the analysed methods of stylus tip probe radius estimation leads to proper results. This method seems to be robust. On-line skid doesn't affect the accuracy of the tip radius prediction. However quantisation errors caused tip estimation difficulties. The use of low-pass digital frequency filtration can be the method of high-frequency noise elimination. It should be used when the noise main wavelength is smaller than the probe tip size. The connection of the profiles causes improvement of stylus tip radius estimation.

Bibliography

[Whitehouse, 1974] D. J. Whitehouse. Theoretical analysis of stylus integration CIRP Annals 23/1 1974 181-182

[Villarubia, 1994] J.S. Villarubia. Morphological estimation of tip geometry for scanned probe microscopy Surface Science, 1994 321 287-300

[Villarubia, 1997] J.S. Villarubia. Algorithms of scanned probe microscope image simulation, surface reconstruction, and tip estimation. J.Res. Natl. Stand. Technol., 1997 102 425-454

[Górka, 2006] S. Górka. The effect of measuring errors on accuracy of surface topography represebtetion. PhD Thesis, Rzeszów University of Technology, 2006, Rzeszów

[Dongmo, 1996] S. Dongmo, M. Troyon, P. Vautrot. Blind restoration method of scanning tunneling and atomic force microscopy images J. Vac. Sci. Technol, 1996 B14(2) 1552-1556

Authors' Information

Sławomir Górka *– PhD, Rzeszów University of Technology, Department of Computer Science; W. Pola 2, 35-959 Rzeszów, Poland ; e-mail: sgorka@prz.edu.pl*

Paweł Pawlus *– Associate professor, DSc, Rzeszów University of Technology, Department of Manufacturing Technology and Production Organisation; W. Pola 2, 35-959 Rzeszów, Poland ; e-mail: ppawlus@prz.rzeszow.pl*

DYNAMIC SYSTEM QUALITY PROVIDING UNDER UNDETERMINED DISTURBANCES. ONE-DIMENSIONAL CASE

Iryna Zbrutska

Abstract: *One-dimensional dynamic system under impact of the undetermined disturbing influences is reviewed. The possibility to affect only the value of disturbances is conceded. An algorithm for system quality estimation and making decision about control aiming to provide required quality is proposed. Control algorithm for one-dimensional system of any order is developed.*

Keywords: *dynamic system quality; undetermined disturbances; condition estimation; resulting disturbance control; control algorithm; quality function.*

Introduction

Quality maintenance of the dynamic systems under random undefined disturbing influences does not have any unequivocal solution. Existing methods presuppose either complete a priori information about disturbances, or their constraints are known [Lin, Su, 2000], [Poliak, Sherbakov, 2002], [Nikiforov, 2003], [Hou, Muller, 1992], while regulators with dynamic disturbance compensators might have high dimensions [Liubchyk, 2007].

System Condition Estimation

Here we review a steady linear system with one input and one output. Input F and output α value correlation is set by the dynamic operator $\Phi(p)$

$$\Phi(p)\alpha = F\,, \tag{1}$$

where $p = \frac{d}{dt}$ – time differentiation operator.

α variable defines system functioning quality when under control, and drift from the needed quality (error) when under disturbances. Without problem contraction, it is possible to review linear system (1) behavior but only under disturbances.

Using system (1) weight function $w(t)$, it becomes possible to estimate input value $\alpha(t)$ under disturbances $F(t)$,

$$\alpha(t) = \int_0^t F(\tau)w(t-\tau)d\tau = \int_0^t \psi(\tau)d\tau\,, \tag{2}$$

$$\psi(\tau) = F(\tau)w(t-\tau)\,. \tag{3}$$

Proceeding from geometrical interpretation of expression (2), it is possible to estimate output value of the system by its square, described in time by the function $\psi(\tau)$ on the observation interval (0, t) (Fig.1).

On the other hand, if the system (1) is observable, we can estimate the disturbance F(t). If $\alpha(t)$ can be differentiated n times (n- system (1) operator p polynomial order (1)), its derivatives can be received after $\alpha(t)$ measuring.

In this way, to estimate system quality it is sufficient to have to have as time function $\psi(\tau)$ depending on the acting disturbances and system (1) dynamic features.

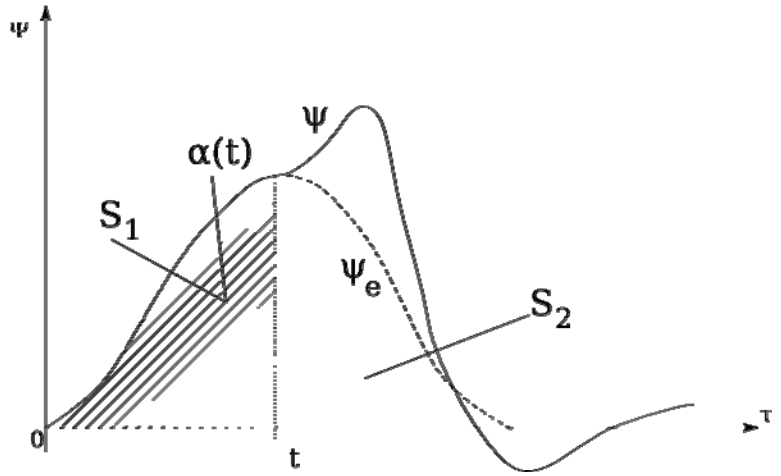


Figure 1. $\alpha(t)$ *estimation*

Making Decision about Starting Disturbance Control

Here we review system (1) quality providing as definition of its drift α (input value bias from the program value) in allowed ranges.

Using geometrical interpretation of (2) to provide needed system quality it is necessary and sufficient to provide function $\psi(\tau)$ (quality function) value in the range allowed (quality range). Its square S_Q does not exceed maximum value α_{al} of on the interval of constant-sign $\psi(\tau)$ function,

$$S_Q \le \alpha_p, \quad t \in (t_1, t_2), \; S_Q \in (S <> 0) \tag{4}$$

Quality range may be constructed as follows. Let us assume that system (1) is a control object in a feedback control system undergoing external disturbance F_{ext}. Then, $F(1)$ is a resulting disturbance, the aggregate of the external disturbance and feedback F_y action. If the system is steady, closed, and of the needed quality when under typical disturbances, resulting disturbance $F(1)$ has variable sign, both under constant-signed and variable signed external disturbances. Let us estimate (or measure) value on interval $t \in (0, t_k)$, where t_k - moment of time, when

$$\alpha(t_k) = \varepsilon\alpha_p, \quad 0 < \varepsilon < 1, \tag{5}$$

and quality function (3)

$$\psi(\tau) = F(\tau)w(t_k - \tau), \quad \tau \in (0, t_k) \tag{6}$$

Range (2) S_1 for the function (6) will thus fill a part of quality range (4) S_Q.

With disturbance type undefined, let us set $\varepsilon = 0.5$. Then the second part S_2 of the quality range can be constructed as function (6) mapping relative to $t = t_k$ at the range $t > t_k$:

$$\psi(t) = \psi(2t_k - t). \tag{7}$$

Meanwhile it is necessary that

$$S_1 + S_2 \le S_Q, \quad S_1 = \alpha(t_k), \quad S_2 = \alpha(t) - \alpha(t_k), \tag{8}$$

and $t = t_k$ is a moment of disturbance control start.

Disturbance Control Start

Let us define the possibility of controlling disturbance value when it acts over the control object (1) included in the closed system. After having implemented a closed control system, as shown in the Fig. 2, under certain constraints on control operator $\Phi_y(p)$ parameters,

$$\alpha = \frac{\Phi^{-1}(p)}{1 + K\Phi^{-1}(p)\Phi_y} F_{ext} = \frac{p}{KH(p)} F_{ext} = \frac{W(p)}{K} F_{ext},$$

$$F = \frac{p\Phi(p)}{KH(p)} F_{ext} = \frac{W(p)\Phi(p)}{K} F_{ext}. \tag{9}$$

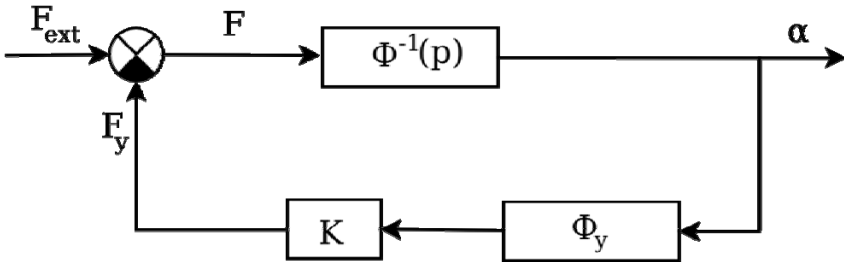


Figure 2. Control system. Example 1

Without constraints on the control operator we can get a similar result in the system with double feedback (Fig. 3), where operator w defines closed system (8):

$$\alpha = \frac{W(p)}{1 + K} F_{ext}, \quad F = \frac{W(p)\Phi(p)}{1 + K} F_{ext}. \tag{10}$$

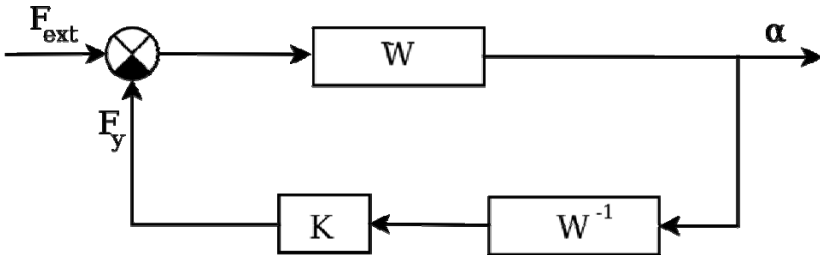


Figure 3. Control system. Example 2

Thus, in the control systems (9), (10) external disturbances are varied with the operator K by its value. In this way disturbance alteration remains unchanged both for the resulting disturbance and for the input value.

Let us define a control algorithm according to (7) and (8). Let us set quality value (7) as etalon.

$$\psi(2t_k - t) = \psi_e(t), \quad t > t_k, \quad \alpha > \alpha_e. \tag{11}$$

Let us name the difference between real α and etalon α value for the input

$$y = \alpha - \alpha_e = \int_0^t [\psi(\tau) - \psi_e(\tau)]d\tau, \quad |\alpha| > |\alpha_e|, \tag{12}$$

and use direct Liapunov method for control $K(t)$ definition in the case (10), or $K_*(t)$, $K(t) = 1 + K_*(t)$ for (9). Let us introduce Liapunov function $V = y^2$. Then it is necessary to provide

$$\frac{dV}{dt} = 2y\frac{dy}{dt} < 0, \tag{13}$$

where

$$y(t) = \int_0^t [\psi_k(\tau) - \psi_e(\tau)]d\tau, \frac{dy}{dt} = \psi_k(t) - \psi_e(t),$$

$$\psi_k = \frac{\psi}{1+K} \qquad (\psi_k = \frac{\psi}{1+K_*}). \tag{14}$$

Condition (13) is met if

$$K(t) > \frac{\psi(t) - \psi_e(t)}{\psi_e(t)}, \quad t > t_k, \quad |\alpha| > |\alpha_e|, \quad K(t) > 0. \tag{15}$$

For that operator K formation rule may be chosen as one of the following:

$$K(t) = \left[\frac{\psi(t)}{\psi_e(t)} - 1\right]\left[1 + sign(|\alpha| - |\alpha_e|)\right]\left[1 + sign(\psi - \psi_e)\right], \tag{16}$$

$$K(t) = \left[\frac{\psi(t)}{\psi_e(t)} - 1\right]\left[1 + sign(|\alpha| - |\alpha_e|)\right]\left[\frac{\alpha}{\alpha_p} - \varepsilon\right]^2\left[1 + sign(\psi - \psi_e)\right] =$$
$$= \frac{y}{\psi_e}\left[1 + sign\frac{dy}{dt}\right]\left[1 + sign(|\alpha| - |\alpha_e|)\right]\left[\frac{\alpha}{\alpha_p} - \varepsilon\right]^2. \tag{17}$$

Operator becomes undefined when $\psi_e = 0, y > 0$.

To avoid uncertainty in (16), (17) algorithms it is sufficient to put

$$K(t) = K(\psi_e \to 0_+, y > 0) = const \tag{18}$$

on the time interval starting from the moment (18) until the moment when $\psi(t) = 0$.

Conclusion

Undefined external disturbance control rule described is based on the estimation not the disturbance itself but dynamic quality function system features, depending on it. Such estimation allows to forecast the possible scale of quality function changes and to make a decision about control start. Control algorithm, developed on the basis of current and estimated quality function values, provides the necessary quality of the one-dimensional dynamic system.

Bibliography

[Hou, Muller, 2002] Hou M., Muller P.C. design of observers for linear systems with unknown inputs. IEEE Trans. On Automatic Control, 37, 871-875.

[Lin, Su, 2000] Lin C.-L., Su H.-W. Intelligent Control Theory in Guidance abd Control System design: an Overview, Proc. Natl. Sci. Counc. ROC(A), 24(1), 15-30.

[Poliak, Sherbakov, 2002] Poliak B.T., Sherbakov P.S. Robust Stability and Control. Moscow, Nauka, RU.

[Nikiforov, 2003] Niliforov V.O. Adaptive and Robust control with Measuring Compensation. St.-P., Nauka, RU.

[Liubchyk, 2007] Liubchyk L.M. Reverse Dynamic Model Method in Problems of Synthesis of the Multi-Dimensional Combination Systems by Disturbance Observers. Radio-electronic and Computer Systems, 5, 77-83, UA.

Authors' Information

Iryna Zbrutska *- Master, National Technical University of Ukraine "Kyiv Politechnic Institute", P.O.Box: 37, Peremogy Av., Kyiv, 03056, Ukraine; e-mail: zbrutsky@cisavd.ntu-kpi.kiev.ua*

ITA 2010: JOINT INTERNATIONAL SCIENTIFIC EVENTS ON INFORMATICS

June 21 - July 03, 2010, Varna (Bulgaria)

ITA 2010 Main Scientific Events:

KDS

XVI[th] International Conference "Knowledge – Dialogue – Solution" is traditionally dedicated to discussions on current research and applications regarding three basic directions of intelligent systems development: Knowledge Processing, Natural Language Interface, and Decision Making.

i.Tech

Eight International Conference "Information Research and Applications" is devoted to discussion of current research and applications regarding the basic directions of computer science.

MeL

Fifth International Conference "Modern (e-) Learning" is oriented to discussion of current research and applications in the basic directions of modern (e-) learning: Philosophy and Methodology of the Modern (e-) Learning and Modern (e-) Learning Technologies.

INFOS

Third International Conference on Intelligent Information and Engineering Systems is devoted to current research and applications regarding the knowledge-based intelligent information and engineering systems.

CFDM

Second International Conference "Classification, Forecasting, Data Mining" is directed to the fields of Classification, Clustering, Pattern Recognition, Forecasting, Features Processing, Transformations, Data Mining, and Knowledge Discovery.

ISSI

Fourth International Summer School on Informatics is devoted to discussion of current research, applications and education regarding the basic directions of informatics.

ITA 2010 is organized by:

ITHEA International Scientific Society
Association of Developers and Users of Intelligent Systems (Ukraine)
Association for Development of the Information Society (Bulgaria)

Institute of Mathematics and Informatics, BAS (Bulgaria)
Institute of Information Technologies, BAS (Bulgaria)
National Laboratory of Computer Virology, BAS (Bulgaria)
Institute of Cybernetics "V.M.Glushkov", NAS (Ukraine)
Institute of Mathematics of SD RAN (Russia)
Institute of Information Theories and Applications FOI ITHEA (Bulgaria)
Dorodnicyn Computing Centre of the Russian Academy of Sciences

Astrakhan State Technical University (Russia)
Ben Gurion University (Israel)
Kharkiv National University of Radioelectronics, (Ukraine)
National University of Kiev "Taras Shevchenko"(Ukraine)
Rzeszow University of Technology (Poland)
Universidad Politecnica de Madrid (Spain)
University of Calgary (Canada)
University of Hasselt (Belgium)
Varna Free University "Chernorizets Hrabar" (Bulgaria)

ITA 2010 is supported by:

The International Journal "Information Theories and Applications"

The International Journal "Information Technologies and Knowledge"

The International Book Series "Information Science and Computing"

Deadlines

March 31, 2010: submission of final paper via submission system http://ita.ithea.org

April 15, 2010: notification of the paper acceptance

April 30, 2010: submission of registration forms and visa application forms to info@foibg.com

Papers

Papers from members of ITHEA International Scientific Society (ITHEA ISS) will be preferably included in the events of ITA 2010. Membership of ITHEA ISS is free and may be done by registration at the www.ithea.org.

Papers need to be up to 8 pages and formatted according the the sample sheet given at http://www.foibg.com/conf/paper_rules.doc.

Papers accepted by the Program Committee will be published in several separate numbers of the **International Book Series "Information Science and Computing" (IBS ISC)**.

The best ideas ***personally presented*** by their authors at the ITA 2010 will be offered a free publishing in English up to 16 A4 pages per idea in the International Journal "Information Theories and Applications"® (IJ ITA) as well as the International Journal "Information Technologies and Knowledge"® (IJ ITK).

Submission

Papers need to be submitted via submission system http://ita.ithea.org.

Official languages: English and Russian