ITHEA

International Journal INFORMATION THEORIES & APPLICATIONS

2008 Volume 15 Number 2

International Journal
INFORMATION THEORIES & APPLICATIONS
Volume 15 / 2008, Number 2

IJ ITA is official publisher of the scientific papers of the members of the ITHEA International Scientific Society, the Association of Developers and Users of Intellectualized Systems (ADUIS) and the Association for Development of the Information Society (ADIS)

IJ ITA welcomes scientific papers connected with any information theory or its application.
IJ ITA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITA as well as the **subscription fees** are given on *www.foibg.com/ijita*.
The camera-ready copy of the paper should be received by e-mail: *info@foibg.com*.
Responsibility for papers published in IJ ITA belongs to authors.

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

International Journal "INFORMATION THEORIES & APPLICATIONS" Vol.15, Number 2, 2008

Printed in Bulgaria

Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria,
in collaboration with the V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,
and the Institute of Mathematics and Informatics, BAS, Bulgaria.

Publisher: **Institute of Information Theories and Applications FOI ITHEA**
Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, www.foibg.com, e-mail: info@foibg.com

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

FORECASTING PROBLEMS FOR ECONOMIC MACROPARAMETERS

Oleksiy Voloshyn, Victoria Satyr

Abstract: The article deals with problems of forecasting of economic macroparameters on the basis of the principle of «subjective multideterminism», i.e. an expert account of maximal amount of interrelated «objective» and «subjective» causes. A description is given of the system of support of decision-making in forecasting the level of inflation and gross domestic product on the basis of the tree solution method.

Keywords: Tree solution method; index of inflation; gross domestic product.

Introduction

[Voloshyn, 2006] offers the technology of forecasting (by applying the system of high-quality forecasting on the basis of multiparameter dependences represented by a solution tree, [Voloshyn, 2005]) - realizing conception of «subjective multideterminism (in [Voloshyn, 2006] - «plural subjective determinism»). In the basis of this conception the effect is determined by a multitude of interdependent causes (objective and subjective, in particular, by activity of a subject, which, in its turn, is determined by his opportunities, will, desires, preferences, etc.). Such approach is by no means an original one - similar views were expressed, in particular, by the Nobel laureate V.V. Leontief («To the issue of pluralistic interpretation of history and problem of inter-disciplinary co-operation», Harvard, 1948), who made an attempt to develop his own methodology of «understanding of history», - «neither economic nor anthropological nor, say, a geographical analysis can, at the modern stage of development of the corresponding sciences, bring to a solely correct assertion». Certainly, as V.V. Leontief notes with irony, it will be much simpler when all sciences have merged into a certain unified field of knowledge. But as «the golden age» of science has not yet come, it is necessary to develop methods of « inter-disciplinary co-operation». V.V. Leontief suggests that several experts should be involved simultaneously in independent explanation (forecasting) of this or that phenomenon, whereupon «the project manager» must compare the results of their analyses to produce a certain maximally complete and objective research. In [Voloshyn, 2006] an approach like that is interpreted as «subjectivization of objectivity» and another step is suggested – «objectivization of subjectivity» (taking into account psychosomatic features of the experts). And the most important, if in 1948 it was possible only to declare such «pluralistic method of forecasting», the present-day development of computing engineering, mathematical methods (in particular, the approaches known under the common name of the «artificial intelligence»), sociology, psychology, etc. allows to a great extent to realize this approach ([Voloshyn, 1999, 2001, 2003, 2005]).

[Voloshyn, 2002, 2003] offer a description of instruments for creating applied systems of support of decision-making in various fields such as, for example, forecasting of economic parameters [Voloshyn, 1999, 2003], medical diagnostics [Voloshyn, 2005]. [Voloshyn, 2006], as an example, quotes the preliminary results of using the instruments created for forecasting the index of inflation in Ukraine. The said work deals with the problem of forecasting of macroeconomic parameters, in particular, the index of inflation and the gross domestic product on the basis of the instruments described in [Voloshyn, 2005, 2006].

Instruments for creating applied systems of forecasting

[Voloshyn, Pykhotnyk, 1999], [Voloshin, Panchenko, 2001], [Voloshyn, Panchenko, 2002], [Voloshin, Panchenko, 2003], [Voloshyn, 2005], [Voloshyn, Holovnia, 2005], most of which were presented at the KDS conferences, develop the conception of «high-quality forecasting on the basis of the multiparameter dependences represented by the solution tree» [Voloshyn, Panchenko, 2002]. It is considered that the effect is determined by a multitude of interdependent causes the degree of influence of which on the cause is determined «subjectively» (by expert measuring). The more parameters «forming» the effect, the better (for adequacy of the model), however, this makes the analysis of the model more complicated (there is the «curse of dimension», which must be overcome [Voloshyn, Panchenko, 2002], in particular, using also methods of artificial intelligence).

The construction of an applied system of support of decision-making is reduced to highlighting by experts of problems and subproblems (tops of the tree) and links between them (arcs of the tree). Experts determine the weights (probabilities) of transitions between tops. It is acceptable to get unclear expert estimations achieved by the Boolean variables described by the values of function of belonging (by vectors of the real numbers from 0 to 1). Every expert sets three estimations – optimistic, realistic and pessimistic, the scaling of which is effected taking into account the psychological type of the expert. The type is determined on the basis of psychological tests built into the system. Such psychological tests help to determine the coefficients of «veracity», «independence», «caution», etc.

The tree is built on the basis of collective estimations of experts using the method of pair comparisons. The construction of a resulting tree requires the use of the algebraic methods of treatment of expert information, and the Hemming metrics and measure of lacks of coincidence of grades of objects is used as distance between ranges. A resulting tree is determined as the Kemeny-Snella median or as a compromise [Voloshyn, 2005]. In the case of setting priorities in an unclear form the elements of matrix are set through the functions of belonging.

The algorithms of successive analysis of variants [Voloshyn, Panchenko, 2002] allowing to process trees with hundred tops are offered for determining optimum ways within the tree.

The solution tree is set by tables. Every table is a separate level of the tree, every line of the table is a separate top at this level. Every element of the line is the probability with which a transition is possible from this top to the top of a lower level. These probabilities are set by the functions of belonging, being vectors of the real numbers from 0 to 1 of any length. A table is filled through questioning of experts. The existent functions allow to add columns, lines, to set a dictionary (which allows to put verbal estimations of an expert in accordance with the probability, by way of setting certain levels), save tables in a file, read tables from a file.

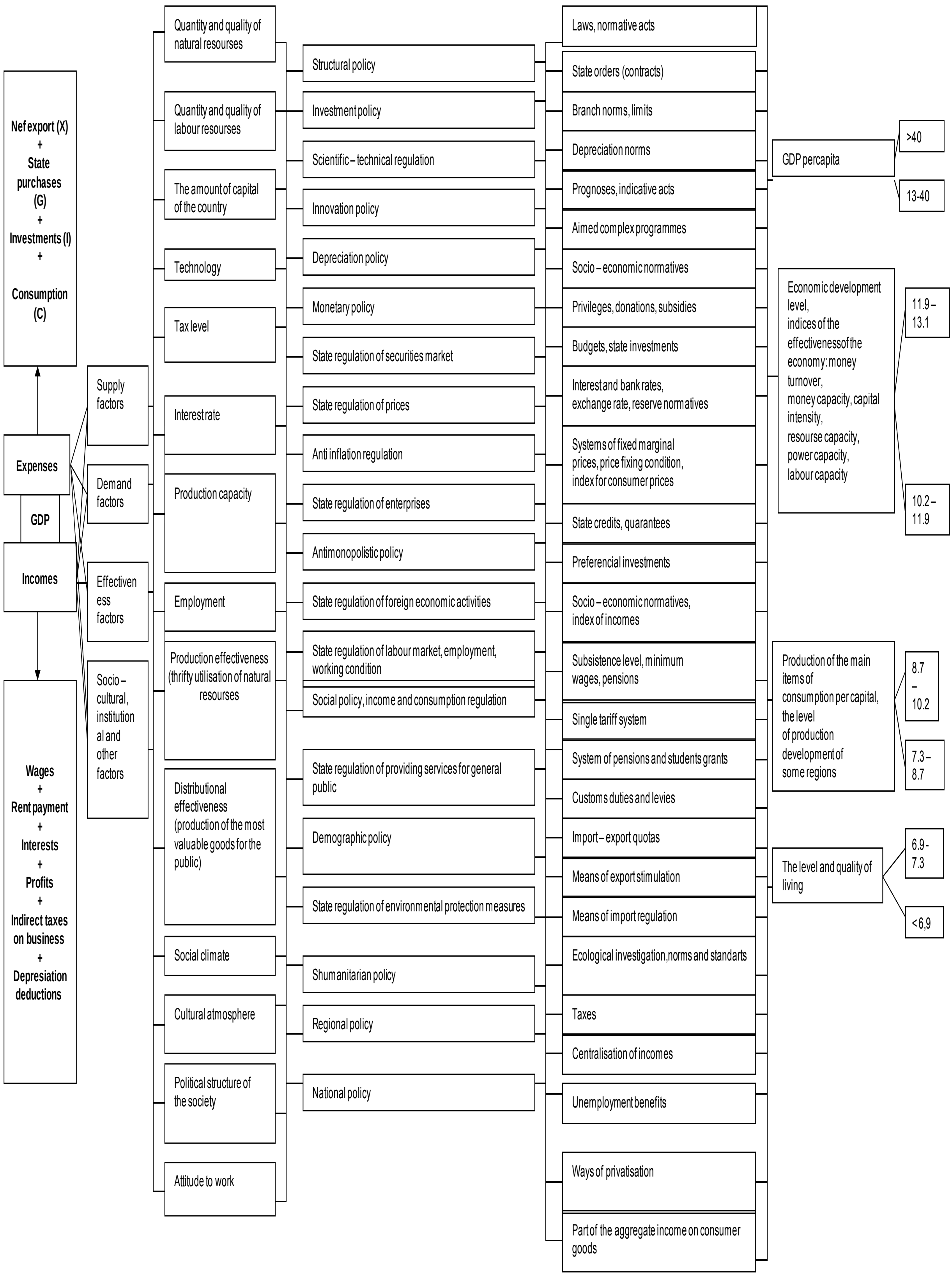
An expert way helps to set matrices that are the result of comparison of variants of tops which can be included into a tree. On the basis of analysis of the matrices tops are determined to be included in a tree as well as probabilities of transition into them from the tops of a higher level. If a solution tree is decomposed into a number of subtrees which have identical leaves, first the probabilities of these leaves in each of them are calculated, and then probabilities for the whole tree are determined.

Forecasting the index of inflation

The index of inflation is one of the basic macroeconomic parameters reflecting the main trends of economic development. No universal and perfect approaches to solving the problem of its forecasting exist today. The methods of quantitative forecasting (timing rows, regressive analysis, imitation modeling, etc.) based on the "continuation of the past" give bad results at forecasting of the "unstable" processes, characterized by the "violation of monotony", based on the saltatory changes not typical for development of the process in the past]. The problem lies in representing the future which cannot be interpreted as the ordinary continuation of the past, as the future can take totally new shapes in principle. Such forecasting ("high-quality forecasting") is based on the idea of direct use of a man's (expert's) knowledge. Thus, above all things, it is necessary to take into account the "unclearness" of the expert information which in its turn depends on his professional and psychological features (competence, independence, objectivity, realism, inclination to risk, etc). Therefore any forecasting of the index of inflation is effected using the instruments described in the previous chapter. A fragment of a solution tree for forecasting of the index of inflation is represented below.

In order to forecast the index of inflation the following basic subproblems are selected:

- The economic situation - the state of industry, the state of the agro-industrial complex, the state of the financial market, trade, etc.;
- The political situation – the share of the shadow economy, the investment, currency, antimonopoly policy, etc.;
- The socially-demographic situation - unemployment, socially-demographic pressure, growth rate of population, etc.;
- The financial situation - financial-budgetary, currency policy, government control of the equity market, government control of prices, etc.



Nef export (X) + State purchases (G) + Investments (I) + Consumption (C)
Expenses
GDP
Incomes
Wages + Rent payment + Interests + Profits + Indirect taxes on business + Depresiation deductions
Supply factors
Demand factors
Effectiveness factors
Socio – cultural, institutional and other factors
Quantity and quality of natural resourses
Quantity and quality of labour resourses
The amount of capital of the country
Technology
Tax level
Interest rate
Production capacity
Employment
Production effectiveness (thrifty utilisation of natural resourses
Distributional effectiveness (production of the most valuable goods for the public)
Social climate
Cultural atmosphere
Political structure of the society
Attitude to work
Structural policy
Investment policy
Scientific – technical regulation
Innovation policy
Depreciation policy
Monetary policy
State regulation of securities market
State regulation of prices
Anti inflation regulation
State regulation of enterprises
Antimonopolistic policy
State regulation of foreign economic activities
State regulation of labour market, employment, working condition
Social policy, income and consumption regulation
State regulation of providing services for general public
Demographic policy
State regulation of environmental protection measures
Shumanitarian policy
Regional policy
National policy
Laws, normative acts
State orders (contracts)
Branch norms, limits
Depreciation norms
Prognoses, indicative acts
Aimed complex programmes
Socio – economic normatives
Privileges, donations, subsidies
Budgets, state investments
Interest and bank rates, exchange rate, reserve normatives
Systems of fixed marginal prices, price fixing condition, index for consumer prices
State credits, guarantees
Preferencial investments
Socio – economic normatives, index of incomes
Subsistence level, minimum wages, pensions
Single tariff system
System of pensions and students grants
Customs duties and levies
Import – export quotas
Means of export stimulation
Means of import regulation
Ecological investigation, norms and standarts
Taxes
Centralisation of incomes
Unemployment benefits
Ways of privatisation
Part of the aggregate income on consumer goods
GDP percapita
>40
13-40
Economic development level, indices of the effectiveness of the economy: money turnover, money capacity, capital intensity, resourse capacity, power capacity, labour capacity
11.9 – 13.1
10.2 – 11.9
Production of the main items of consumption per capital, the level of production development of some regions
8.7 – 10.2
7.3 – 8.7
The level and quality of living
6.9 - 7.3
<6,9

The value of the index of inflation in Ukraine for 2005 that was forecast in June, 2005, made up 12,8%. In the budget of Ukraine this index amounted to 9,8%, the official statistics on the results of the year 2005 gives it as 10,5%, and international organizations assess it as 12,5-13,0%. As we see it, the high accuracy of our forecast is achieved owing to the «objective» reason – taking into account of large number of heterogeneous interrelated causes influencing the result. The second reason, in the authors' opinion, is «subjective», i.e. a «narrowly specialized» expert estimation, an expert frequently is not aware of what he forecasts in the final analysis. The forecast of the index of inflation value for 01.01.2008, received in April, 2007, made up 17,3%. It is interesting to note that the National bank of Ukraine forecasts inflation in Ukraine in the current year at the level of 7%, the government – 8%, President of Ukraine pronounced in March, 2007 the level of inflation of 11-12%. It only remains to wait for KDS-2008 and compare the forecasts!

Determination the gross domestic product index

As is generally known, the GDP index may be determined by three basic methods: by the created products (the production method); by the costs (the method of the final consumption); by the profits (the distributive method).

The first two methods are basically used in most countries. The choice of one method or another is determined by the availability of a reliable data base. Under the present conditions in Ukraine there is no «reliable data base», one can speak not so much about accuracy of statistical information, as about its «complete inaccuracy» (and, frequently, about its absence). One of the main factors «distorting» the statistics is the shadow sector of the economy. Its «integral» expert estimation (from 40% to 60%) is too inaccurate to be used in the GDP calculations. The way out is in applying the principle of «indirect calculation», using (after V.V. Leontief) the method of «inter-disciplinary co-operation» according to which long chains of events are naturally divided into groups of the directly linked events; each of them is studied and is explained separately, the type of explanation will change in transition from one group of links to another; purely «economic» interpretation may prove most suitable for one group of events, purely «political» or «social» for another.

For the purpose of estimation (the GDP forecasting) the maximal possible number of groups-factors («subproblems», «subtrees» in solution tree) were determined that influence directly or indirectly the GDP value. Thereupon, in subtrees the maximum possible amount of tops was singled out that may potentially determine the given group of factors. While making an expert decision on including each particular top into a subtree, we took into account, in particular, availability (accuracy, authenticity, possibility of receipt) of evaluation of this factor.

The following groups of factors and factors in a group were taken into account:

1. The real sector (expenses by the categories of expenses, expenses of households on the final consumption, the final consumption of the general state administration sector, individual final consumption, collective final consumption, gross fixed assets accumulation, change of stocks of the material and technical facilities, export of goods and services, import of goods and services, index of industrial production, employment, unemployment, average wages, consumer price index, price index of producers);

2. Budgetary-tax sector (operations of the general state administration sector, i.e. the summary balance, operations of the central organs of state administration, like the summary balance, financing due to outsourcings, due to internal sources, by bank institutions, non-banking institutions, income from privatization of state property, national and guaranteed by the state internal debt, government internal liabilities on the initial terms of settlement, short-term government internal liabilities (within 1 year), medium-term and long-term government internal liabilities, state securities, credits, debt obligations guaranteed by the government of Ukraine, loans of international financial organizations, official loans);

3. The financial sector (analytical accounts of the bank sector as of the end of the period, amount of money, internal credit, requirements to the government, requirements to other sectors, net external assets, analytical accounts of the central bank, as of the end of the period, monetary base, net requirements to the government, obligations before the government, requirements to the banks, requirements to other sectors, external assets, external liabilities, interest rates, the National Bank of Ukraine, the current bank rate, actual rate by the instruments, banks of Ukraine as to the credits allotted at the bank market in the national currency, rate by the attracted deposits in the national currency, rate by the given credits in the national currency).

4. The external sector (the balance of payments as of the end of the period, the account of current operations, the balance of goods and services, profits, current transfers, the account of operations with capital, the financial

account, investments from Ukraine (direct + portfolio + other), investments into Ukraine (direct + portfolio + other), errors and omissions, international reserves as of the end of the period, reserves in foreign currency, reserve position in the IMF, special rights of borrowing, gold, other reserve assets, foreign commodity trade, gross external debt as of the end of the period, the rate of exchange of a US dollar to an euro, annual average).

5. Social conditions for the population (average size of monthly pension, average size of monthly wages, state help to families with children, the labor-market, economic activity of the population, registered labor-market, unemployment, demographic situation, quantity of the population, natural movement of the population, migratory movement of the population, education, preschool establishments, vocational educational establishments, higher educational establishments).

The GDP estimates for the 1st quarter of 2007 amounted to 150 billion UAH, the official statistics (taking into account expert estimation of the shadow sector) is 10% lower. This fact can be interpreted as an error of the «integral official» expert estimation. It is interestingly to note, that the forecast GDP values for the year 2006 estimated in May-June, 2006, gave the same 10% divergence with the official estimation of the State Committee on Statistics. Undoubtedly, there is no question here of the accuracy of our forecast to the «true» value (merely for the lack of the latter. It would be interesting to conduct a research using the «true» data (but for one sector) for «training» the system i.e. for determining reliable values of interaction of the factors.

Conclusion

The name and contents of the KDS conference fully corresponds to the principle of the «subjective multideterminism» used in forecasting of economic macroparameters:

KNOWLEDGE (of experts in different branches of science and practice), - DIALOG (between them, with the help of experts on decision-making and artificial intelligence), - SOLUTION (of problems of humanity).

Bibliography

[Voloshyn, 1999] Voloshyn And., Pykhotnyk E. Expert system of forecasting of exchange rate of gryvnya. "Artificial intelligence", 1999, №2. -P.354-359 (Ukrainian).

[Voloshin, 2001] Voloshin O.F., Panchenko M.V. The Forecasting of Stable Processes by a solution tree Method using a Pairwise Comparison Method for Analysis of Expert Information. Materials of the international conference «KDS-2001», Vol. 1, Saint Petersburg, 2001. -P.50-53 (English).

[Voloshyn, 2002] Voloshyn O.F., Panchenko M.V. The Expert System of High-quality Evaluation on the Basis of Multiparameter Dependences. "Problems of mathematical machines and systems", 2002, №2. -P.83-89 (Ukrainian).

[Voloshin, 2003] Voloshin O.F., Panchenko M.V. The System of Quality Prediction on the Basis of and Fuzzy Data and Psychografphy of the Experts. International Journal "Information Theories & Applications", 2003, Vol.10, №3. -P.261-265.

[Voloshyn, 2005] Voloshyn O.F. On problems of decision-making in the socio-economic systems. Materials of the conference «KDS-2005», Vol. 1, Sofia, 2005.-P.205-212.

[Voloshyn, Holovnya, 2005] Voloshyn O.F., Holovnya V.M. The System of high-quality forecasting on the basis of unclear data and psychography of experts. Materials of the conference «KDS-2005», Volume 1, Sofia, 2005. -P.237-243.

[Voloshyn, 2006] Voloshyn O.F. Systems of support of decision-making as the personal intellectual toolkit of a person making a decision. Materials of the conference «KDS-2006», Sofia, 2006. -P.149-153.

Authors' Information

Oleksiy Fedorovych Voloshyn, Kyiv Taras Shevchenko National University, Faculty of Cybernetics, Professor. Kyiv, Ukraine. E-mail: ovoloshin@unicyb.kiev.ua

Victoria Valeriyivna Satyr, Kyiv Taras Shevchenko National University, Faculty of Cybernetics, master's degree. Kyiv, Ukraine. E-mail: Brili@ukr.net

MULTICRITERIA EVALUATION AND OPTIMIZATION OF HIERARCHICAL SYSTEMS

Albert Voronin

Abstract. It is shown that any multicriteria problem can be represented by a hierarchical system. Separate properties of the object are evaluated at the lower level of the system, using a criteria vector, and a composition mechanism is used to evaluate the object as a whole at the upper level. The paper proposes a method to solve complex multicriteria problems of evaluation and optimization. It is based on nested scalar convolutions of vector-valued criteria and allows simple structural and parametrical synthesis of multicriteria hierarchical systems.

Keywords: alternative choice, multicriteriality, hierarchical systems, composition of criteria, nested scalar convolutions

Introduction

In decision-making theory [1,2], there are two different approaches to evaluating objects (alternatives) subject to choice. One of them is to evaluate an object as a whole and to choose an alternative by comparing objects as gestalts (holistic images of objects without detailing their properties). The second approach is to detail and evaluate certain vectors of characteristics (properties) of objects and to make a decision after comparing these properties. The scheme of decision making can be represented by the formula [3]

$$\{\{\chi\},\Phi\} \to \chi^*,$$

where $\{\chi\}$ is a set of objects (alternatives); Φ is a choice function, i.e., a rule that establishes preferability in the set of alternatives; and χ^* are chosen alternatives (one or more).

The holistic approach implies choosing χ^* using the choice function Φ. The vector approach requires decomposing (expanding) the function Φ into a set (vector) of some choice functions φ. By decomposition of the choice function Φ is meant [1] its equivalent representation by a certain set of other choice functions φ whose composition is the initial choice function Φ.

Both approaches have their advantages and shortcomings. Choice after comparing objects may substantially differ from choice after comparing the vector characteristics of objects [2]. This is because information on vectors sometimes gives an insufficiently adequate description of objects, even with the most careful choice of characteristics of objects. Some portion of the information on objects is lost when the objects are described by a set of characteristics. Another portion of the information, which is not directly concerned with the objects being compared, is introduced into the model. The choice of an appropriate set of properties (characteristics) of an object is subjective to a certain extent. Moreover, there is an assumption [2] that human thinking is not has been evolved to adapt for a natural (from a formal standpoint) changeover from preferences on a set of objects to preferences on a set of their characteristics.

Nevertheless, modern decision theory is inclined to using the vector approach since it is objective and comprehensive and allows employing formalized methods. The concreteness and clearness of an approach are also taken into account since it is easier to collect indisputable facts and reach a consensus on a specific issue [4].

It is assumed that it is much easier for a decision-maker to reveal a preferable alternative for a certain property of an object. For example, in choosing the best design of an aircraft, it is easier to compare design A over an design B in comfort, or reliability, or weight-lifting ability, than to compare the whole designs A and B [3]. Selecting properties of alternatives is a decomposition that leads to a hierarchical structure of properties. The properties of the first hierarchical level can be subdivided into sets of next specific properties, etc. The division depth is determined by tending to reach the properties that are convenient to be compared to each other. Indeed, in the example with an aircraft, it is easier to judge its comfort rather than the aircraft as a whole; however, such a qualitative property is also not always convenient for comparison and has to be decomposed for convenient and objective comparison of properties. Therefore, the property of comfort, in turn, is subjected to hierarchical decomposition into levels: (a) cabin noisiness, (b) floor vibrations, (c) seat spacing, etc. These characteristics can be evaluated and are objective.

Properties for which objective numerical characteristics exist are called criteria. More strictly, quantitative characteristics of properties of an object, whose numerical values are a quality measure of an object of evaluation with respect to the given property, are called criteria. Deriving a set of criteria is the final result of hierarchical decomposition. The number of levels depends on the decomposition depth required. Complexity is that decomposition depth can be different for different initial properties, and heterogeneous sets of criteria should be normalized at each hierarchy level.

Though attractive, the approach comparing individual properties involves the serious problem of inverse passage to the required comparison alternatives in general. This problem assumes a composition of criteria based on hierarchy levels, which is difficult, especially for a significant decomposition depth of properties. In an elementary and most popular case (two-level hierarchy), the composition problem can be solved traditionally, by deriving a single scalar convolution of criteria. However, other approaches are required for three (and more)-level hierarchy.

The aforesaid makes it reasonable to state that any multicriteria problem can be represented as a hierarchical system, where at the lower level, the object is evaluated in separate characteristics using a vector of criteria, and at the upper level, a composition is used to evaluate the object as a whole. The key problem is composition of criteria at hierarchy levels.

Analysis of the problem state

The case where a multicriteria problem can be represented by a two-level hierarchical system is developed most thoroughly in decision theory. The composition problem is usually solved either using a main criterion (criteria constraints) method or by using a single scalar convolution of a vector –valued criterion [3]. The latter method is used more often, the numerical value of convolution being the quality evaluation of the given object (alternative) as a whole.

It is convenient to use scalar convolution in a traditional form when the number of partial criteria is not too great (usually, $s \le 10$). Then each criterion plays a self-depended role and all of them are comparable in importance. However, there exist complex multicriteria problems with a large number (say, several tens) of partial criteria. Then the value of each criterion separately has a weak effect on the solution of the multicriteria problem. It is expedient to group them into headings (groups, clusters) where scalar convolutions are considered as new, higher-weight criteria. These aggregated criteria, in turn, are subjected to scalar convolution and then are compared with higher weights during the solution of the multicriteria problem. Thus, as the dimension of the criteria space increases, the initially two-level hierarchical system of criteria is transformed into multilevel one and requires a mechanism to compose criteria into hierarchy levels.

As an example, let us consider alternate evaluation of research projects in biological studies in space [5]. To evaluate the efficiency of such projects, 28 partial criteria are used. Consultations with experts has allowed grouping these criteria into four headings (groups): (i) general criteria, (ii) scientific development criteria, (iii) economic criteria and (iv) social criteria (see Table 1).

Table 1

Criterion No.	Project quality criteria	Points (10-point scale)
	General Criteria	
1	Compliance of the project with the Space Program of Ukraine	10.00
2	Integration of the project into international programs of biological investigations in space	9.00
3	Probability that the approach will lead to the desired results	7.50
4	Completeness of feasibility of the project under given conditions of space experiment	8.30
	Scientific Development Criteria	
5	Compliance of the job structure and investigation methods with project tasks	9.75
6	Furtherance of gathering knowledge on the influence of space flight factors on fundamental physiological processes	8.25

7	Furtherance of gathering knowledge on the adaptation of biological objects for space flight conditions	8.25
8	Novelty of investigations	8.00
9	Originality and innovation of the purposes formulated	8.75
10	Influence of investigation on scientific concepts and methods in space biology and medicine	9.75
11	Probability that investigation will allow new break-through projects	8.50
12	Refuting the paradigms available	5.30
13	Share of worldwide investigations in the project	8.00
14	Enhancing the prestige of Ukraine in the world	8.50
15	International support of the project	9.70
16	Coverage in the scientific literature	9.70
17	Using the results in academic activity	7.70
18	Popularization and propagation of knowledge	7.75
	Economic Criteria	
19	Probability of introducing the technologies into Ukrainian economy	9.33
20	Adequate number of specialists to bring the research to practical implementation	10.00
21	Attracting investments	9.33
22	Reducing production expenses	9.33
23	Increasing sales	8.67
24	Adequacy of financing the tasks planned	9.33
	Social Criteria	
25	Increasing the number of worksites	10.00
26	Increasing the level of staff qualification	8.00
27	Promoting development of small and medium-scale business	7.67
28	Influence on the activity of social and youth organizations	10.00

The right-hand column of the table contains the results of expert evaluation of the Biosorbent space research project, which has been included in the program of onboard experiments at the International Space Station. These data are the lower level of a three-level hierarchical system of criteria for evaluating the project as a whole.

Totsenko considered in [6] the general case of developing a decision-making support technology when an alternative should be chosen from a set of inhomogeneous alternatives for which it is impossible to formulate a unified set of quantitative evaluation criteria. In this case, the problem can be solved by methods based on hierarchical objective evaluation of alternatives without criteria analysis. Given qualitative properties, the problem of composition can be solved using binary relations, for example, by the hierarchy analysis [7]. But the problem is facilitated substantially if quantitative (or reducible to them) criteria that permit operations in a normalized criteria space are employed to evaluate alternatives. The theory of multicriteria evaluation and optimization is applicable to such problems. The present paper addresses such a class of problems.

Formulation of the problem

The state of a hierarchical system for a given alternative is defined by the following parameters:

$I = \{1,2,...,n\}$ is a set of elementary subsystems evaluated using lower-level hierarchy criteria;

$\{y_i\}_{i\in I}$, $y = \{y_1,..., y_n\}$ are the estimates of elementary subsystems based on scalar criteria and a vector-valued criterion of the lower hierarchy level. The efficiency of each highest level depends on the estimates according to the lowest-level hierarchy criteria.

The additional conditions that define the hierarchical structure are as follows:

$J = \{1,2,..., m\}$ is the set of hierarchical levels;

$\{I_j\}_{j\in J}$ is the distribution of subsystems into levels, $I_j = \{1,2,..., n_j\}$;

$\{\lambda_j\}_{j\in J}$ are priority vectors.

It is required to find an analytical estimate $\varphi*$ and a qualitative efficiency evaluation of the hierarchical structure and to choose the best alternative from available ones.

Solution technique

To analytically evaluate the hierarchical structures for efficiency, we propose to apply the method of nested scalar convolutions [8]. A composition is carried out by a nested doll ("matreshka") principle: scalar convolutions of weighed components of vector criteria of the lowest level are components of vector criteria of the highest level. The scalar convolution of criteria obtained at the uppermost level automatically becomes the expression for the efficiency evaluation for the whole hierarchical system.

The algorithm of nested scalar convolutions can be represented as sequential weighed scalar convolutions of vector criteria at each hierarchy level in view of priority vectors based on the compromise (trade-off) scheme selected

$$\{(\varphi^{(j-1)},\lambda^{(j-1)})\to\varphi^{(j)}\}_{j\in J}\,,\ \varphi^{(1)}\equiv y\,,$$

and efficiency evaluation of the whole hierarchical system can be expressed by determining scalar convolution of the upper hierarchy level:

$$\varphi*=\varphi^{(m)}\,.$$

In choosing solutions, the number of alternatives is $n_a\geq 1$. Each alternative is characterized by a hierarchical structure. For $n_a=1$, the problem posed is transformed into the evaluation of the given hierarchical structure. If $n_a>1$, then each structure is evaluated as a given one and the alternative whose hierarchical structure has been evaluated best, is chosen. Therefore. in case of discrete multicriteria optimization, the base problem is to evaluate a given hierarchical structure. However, this method can only be used if the number of alternatives n_a is relatively small, when simple enumeration does not involve significant computational difficulties. For large sets of alternatives, other optimization methods should be applied, for example those stated in [9].

Compromise scheme

As a base trade-off scheme for the method of nested scalar convolutions, we propose to use the nonlinear scheme described in [10]. It was established that without loss of generality, a premise for its application is that all of the partial criteria are subject to minimization and are bounded:

$$y_i\leq A_i,\,A=\{A_i\}_{i=1}^n,\ i\in[1,n],$$

where *A* is the vector of constraints.
The scalar convolution

$$\varphi(\lambda,y)=\sum_{i=1}^{n}\lambda_i[A_i-y_i]^{-1}$$

or

$$\varphi(\lambda,y_0)=\sum_{i=1}^{n}\lambda_i[1-y_{0i}]^{-1},$$

if quantitative criteria are normalized by the formula $y_0=y/A$, is a simple informative model of utility function of the decision maker at the lower level of hierarchy for criteria being minimized according to the concept of nonlinear trade-off scheme.

Qualitative (but reducible to quantitative) criteria are usually determined by experts using scale points. A questionnaire with partial criteria is given to experts. Criteria are associated with a continuous scale divided, for example, into ten intervals. Zero on the scale is indicative of no weight, 10 corresponds to the maximum weight. An expert should estimate the relative influence of each partial criterion on the general estimate under given conditions and to associate it with the corresponding point on the scale, characterized by a number *f*. It is admissible to select points between numbers or to assign some criteria to one point on the scale.

An analysis of decision-making processes has shown that evaluating objects on a 10-points scale, experts are guided by gradations of a so-called fundamental scale represented in a general form in Table 2 and described in [7], where qualitative gradations of properties of objects are associated with the corresponding quantitative estimates *f*. It is possible to say that in terms of fuzzy-set theory [11], the fundamental scale appears as a universal membership function for a passage from a number to the corresponding qualitative gradation and back. A passage from a linguistic variable (satisfactory quality, excellent quality, etc) to corresponding quantitative estimates *f* according to point scale (5.5, 7.0), i.e. passage from fuzzy qualitative gradations to numbers and back, is carried out.

Table 2

Quality category	**Ranges of fundamental scale for f**	**Ranges of normalized inverse fundamental scale for y_0, φ_0**
Unacceptable	0-3	1.0-0.7
Low	3-5	0.7-0.5
Satisfactory	5-6	0.5-0.4
Good	6-8	0.4-0.2
High	8-10	0.2-0.0

Estimates *f* are determined according to a direct 10-point fundamental scale for the criteria being maximized. The technique applied in the paper for multicriteria evaluation according to a nonlinear trade-off scheme is developed for normalized minimized criteria y_0 whose estimates are obtained from *f* by the formula [12]

$$y_0=1- 0,1 \cdot f, \; y_0\in[0;1].$$

This is reflected in Table 2 by an inverse normalized scale. This scale is used to measure normalized scalar convolutions of criteria φ_0 as well.

Allowance for priorities

The simplex

$$\Gamma_\lambda = \{\lambda \mid \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = 1\} \quad (1)$$

is the domain of definition of priority coefficients $\lambda \in \Gamma_\lambda$, where λ_i =const are formal parameters with double physical meaning. On the one hand, these are priority coefficients that express the preference of a decision-maker according to certain criteria. On the other hand, these are coefficients of an informative regression model constructed based on the concept of nonlinear trade-off scheme. The coefficients *λ* can be determined at each hierarchy level through optimization on a simplex using the dual approach described in [10] or by the formula

$$\lambda_{ik}^{(j-1)} = \frac{f_{ik}}{\sum_{i=1}^{n_k^{(j)}} f_{ik}}, k \in I_j,$$

where $\lambda_{ik}^{(j-1)}$ is the *i*-th component of the priority vector at the $(j-1)$-th hierarchy level in evaluating the efficiency of the *k*-th subsystem on the *j*-th level; f_{ik} is the significance parameter of the *i*-th subsystem of the $(j-1)$-th level for the *k*-th subsystem of the *j*-th level (determined by experts using a 10-point scale); and $n_k^{(j)}$ is the number of subsystems of the $(j-1)$-th level that support the *k*-th subsystem of the *j*-th level.

In the most simple and popular case, a multicriteria problem without priorities is formulated and solved, where the decision-maker assumes that all of the significance parameters are identical for all of the subsystems. In this case, an elementary scalar convolution under a nonlinear trade-off scheme in a unified form [10] is used.

For a composition of criteria on hierarchy levels, it is expedient to calculate all of the scalar convolutions from top to bottom based on the concept of a nonlinear trade-off scheme. In this case, the efficiency of the *k*-th subsystem at the *j*-th hierarchy level as a normalized nested scalar convolution, in view of priority coefficients, can be evaluated by the formula

$$\varphi_{0k}^{(j)} = N_k^{(j)} \sum_{i=1}^{n_k^{(j)}} \lambda_{ik}^{(j-1)} [1 - \varphi_{0ik}^{(j-1)}]^{-1} \,,\; k \in I_j \,,\; \varphi_0^{(1)} \equiv y_0 \,, \tag{2}$$

where $\varphi_{0ik}^{(j-1)}$ is the estimate of the *i*-th component of the normalized vector-valued criterion at the $(j-1)$-th hierarchy level for evaluating the *k*-th subsystem of the *j*-th level; and $N_k^{(j)}$ is a normalizing factor.

Normalizing conditions

Normalization of nested scalar convolutions at each hierarchy level is of importance for the theory presented here. In [5,8], the possibility of calculating normalizing conditions based on the principle of joint liability of criteria is considered. The value of the greatest (worst) criterion is separated out in the set of normalized criteria. It is agreed that if this criterion has attained the worst value, the remaining normalized criteria are assigned the possibility of attaining the same values, which constitute components of the normalizing vector. Such an approach is simple but works only if the criteria are really "equivalent".

It is logical that if estimates with respect to all of the partial relative criteria $\varphi_{0ik}^{(j-1)}, i \in [1, n_k^{(j)}]$ are identical and equal to $\varphi_{0ik}^{(j-1)} \equiv \varphi_{0k}$, then their normalized scalar convolution in formula (2) should express the same analytical and qualitative estimate according to the inverted normalized fundamental scale:

$$\varphi_{0k} = \frac{N_k^{(j)}}{(1 - \varphi_{0k})} \sum_{i=1}^{n_k^{(j)}} \lambda_{ik}^{(j-1)}.$$

Since by normalizing conditions (1) $\sum_{i=1}^{n_k^{(j)}} \lambda_{ik}^{(j-1)} = 1,$ the expression for the normalizing factor becomes

$$N_k^{(j)} = \varphi_{0k} \left(1 - \varphi_{0k}\right). \tag{3}$$

Let us use formula (3) to perform calibration calculations of the normalizing factor $N_k^{(j)}(\varphi_{0ik}^{(j-1)})$ for the estimates $\varphi_{0ik}^{(j-1)}$, $i \in [1, n_k^{(j)}]$. Let us compose the measure of the total quadratic error that occurs since the unknown factor $N_k^{(j)}$ is used rather than exact values of the normalizing factor at calibration points $N_k^{(j)}(\varphi_{0ik}^{(j-1)})$:

$$M = \sum_{i=1}^{n_k^{(j)}} [N_k^{(j)} - N_k^{(j)}(\varphi_{0ik}^{(j-1)})]^2.$$

Using the necessary extremum condition for the function

$$\frac{\partial M}{\partial N_k^{(j)}} = 0,$$

we get the normalizing factor

$$N_k^{(j)} = \frac{1}{n_k^{(j)}} \sum_{i=1}^{n_k^{(j)}} N_k^{(j)}(\varphi_{0ik}^{(j-1)})$$

and with allowance for (3), we get

$$N_k^{(j)} = \frac{1}{n_k^{(j)}} \sum_{i=1}^{n_k^{(j)}} \varphi_{0ik}^{(j-1)} (1 - \varphi_{0ik}^{(j-1)}).$$

Conclusions

The recurrence formula (2) allows us to evaluate qualitatively and quantitatively the scalar convolutions of criteria, normalized using the inverted fundamental scale, with respect to all of the hierarchy levels up to the upper one: $\varphi_0^*=\varphi_0^{(m)}$.

The solution of the multicriteria evaluation problem for the example represented by Table 1 has resulted in the following. From the scalar convolution of partial criteria of the lower level of hierarchy, going under the heading General Criteria, we obtained an aggregated criterion of the second hierarchy level $\varphi_{01}^{(2)}=0{,}121.$ Similarly, we obtained the value of the aggregated criterion $\varphi_{02}^{(2)}=0{,}143$ for Scientific Development Criteria, $\varphi_{03}^{(2)}=0{,}022$ for Economic Criteria, and $\varphi_{04}^{(2)}=0{,}100$ for Social Criteria.

The scalar convolution of the indicated aggregated criteria of the second hierarchy level has allowed obtaining the normalized estimate for the whole Biosorbent space research project as an aggregated criterion of the third hierarchy level $\varphi_0{*}=\varphi_0^{(3)}=0{,}094.$ Comparison of the analytical estimates to the converted normalized fundamental scale (Table 2) allows concluding that all of the aggregated criteria are within the limits of the High Quality gradation. Note that in calculations of this example, all of the criteria were assumed to be of identical significance, i.e., a multicriteria problem without priorities was solved.

Evaluation of a given alternative and selection of the best one pertain to the class of problems of structural synthesis. Problems of parametrical synthesis solved by the method of nested scalar convolutions are described in [12]. Thus, any multicriteria problem can be represented as a hierarchical system; at its lower level partial properties of the object are evaluated using a vector of criteria, and at the upper level, the object is evaluated as a whole by means of composition.

Bibliography

1. I.M. Makarov, T.M. Vinogradskaya, A.A. Rubchinskij, and V.B. Sokolov, The Theory of Choice and Decision Making [in Russian], Nauka, Moscow (1982).
2. D.B. Yudin, Computational Methods in Decision Theory [in Russian], Nauka, Moscow (1989).
3. V.A. Gubanov, V.V. Zakharov, and A.N. Kovalenko, Introduction to Systems Analysis [in Russian], Izd. LGU, Leningrad (1988).
4. O.I. Larichev, "Problems of collective decision-making in small groups", in: Nonlinear Dynamics and Control (To the 70-th Anniversary of Acad. S.V. Yemel'yanov [in Russian], Editorial URSS, Moscow (1999), pp. 91-103.
5. A.N. Voronin, "Systems analysis and multicriteria evaluation of space projects by expert methods", Probl. Upravl. Inform., No. 1, 121-135 (2004).
6. V.G. Totsenko, Decision Support Systems and Methods [in Russian], Naukova Dumka, Kiev (2002).
7. T.L. Saaty, Multicriteria Decision Making: The Analytical Hierarchy Process, McGraw-Hill, New York (1990).
8. A.N. Voronin, "Nested scalar convolutions of a vector-valued criterion", Probl. Upravl. Inform., No. 5, 10-21 (2003).
9. A.N. Voronin, "Adaptive approximation models in optimization", Cybern. Syst. Analysis, **30**, No. 5, 700-708 (1994).
10. A.N. Voronin, Yu. K. Ziatdinov, and A.I. Kozlov, Vector Optimization of Dynamic Systems [in Russian], Tekchnika, Kiev, (1999).
11. L.A. Zadeh, The Concept of a Linguistic Variable and its Application to Approximate Reasoning [in Russian], Mir, Moscow, (1976).
12. A.N. Voronin, "Vector optimization of hierarchical structures", Probl. Upravl. Inform., No. 6, 26-34 (2004).

Authors' Information

Albert N. Voronin – National aviation university, faculty of computer information technologies, Dr.Sci.Eng., professor; 03058, Kiev-58, Kosmonavt Komarov avenue, 1, Ukraine

LIMIT BEHAVIOUR OF DYNAMIC RULE-BASED SYSTEMS

Gennady Osipov

Abstract: The paper suggests a classification of dynamic rule-based systems. For each class of systems, limit behavior is studied. Systems with stabilizing limit states or stabilizing limit trajectories are identified, and such states and trajectories are found. The structure of the set of limit states and trajectories is investigated.

Keywords: Dynamic rule-based systems, set of attainable states, limit trajectories.

ACM Classification Keywords: I.2.8 Problem Solving, Control Methods, and Search

Introduction

Dynamic intelligent systems and dynamic knowledge bases are typically understood as a result of integrating expert systems with simulation systems and automatically updated knowledge bases [1].

Although research in this area has a rather long history, there are some issues that still remain unaddressed — those of global behavior of dynamic intelligent systems, their attainable sets, stability, and other issues that usually come up when studying dynamic systems [2].

In this paper, we focus on the kind of systems that are dynamic systems whose states, behavioral laws and other dynamic system “attributes” are described in a special way.

This special way consists in using intelligent system techniques for describing both states and behavioral laws for such systems (by intelligent system techniques we mean methods for knowledge representation, modeling of reasoning and behavior modeling that are common in artificial intelligence). The said does not mean that functions and variables defined in any other way cannot be used as components of such systems; what is more even, it is supposed that the systems in question allow for integration with various models, such as differential equation systems, finite automata, and others.

In a general case, attainability of knowledge-based systems is determined by their knowledge bases and control strategies [3, 4]. In the case when a set of rules is used as knowledge representation in a system [5], attainability is entirely determined by the structure of the set of rules, by the general principles of rule organization, and the control strategy being used [6].

If by a rule base architecture we mean ‘structure of the rule set + rule structure + rule application logic’ then, consequently, attainability of rule-based systems is entirely determined by the rule base architecture.

Let us remind the basic definitions following [4].

1 Rule-Based Systems

A rule [6] is said to be a triple of sets:

$\Pi = \langle C, A, D \rangle$, where:

C is the applicability condition for the rule;

A is the set of facts to be added by the rule Π;

D is the set of facts to be removed by the rule Π.

C, A and D are sets of formulas of a language L, e.g. a multi-sorted first-order predicate calculus language, whose alphabet contains variables of sort t that take values from a linearly-ordered discrete set T. $A \cap D = \varnothing$ for every rule.

The word “fact” is used here as a synonym for the expression “closed atomic formula of a first-order predicate calculus language”.

Formulas from C, A and D are turned into facts by some substitutions that will be described below.

Every rule will be assigned to one of two classes τ or θ and denoted as a τ-rule or θ-rule, respectively.

With each τ-rule, we associate either an action which is performed by an actuator in the environment or a procedure that computes and assigns to a variable the values of certain database attributes based on values taken by other attributes in the current state.

No actions are associated with θ-rules, as the latter do not affect the real world and merely update our knowledge of it.

It should be stressed that in conditions C of θ-rules first-order language formulas are such that the value of t ($t \in T$) in the formulas from the condition is the same as the value of t in the formulas from the sets A and D. This means that the result of a θ-rule execution changes the state in which its condition is satisfied.

As far as τ-rules are concerned, if $(\forall t)$ $(n \leq t \leq m)$ $(\exists x)$ $F_C(t,x)$ is a formula from condition C and $(\forall t)$ $(p \leq t \leq q)$ $(\exists y)$ $P_A(t,y)$ is a formula from the list A of formulas to be added (where n, p and m, q are the discrete start and end time points, respectively, for the "validity" term of facts $F_C(t,x)$ and $P_A(t,y)$), then the integer $v = p - n$, which is the time lag between the start point of the fact $P_A(t,y)$ being added and the start point of the period when the condition $F_C(t,x)$ is true, is a characteristic of each rule and is associated with it.

Things are the same with the sets of formulas to be removed.

Let us now look at the basic computational process in rule-based systems.

For this purpose, we need the following concepts to be introduced [3]: database and strategy of control over the system's rules.

1.1 Database

Database is a collection of finite relations, or tables (e.g. like those in relational databases), the number of which equals the number of different predicate symbols in the rules. Table columns correspond to the sorts of individual variables in atomic formulas. Interpretation of language L in the database is taken to be defined in a standard way.

One can therefore talk of satisfiability or non-satisfiability of rules' conditions.

1.2 Control Strategy

Control strategy picks up a rule from the set of rules, checks if its condition is satisfied in the current state of the working memory and, if so, applies the rule, i.e. performs the actions as prescribed by the rule; otherwise, it picks up the next rule and carries out the same manipulations on it.

For the sake of definiteness, we assume that the set of rules is ordered, e.g. in a lexicographic way.

Then the control strategy looks as follows:

1. Pick up the next rule Π_i from the set of rules.

2. Check whether condition C_i is true in the current state of the working memory.

3. If C_i is true, then substitute all free variables in formulas from C_i, A_i and D_i by the corresponding values from the database. Otherwise go to 1.

4. Apply the rule, i.e. write down to the working memory the values that make true the formulas from A_i and remove from the working memory the values that make true the formulas from D_i.

5. Go to 1.

The condition for the completion of the process is either stabilization of the working memory or exhaustion of the set of applicable rules.

Typically, the choice of rule depends on the task or domain specifics; the general principle consists in that the rule's condition should hold. If there is more than one such rule in a current state then the so-called conflict set resolution strategies are applied. With the latter not being the subject of this paper, we take the control strategy to be such that the choice of rule will only affect the computational complexity and not the result of the process. To put it differently, in what follows we are not going to be concerned with rule applicability, and we will get back to this later.

2 Dynamic Rule-Based Systems

Let X be a set of facts, $\chi \in 2^X$, $\Pi \in (\tau \cup \theta)$. Let $K(\chi,\Pi)$ denote the control strategy we described above, and we assume

$$K(\chi,\Pi^\theta) = \Phi(\chi), \text{ where } \Pi^\theta \in \theta,$$
$$K(\chi,\Pi^\tau) = \Psi(\chi), \text{ where } \Pi^\tau \in \tau.$$

$\Phi(\chi)$ will be referred to as a closure function, $\Psi(\chi)$ as a transition function.

Then

$$H = <X, T, \Phi, \Psi> \quad (1)$$

will be referred to as a dynamic rule-based system.

The fixed point of equation

$$\Phi(\chi) = \chi$$

will be referred to as a state of the system (1), and the fixed point of equation

$$\Psi(\Phi(\chi)) = \chi \text{ (if such exists) with } t\to\infty,$$

will be referred to as the limit state of the system (1).

3 Classification of Rule-Based Dynamic Systems

As the basis for classification we will take the form of system's rules and certain correlations on the sets of rule components.

First, we identify classes of systems which differ in the form of rules.

In system H1 the rules are of the form:

$$\Pi_1=<C, \{P(t,y)\}, \varnothing>$$

(here P(t, y) is a fact to be added).

In system H2 the rules are of the form:

$$\Pi_1=<C, \{P(t,y)\}, \{\Phi(t,z)\}>$$

(here P(t, y) is a fact to be added, Φ(t, z) is a fact to be removed).

In system H3 the rules are of the form:

$$\Pi_1=<C, P(t,y), F(t,z)>$$

(here P(t, y) is a set of facts to be added, F(t,z) is a set of facts to be removed).

Let us now identify classes of systems, based on some correlations on the sets of rule components. Let S_0 be the initial state.

Then system H21 is a system H2, such that:

$(\cup\{P\}) \cap (\cup\{\Phi\}) = \varnothing$ (where $(\cup\{P\})$ and $(\cup\{\Phi\})$ is the union of facts being added and removed, respectively, over all of the rules of system H2);

system H22 is a system H2, such that $S0 \cap (\cup\{\Phi\}) = \varnothing$;

system H23 is a system H2, such that $(\cup\{P\}) \cap (\cup\{\Phi\}) \neq \varnothing$ and $S0 \cap (\cup\{\Phi\}) \neq \varnothing$;

system H31 is a system H3, such that $(\cup P) \cap (\cup F) = \varnothing$;

system H32 is a system H3, such that $S0 \cap (\cup F) = \varnothing$;

system H33 is a system H3, such that $(\cup P) \cap (\cup F) \neq \varnothing$ and $S0 \cap (\cup F) \neq \varnothing$ (here $(\cup P)$ and $(\cup F)$ stand for the union of the sets of facts being added and removed, respectively, over the entire set of the rules of system H3).

4 Limit States of Dynamic Rule-Based Systems

Let us give a few rather simple statements without proof:

Statement 1. The limit state of system H1 equals $S0 \cup (\cup\{P\})$.

Statement 2. The limit state of system H21 equals $(S0 / (\cup\{\Phi\})) \cup (\cup\{P\})$.

Statement 3. The limit state of system H31 equals $(S0 / (\cup F)) \cup (\cup P)$.

Statement 4. In systems H22, H23, H32, H33 stabilization of states never occurs, as a matter of fact, but every state of every one of these systems lies in the set $S0 \cup (\cup\{P\})$ — for systems H22 and H23 or the set $S0 \cup (\cup P)$ — for systems H32 and H33.

Statement 5. The trajectories of systems H22, H23, H32 or H33, with t large enough, look as shown by diagram 1, where for H22 and H32 for all i: $Si0 \subseteq Si1$.

It is appropriate to call such trajectories *limit trajectories*.

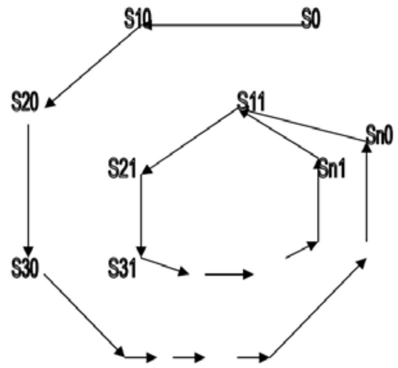


Diagram 1

5. Structure of the Set of Limit States

Let us represent the structure of the set of limit states as set inclusion diagrams. The arrow pointing from a smaller set to a larger one plays the role of the inclusion relation. We also assume that all facts to be added and removed of all the systems under consideration belong to set X. Then it is evident enough that the following diagram 2 holds for systems H1, H21 and H22:

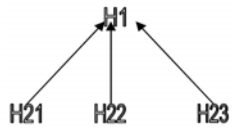


Diagram 2

Let now systems H31, H32 and H33 be such that to each rule of system H31 a rule of system H32 is related in such a way that for each rule Π(H31) of system H31 there is such a rule Π(H32) in system H32 that C(H31) = C(H32), P(H31) = P(H32), and the inverse holds true (where C and P are the applicability conditions and the sets of facts to be added of systems H31 and H32, respectively). Then, if system H33 is such that for each rule Π(H33) of system H33 there is a rule Π(H31) in system H31 and there is a rule Π(H32) in system H32 such that C(H31) = C(H32) = C(H33), P(H31) = P(H32) = P(H33) and $F(H31) \subseteq F(H33)$ and $F(H32) \subseteq F(H33)$, then the following diagram 3 holds:

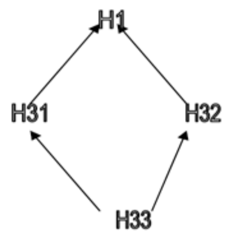


Diagram 3

In the last two diagrams, the inclusions of H22, H23, H32 and H33 in H1 have a slightly different meaning from others: they mean inclusion in H1 of every state of the limit trajectory.

Now let us come back to the postponed question of rule applicability. The situation is as follows: taking rule applicability into account may lead to some rules proving inapplicable on a certain step. It can be shown that if one sticks to the control strategy described in Section 1.2 then diagram 2 will remain the same. Diagram 3 will change its appearance to that of diagram 4:

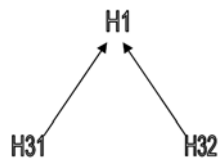


Diagram 4

Conclusion

In the paper, classes of dynamic systems have been identified from the point of view of their architecture. It has been shown that it is precisely the architectural specifics of such systems that determine their behaviour. The classes of systems with stabilizing limit states have been specified, and these states have been found. For systems with no stable limit states, the limit trajectories have been found (in case of a finite rule set).

The structures of limit states and trajectories have been established based on the criterion of attainable set inclusion.

Bibliography

[1] V.L.Stefanuk. Dynamic Expert Systems. The International Journal of Systems & Cybernetics. V. 29, Issue 5, pp. 702-709, MCB University Press, 2000.

[2] M.Mesarovich, Ya.Takakhara. General Theory of Systems: Mathematical Foundations. Moscow: Mir, 1978 (in Russian).

[3] G.S.Osipov. Dynamics in Integrated Knowledge-Based Systems. Proc. of the 1998 IEEE International Symposium on Intelligent Control (September 14-17,1998). Gaithersburg, Maryland, USA, 1998, pp.199-2003.

[4] T.G.Lebedeva, G.S.Osipov. Architecture and Controllability of Knowledge-Based Discrete Dynamical Systems. Journal of Computer and System Sciences International. NY, Vol. 39, No.5, 2000.

[5] N.Nilson. Principles of Artificial Intelligence. Moscow: Radio i svyaz', 1985.

[6] A.N.Vinogradov, L..Yu.Zhilyakova, and G.S.Osipov. Dynamic Intelligent Systems: II. Computer Simulation of Task-Oriented Behavior. Journal of Computer and System Sciences International. NY, Vol. 42, No. 1, 2003

Author's Information

Gennady Osipov – Institute of Systems Analysis, RAS, 9 Prospekt 60-letiya Oktyabrya, Moscow 117312; e-mail: gos@isa.ru

CASE-BASED REASONING METHOD FOR REAL-TIME EXPERT DIAGNOSTICS SYSTEMS

Alexander Eremeev, Pavel Varshavskiy

Abstract: The method of case-based reasoning for a solution of problems of real-time diagnostics and forecasting in intelligent decision support systems (IDSS) is considered. Special attention is drawn to case library structure for real-time IDSS (RT IDSS) and algorithm of k-nearest neighbors type. This work was supported by RFBR.

Keywords: Intelligent decision support systems, expert diagnostics systems, analogous reasoning, case-based reasoning.

ACM Classification Keywords: H.4.2 [Information systems applications]: Types of systems – Decision support; I.2.5 [Artificial intelligence]: Programming Languages and Software – Expert system tools and techniques; I.2.6 [Artificial intelligence]: Learning – Analogies.

Introduction

The problem of human reasoning simulating (so called "common sense" reasoning) in artificial intelligence (AI) systems and especially in IDSS is very actual nowadays [1,2]. That is why special attention is turned to case-based and analogous reasoning methods and models. The analogy and precedents (cases) can be used in various applications of AI and for solving various problems [3-7], e.g., for diagnostics and forecasting or for machine learning. AI experts model case-based reasoning by computers in order to develop more flexible models of search for solutions and learning.

In this paper, we consider method of case-based reasoning for a solution of problems of real-time diagnostics and forecasting in RT IDSS [5]. These systems are usually characterized by strict constraints on the duration of the search for the solution. One should note that, when involving models of case-based and analogous reasoning in RT IDSS, it is necessary to take into account a number of the following requirements to systems of this kind [2]:

– The necessity of obtaining a solution under time constraints defined by real controlled process;

- The necessity of taking into account time in describing the problem situation and in the course of the search for a solution;
- The impossibility of obtaining all objective information related to a decision and, in accordance with this, the use of subjective expert information;
- Multiple variants of a search, the necessity to apply methods of plausible (fuzzy) search for solutions with active participation of a decision making person (DMP);
- Nondeterminism, the possibility of correction and introduction of additional information in the knowledge base of the system.

The methods of case-based reasoning may be applied in units of analysis of the problem situation, search for solutions, learning, adaptation and modification, modeling and forecasting. The use of the respective methods in IDSS broadens the possibilities of IDSS and increases the efficiency of making decisions in various problem (abnormal) situations.

Case-based reasoning

Case-based reasoning, like analogous reasoning, is based on analogy; however, there are certain differences in their implementation [5, 8]. In the most encyclopedias, a precedent (from Latin, precedentis) is defined as a case that took place earlier and is an example or justification for subsequent events of this kind. To create a precedent means to give grounds for similar cases in the future, and to establish a precedent is to find a similar case in the past.

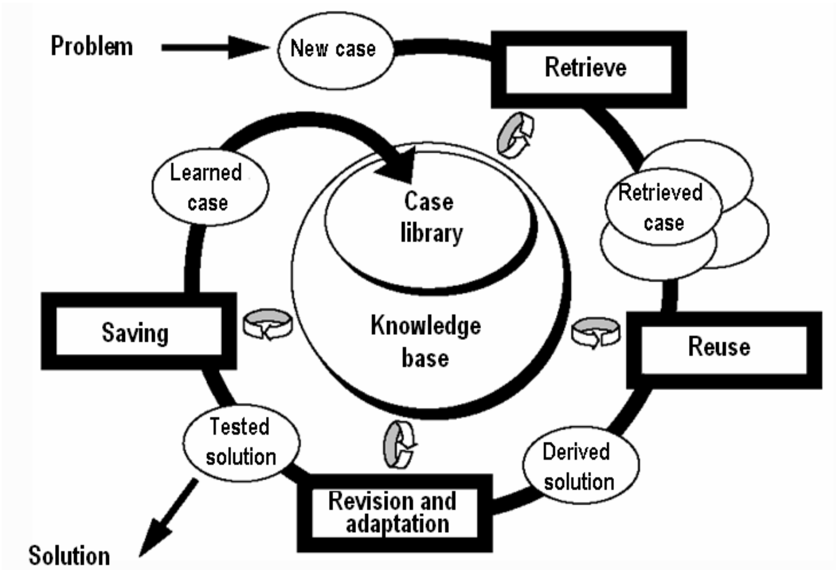


Fig. 1 The structure of CBR-cycle

As the practice shows, when a new problem situation arises, it is reasonable to use this method of case-based reasoning without drawing an analogy. This is caused by the fact that humans operate with these reasoning schemes at the first stages, when they encounter a new unknown problem.

Case-based reasoning is an approach that allows one to solve a new problem using or adapting a solution of a similar well-known problem. As a rule, case-based reasoning methods include four main stages that form a CBR-cycle, the structure of which is represented in Fig. 1 [9].

The main stages are as follows:

- Retrieving the closest (most similar) case (or cases) for the situation from the case library;
- Using the retrieved case (precedent) for solving the current problem;
- If necessary, reconsidering and adaptation of the obtained result in accordance with the current problem;
- Saving the newly made solution as part of a new case.

It is necessary to take into account that a solution on the basis of cases may not attain the goal for the current situation, e.g., in the absence of a similar (analogous) case in the case library. This problem can be solved if one presupposes in the CBR-cycle the possibility to update the case library in the reasoning process (inference) [5, 8]. A more powerful (in detecting new facts or new information) method of reasoning by analogy is a means of updating case libraries. We also note that the elements of case-based reasoning may be used successfully in analogy-based reasoning methods; i.e., these methods successfully complement each other and their integration in IDSS is very promising.

Use of the mechanism of cases for RT IDSS consists in output of the decision to the operator (DMP) for the current situation on the basis of cases which is contained in system. As a rule, the last stage in a CBR-cycle is excluded and performed by the expert (DMP) because the case library should contain only reliable information confirmed by the expert. Reconsidering and adaptation of the taken decision is required seldom because the same object (subsystem) is considered.

The modified CBR-cycle for RT IDSS includes following stages:

– Retrieving the closest (most similar) case (or cases) for the situation from the case library;
– Using the retrieved case (precedent) for solving the current problem.

Case-based reasoning for IDSS consists in definition of similarity degree of the current situation with cases from case library. For definition of similarity degree, the nearest neighbor algorithm (k-nearest neighbors algorithm) is used [10].

The k-nearest neighbors algorithm

The class of selection algorithms used by the most CBR products is called nearest neighbor (k-nearest neighbors). Let's explain the work of algorithm on a simple example. Consider an item that has two attributes: a **temperature** and a **liquid level**. Let us draw all the items at diagram. On the one (x-axis) there is the temperature, say, from **10 – 50°C**, while other (y-axis) contain a value of **liquid level** (for simplicity, let's make the range **1000 – 5000 mm**).

In the library, there are two cases:

– **C_1: t = 30°C**, **h = 3500mm**;
– **C_2: t = 40°C**, **h = 1500mm**.

For the current situation (**Target**): **t = 20°C**, **h = 3000mm**.

If we plotted them on the chart, it might look like fig. 2.

Now we shall calculate the distance from the Target up to C_1 and C_2:

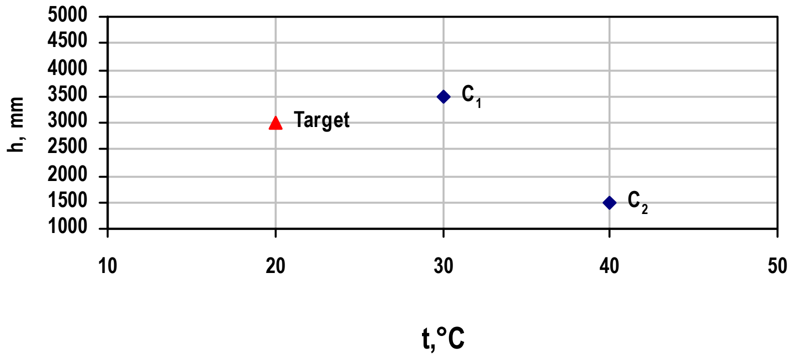


Fig. 2 Coordinate plane

$D_1 = \sqrt{(20-30)^2 + (3000-3500)^2} = 500{,}10$; $D_2 = \sqrt{(20-40)^2 + (3000-1500)^2} = 1500{,}13$.

The maximal distance D_{MAX} between points with coordinates (10,1000) and (50,5000) is similarly calculated (D_{MAX} = 4000,20).

Then values of similarity degree (SIM) of the current situation with two cases from case library are calculated:

– for C_1: $SIM_1=(1 - D_1/D_{MAX})=(1 - 500{,}10/4000{,}20)=0{,}8750$ (87,50%);
– for C_2: $SIM_2=(1 - D_2/D_{MAX})=(1 - 1500{,}13/4000{,}20)=0{,}6250$ (62,50%).

In case there are **n** (**n**>2) parameters for the description of a situation and cases, a more complex variant is considered, and it is differed from the presented one only that **n** coordinates are used.

Further, we shall view the structure of case library for RT IDSS on the basis of nonclassical logics for monitoring and control of complex objects like power units.

The structure of case library for RT IDSS

The case library for RT IDSS should join in itself the cases concerning a particular subsystem of complex object, and also contain the information on each parameter which is used for the description of cases (parameter type and range).

Besides, the case library should include such adjustments, as:
- the significance of parameter;
- a threshold value of similarity;
- a value which limits quantity of considered cases.

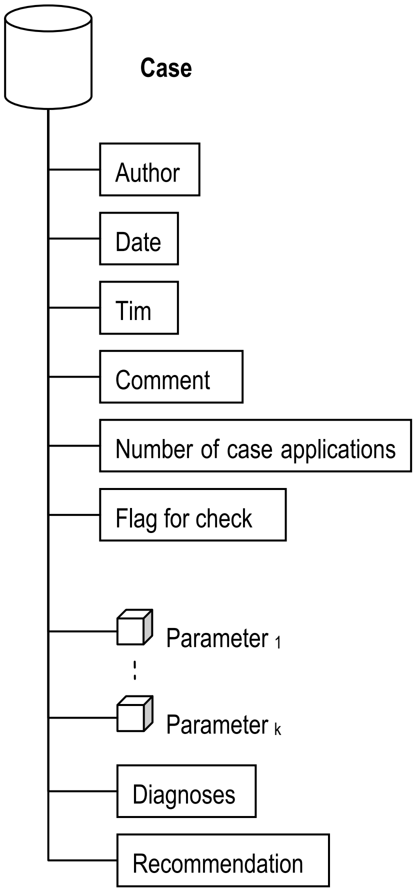


Fig. 3 Case structure

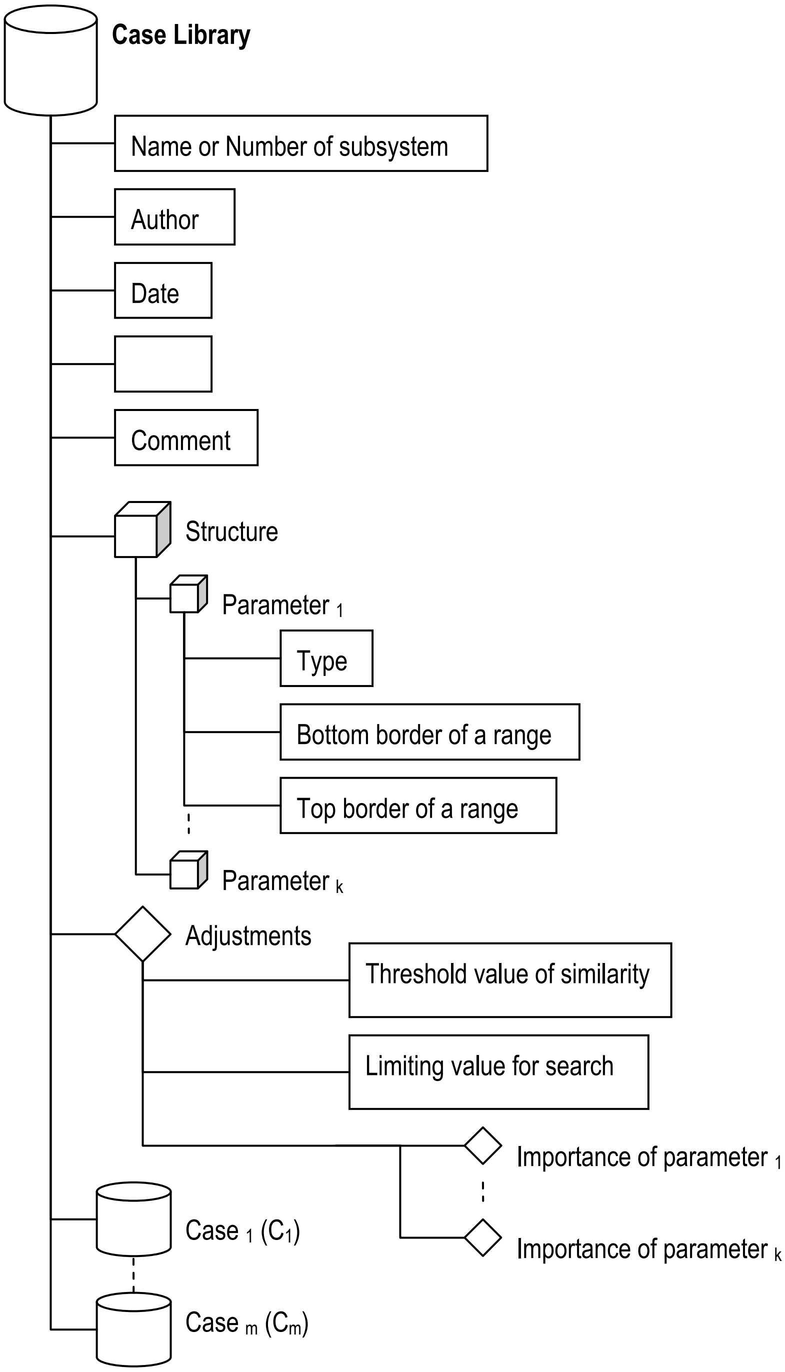


Fig. 4 The structure of case library for RT IDSS

It is necessary to emphasize, that the case library can be formed on the basis of:

- the experience, accumulated by the expert;
- analysis of the system archive;
- the analysis of emergencies;
- operative instructions;
- technological requirements.

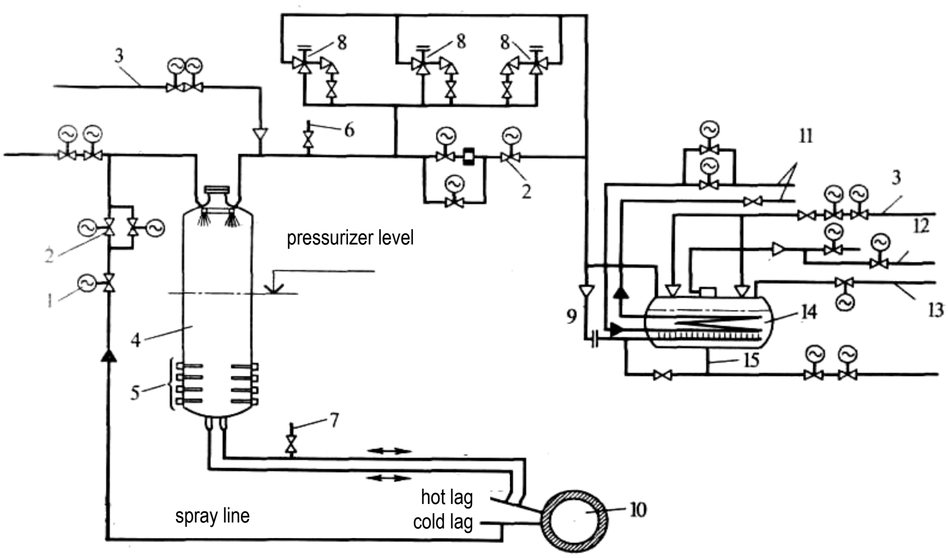
The case library can be included in the structure of the knowledge base of RT IDSS or act as a separate component of the system. Case structure is presented in fig. 3, and the structure of case library in fig. 4.

Application of case-based reasoning for diagnostics of complex object states

As a complex object, we shall understand an object which has a complex architecture with various interrelations, with a lot of controllable and operated parameters and small time for acceptance of operating influences. As a rule, such complex objects as the power unit are subdivided into technological subsystems and can function in various modes (in regular, emergency, etc.).

For the description of such complex object and its subsystems, the set of parameters is used. The state of object is characterized by a set of concrete values of parameters.

In the operative mode reading of parameters values from sensors for all object is made by the system of controllers with an interval in 4 seconds. For this time interval, it is necessary to give out to the DMP (operator) the diagnosis and the recommendation on the developed situation.



1 – isolation valve;
2 – control valve;
3 – nitrogen injection;
4 – pressurizer;
5 – electrical heaters;
6 – sampling of steam;
7 – sampling of water;
8 – safety relief valves;
9 – connector;
10 – reactor;
11 – cooling system;
12 – gas blowdown pipe;
13 – water sparger;
14 – relief tank;
15 – sampling system.

Fig. 5 The technological scheme of the steam pressurizer

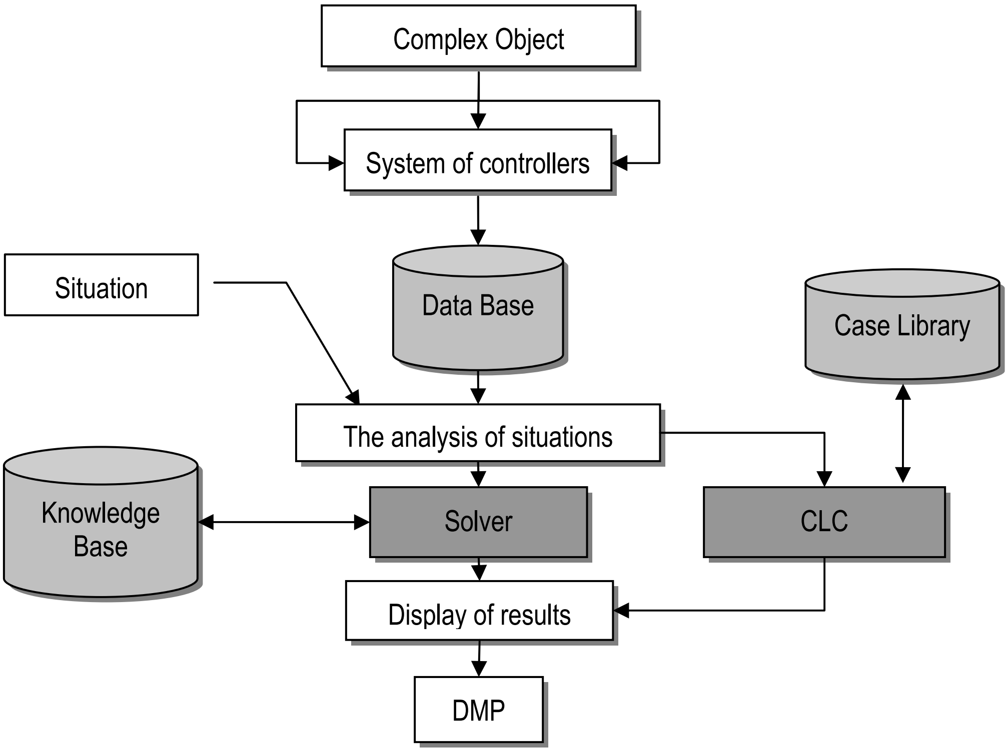


Fig. 6 The scheme of functioning for RT IDSS with use of CLC

Diagnosing and detection of operating influences is carried out on the basis of expert knowledge, technological requirements and operative instructions. The developed software (Case Libraries Constructor – CLC) can be applied to the decision of the specified problems.

Basic components of CLC are:

- module for storage and loadings case libraries and for data import;
- a subsystem of visualization for browsing the structure of case libraries;
- a subsystem of editing and adjustment of case libraries;
- a module of new cases check;
- a subsystem of case library testing and case-based reasoning.

CLC was implemented in Borland C++ Builder 6.0 for Windows NT/2000/XP.

This tool was applied in the prototype of a RT IDSS for monitoring and control of complex objects like power units on an example of a pressurizer in pressurized water reactor (PWR) of the atomic power station (fig. 5, 6).

Implementation of case libraries with use of CLC for systems of expert diagnosing is subdivided into the following main stages:

- Creation of case libraries for subsystems of complex object;
- Adjustment of the created case libraries;
- Addition of cases in case libraries;
- Check of the added cases;
- Testing of the filled case libraries with using case-based reasoning;
- Reservation of the created case libraries for their subsequent transfer to operative maintenance.

Conclusion

The method of case-based reasoning was considered from the aspect of its application in modern IDSS and RT IDSS, in particular, for a solution of problems of real-time diagnostics and forecasting.

The CBR-cycle is considered and its modification for application in RT IDSS is offered.

The k-nearest neighbors algorithm for definition of similarity degree of the current situation with cases from case library is described.

The structure of case library for RT IDSS is proposed.

The proposed method of case-based reasoning was implemented in Borland C++ Builder 6.0 for Windows NT/2000/XP. The main functional components of the implemented tool (Case Libraries Constructor – CLC) are specified.

The presented tool was applied in the prototype of a RT IDSS on the basis of non-classical logics for monitoring and control of complex objects like power units.

The possibility of application of analogous reasoning in case-based reasoning is underlined. We also note that the elements of case-based reasoning may be used successfully in analogy-based reasoning methods; i.e., these methods successfully complement each other and their integration in IDSS is very promising [8, 11].

Bibliography

1. Vagin V.N., Yeremeyev A.P. Modeling Human Reasoning in Intelligent Decision Support Systems // Proc. of the Ninth International Conference on Enterprise Information Systems. Volume AIDSS. Funchal, Madeira, Portugal, June 12-16, INSTICC, 2007, pp.277-282.
2. Vagin V.N., Eremeev A.P. Certain Basic Principles of Designing Real-Time Intelligent Decision Systems // Journal of Computer and Systems Sciences International, v. 40(6), 2001, pp. 953-961.
3. Kolodner J.L. Improving human decision making through case-based decision aiding // AI Magazine, 12(2), 1991, pp. 52–68.
4. Leake D.B. CBR in Context: The Present and Future // Case-Based Reasoning: Experiences, Lessons and Future Directions, AAAI Press / The MIT Press, 1996, pp. 3–31.
5. P.R. Varshavskii, A.P. Eremeev Analogy-Based Search for Solutions in Intelligent Systems of Decision Support. Integrated models and flexible calculations in artificial intelligence. Journal of Computer and Systems Sciences International, Vol. 44, No. 1, 2005, pp. 90–101.
6. Eremeev, P. Varshavsky Analogous Reasoning for Intelligent Decision Support Systems. Proceedings of the XI-th International Conference “Knowledge-Dialogue-Solution” – Varna, vol. 1, 2005, pp. 272-279.
7. A.P. Eremeev, P.R. Varshavsky Reasoning by structural analogy in intelligent decision support systems // Proceedings of the Seventh Joint Conference on Knowledge-Based Software Engineering JCKBSE’06. IOS Press, 2006, pp. 303-306.
8. Alexander Eremeev, Pavel Varshavsky Analogous Reasoning and Case-Based Reasoning for Intelligent Decision Support Systems // International Journal INFORMATION Theories & Applications (ITHEA) 2006, Vol.13 № 4, pp. 316-324.
9. Aamodt, E. Plaza “Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches”. AI Communications, No. 7, 1994.
10. Alexander Eremeev, Pavel Varshavsky Application of Case-based reasoning for Intelligent Decision Support Systems // Proceedings of the XIII-th International Conference “Knowledge-Dialogue-Solution” – Varna, vol. 1, 2007, pp. 163-169.
11. Eremeev A., Varshavsky P. Methods and Tools for Reasoning by Analogy in Intelligent Decision Support Systems // Proc. of the International Conference on Dependability of Computer Systems. Szklarska Poreba, Poland, 14-16 June, 2007, IEEE, pp.161-168.

Authors' Information

A.P. Eremeev – Applied Mathematics Department of the Moscow Power Engineering Institute (Technical University), Krasnokazarmennaya str., 14, Moscow, 111250, Russia; e-mail: eremeev@appmat.ru

P.R. Varshavsky – Applied Mathematics Department of the Moscow Power Engineering Institute (Technical University), Krasnokazarmennaya str., 14, Moscow, 111250, Russia; e-mail: VarshavskyPR@mpei.ru

KNOWLEDGE-BASED ROBOT CONTROL

Agris Nikitenko

Abstract: The paper is related with the problem of developing autonomous intelligent robots for complex environments. In details it outlines a knowledge-based robot control architecture that combines several techniques in order to supply an ability to adapt and act autonomously in complex environments. The described architecture has been implemented as a robotic system that demonstrates its operation in dynamic environment.

Although the robotic system demonstrates a certain level of autonomy, the experiments show that there are situation, in which the developed base architecture should be complemented with additional modules. The last few chapters of the paper describe the experimentation results and the current state of further research towards the developed architecture.

Keywords: Intelligent robots, autonomous intelligent systems, autonomous robots. Artificial intelligence.

ACM Keywords: I.2.4.J – Representations, I.2.6.A – Analogies, I.2.6.E – Induction, I.2.8.G – Plan execution, formation, and generation, I.2.9.A – Autonomous Vehicles

Introduction

Due to the constantly increasing interest about autonomous systems for application in various fields that require a certain degree of autonomy, it is necessary to develop platforms or architectures, which fit the demand. This paper describes an alternative knowledge-based architecture that combines several well known techniques of artificial intelligence in order to increase the system's autonomy. The most important advantage of the described architecture hides in the usage of the symbolic representation of the system's knowledge that is easy to use by the researcher and the system itself.

While the paper relies on the previous research in the field, only the most fundamental definitions are given [Nikitenko EMS2006, Nikitenko KDS 2005, Nikitenko 2005].

A complex environment is described with the following fundamental properties [Druzinin 1985]:

– uniqueness – usually complex systems are unique or number of similar systems is insignificant.

– hardly predictable – complex systems are very hard to predict.

– ability to maintain a certain progress resisting against some outer influence.

According to the sources used [Russell 2003, Huang 2003, Antsaklis 1996, Knapik 1997], an autonomous intelligent system is defined as any artificial intelligent system that can achieve its goals using its own knowledge, experience and available decision alternatives as well as operating without any outer assistance.

According to the definitions of the complex environment and autonomous system, the previous research resulted in development of knowledge-based architecture that supplies the basic functionality for autonomous system operating in a complex environment. The basic features of the developed architecture are outlined in the next section.

Basic Features of an Intelligent System

Summarizing the basic qualities of the proposed architecture are as follows [Nikitenko EMS 2006, Nikitenko KDS 2005, Nikitenko 2005]:

– Ability to reason about facts that are not observable directly by the system. This ability is achieved by means of the deductive reasoning. The proposed architecture does not state the kind of deductive reasoning that should be used. The only rule is that the selected deductive reasoning method has to address demands of a particular task. As it is described above the complex environments may be very dynamic and even with stochastic features. Therefore some uncertain reasoning techniques may be the most suitable. For example the experimental system implements a certainty factor based reasoning [Buchanan 1982].

– Ability to learn. As it is assumed above the intelligent system eventually will not have a complete model of the environment. Therefore the environment will be hardly predictable. Also complex environments are dynamic – in other words the system will face with new situations very often. Obviously, some adaptation mechanisms should be utilized. From point of view of intelligence the adaptation includes the following main capabilities: a capability of acquiring new knowledge and adjustment of the existing knowledge. The inductive reasoning module refers to the capability of learning. During an operation the intelligent system collects a set of facts (observations) through sensing the environment that forms an input for the learning mechanism.

– Ability to reason associatively. This feature is necessary due to the huge set of different possible situations that the intelligent system may face with in the complex environments. For example, there may be two different situations that can be described by n parameters (n is big enough number) where only k parameters are different (k is small enough number). Obviously these situations may be assumed as similar. Therefore an associative reasoning is used – to reason about situations that are observed for the first time by the intelligent system similarly to reasoning about experienced situations. The associative reasoning is realized through using associative links among similar situations (descriptions). Each situation may be accessed or identified by a set of features thus this mechanism operates in similar manner to the associative memory [Kokinov 1988, Wichert 2000]. An issue about which situations should be linked is conditioned by a particular task or goals of the system's designer.

– Ability to sense an environment. This feature is essential for any intelligent system that is built to be more or less autonomous. This feature also includes an ability to recognize situations that the system has faced with as well as an ability to obtain data about unknown situations. All sensed data is structured in frames (see below). During the frame formation process the sensed environment's state is combined with system's inner state thereby allowing the system to reason about the system itself. Also the sensed system's and environment's states are used to realize a feedback in order to adjust the system's knowledge. Thereby the system's flexibility is increased.

– Ability to act. This feature is essential for any intelligent system that is designed to do something. If the system (autonomous) is unable to act, it will not be able to achieve its goals. The way of acting and the purpose of acting vary depending on the goals of the system's designer.

The listed above features form a basis for an intelligent system that operates in a sophisticated environment. According to the features of complex systems that are listed above, any of them may be implemented as it is needed for a particular task. In other words the implementation methods and approaches are dependent on the purposes of the system itself. Nevertheless the main question is how to bind all of them in one whole - one intelligent system. Obviously, there is a necessity for some kind of integration. There are many good examples of different kinds of integration that may be found in widely available literature devoted to hybrid intelligent systems [Goonatilake 1995].

The developed architecture is based on so called intercommunicating hybrid architecture where each of the integrated modules is independent, self-contained, intelligent processing unit that exchanges information and performs separate functions to generate solutions [Goonatilake 1995]. The developed architecture in more details is described in the next section

Architecture of the Intelligent System

According to the list of the basic features there can be outlined the basic modules that correspond to the related reasoning techniques. The modules are outlined in the figure 1.

According to the figure 1, there are four basic modules that form system's kernel. The modules fulfill the following basic functions

- Deductive reasoning module This module performs deductive reasoning using if..then style rules [Luger 2002, Russell 2003]. In order to implement the adaptation functionality, this module may exploit some particular uncertain reasoning technique. In the proposed architecture the main purpose of this module is to predict (forecast) future states of the environment as well as the inner state of the system. During the reasoning process if..then rules are used in a combination with the input data obtained from the sensors.

– Inductive reasoning module This module performs an inductive reasoning. It learns new rules and adds them to the rule base. Again the proposed architecture does not state what kind of inductive learning technique is used. The only limitation is the requirement to produce rules that could be used by the deductive reasoning module. For example, if the fuzzy reasoning is used, then the result is a set of fuzzy rules.

– Case based reasoning module Case based reasoning operates with "best practice" information that helps to reduce planning time as well as provides this information to modeler in an explicit manner.

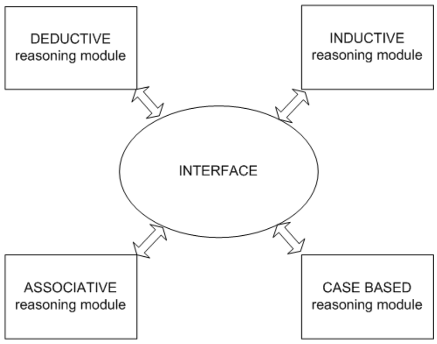


Figure 1. Basic modules

– Associative reasoning module. This module links situations according to the similarity among them thus allowing to reason associatively. In the robotic system the similarity measure is calculated using the following formula:

$$yi = \begin{cases} 1, if \sum_{j=1}^{n} \partial(x_j) \geq T \\ 0, if \sum_{j=1}^{n} \partial(x_j) < T \end{cases} \quad (1)$$

In the formula (1):

yi – 1 – if the i-th situation is similar to the given;

n – a number of attributes that describes each situation;

xj – value of the j-th attribute of the given situation;

T – a threshold value – an number of attributes which values are equal for the given and the i-th situation. In the robotic system (see below) is applied only for 8 IR sensors, all other situation parameters have to be equal. Therefore the T's value may be changed from 1 to 8.

$\partial(x)$ - 1 if the value of the j-th attribute is equal for both situations;

This module considerably reduces the overall amount of knowledge processed by the system because there is no need to store every experienced situation just unique situations, which are not like any other experienced before.

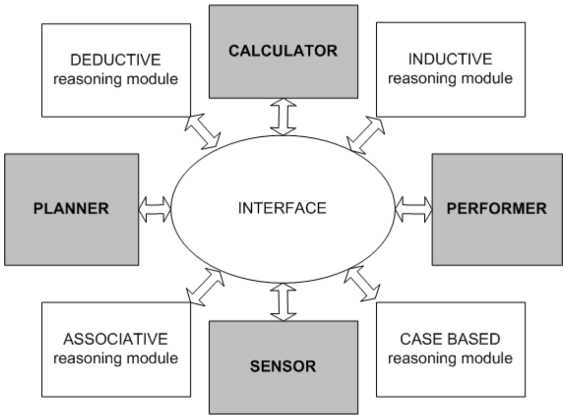


Figure 2. Enhanced architecture

Of course, the intelligent system needs additional modules that would supply it with the necessary information about the environment and mechanisms to perform some actions. Therefore the basic architecture shown in figure 1 is complemented with few additional modules. The enhanced architecture is depicted in the figure 2. The additional modules (drawn in grey) have the following basic functions:

– Planner module. This module is one of the central elements of the system. Its main function is to plan future actions that lead to achievement of the system's goals. During the planning process three of the basic reasoning techniques are involved – deductive, case based and associative reasoning. A result of the planner is a sequence of actions that are expected to be accomplished by the system thereby achieving its goals.

– Sensor module. The module's purpose is to collect information from the sensors about the environment's and the system's states. The sensed information is portioned in separate frames (see below) and forwarded to the interface (see figure 3). Once the information is forwarded, it is available for the other modules.

– Performer module. This module performs a sequence of actions that are included in the plan. Also this module uses information about the system's and the environment's current states in order to determine whether the instant actions can be accomplished.

– Calculator module. This module collects and produces any reasoning relevant quantitative data. For example, in the robotic system (see below) this module is used to calculate certainties of the rules including those rules that are newly generated by the inductive reasoning module. Functionality of the module may be enhanced according to the necessities of the particular tasks or goals of the system's designer.

As it is depicted in figures 1 and 2, all of the modules use the central element – Interface in order to communicate to each other. They are not communicating to each other directly thereby a number of communication links is reduced as well as all of the information circulating in the system is available for any module. A architecture of the interface is depicted in the following figure 3:

The architecture consists of several basic elements. The fundamental element of the whole architecture is situation.

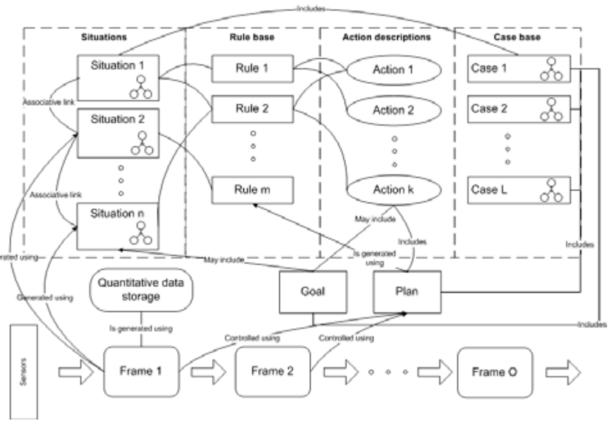


Figure 3 Architecture of the interface.

Situations. Situations are the key elements in the interface structure. They correspond to the situations which are experienced by the system. Every situation is described with a set of features (attributes). Each attribute is described with its value. As it is depicted in the figure 3 situations are linked to each other by associative links. These links form the basis for associative reasoning. When the intelligent system runs into a certain case the most likely situation is activated. It is used for reference to rules and cases. If there is no rule that can be triggered, then the system tries to trigger rules that refer to the bounded similar situations (with same degree of likeness). The result may be less feasible, but using association among situation the system can run out of the dead end cases. The idea is obtained from the associative memory mechanism [Kokinov 1988, Wichert 2000]. It also reduces the impact of the sensor errors on the reasoning process, because allows to use similar not directly matching situation. Thereby the overall amount of knowledge is reduced as well.

Rules. Rules are any kind of notation that represents causalities. In the practical experimentations a well known if..then notation was used [Luger 2002, Russell 2003]. As it is depicted in the figure 3 rules are linked to situations and actions. When the system activates several situations by using associative links the appropriate rules are also activated thus system can scan a set of "associated" rules as well. This simple mechanism improves the ability to adapt. As it is depicted in the figure 3 rules are linked to actions. Thereby rules through the deductive reasoning are used in planning process. In the robotic system each rule is complemented with a certainty factor which is used during the reasoning process (the certainty theory's simplified practical model is used) [Buchanan 1982]. The certainty factor is calculated using the following formula:

$$CF(Rule) = \frac{S}{N} \quad (2)$$

In the formula (2):

CF(Rule) – certainty factor of the rule;

S – number of times when the rule's forecasted values of the situation attributes (one or many) were observed by the system;

N – number of times when the rule has been used;

Case quality or value is calculated using the same formula. In that case:

S – number of times when case was used and appropriate goal was achieved;

N – number of times when the case was used;

Actions. Actions are symbolic representations that can be translated by the intelligent system and cause the system to do something. For example “turn to the right” causes the system to turn to the right by 900. Each action consists of three parts: precondition, body and postcondition. Precondition is every fact (attribute of the situation) that should be true before the action is executed. For example, before opening the door it has to be unlocked. Body is a sequence of basic (or lover level) actions that are executed directly – for example a binary code that forwarded to a motor controller causes the motors to turn (in the case of robotic system). Post conditions are factors that will be true after the execution. For example after opening the door, the door is opened.

Cases. Cases are direct descriptions of the system’s experience. Mathematically a case is described as follows:

$$Case = \{E, Pl, G\} \quad (3)$$

In the formula (3):

Case – The case;

E – situation or input;

Pl – plan which leads to achievement of the goal;

G – goal;

In the robotic system each case was complemented with reliability factor that is calculated using the formula (2), with difference that S is a number times when the case has been used and the goal was achieved.

Frames. Frames are data structures that contain the sense array from the environment and the system. It means that frames contain snapshots of the environment’s and the system’s states. Mathematically it may be described as follows:

$$Frame = \{En, Sy\} \quad (4)$$

In the formula (4):

En – a snapshot of the environment’s state;

Sy – a snapshot of the system’s state;

As it is depicted in figure 3 frames are chained one after another thus forming a historical sequence of the environment’s and the system’s states. Frames form an input data for the learning (induction module) algorithms as well.

Goal. The goal is a task that has to be accomplished by the system. It can be defined in three different ways: as a sequence of actions that should be done, as some particular state that should be achieved or as a combination of the actions and the states. The third option is implemented in the robotic system described below. Thereby the goal is described as:

$$G = \{S, M, C\} \quad (5)$$

In the formula (5)

G – the goal;

S – a set of states that has to be achieved;

M – a set of actions that has to be performed;

C – order constraints that order elements of the sets S and M;

Plan. Plan is a sequence of actions that is currently executed by the system. It may be formed using both basic and complex actions. After the plan is accomplished it is evaluated depending on whether the goal is achieved or not thereby forming feedback information for the calculator module.

Quantitative data. This element is used to maintain any kind of quantitative data that is produced by the calculator module and is used during the reasoning process. For example it may contain certainties about facts or rules, possibilities etc. Quantitative data is collected during the reasoning process as well as during the analysis of the input data - feedback data. All of those components together form an interface for the basic modules: Inductive, Deductive, Case based and Associative reasoning. The architecture is implemented as experimental robotic system that is shortly described below.

Experimental Robotic System

The robotic system is an autonomous intelligent system that encapsulates all of the mentioned above elements of the proposed architecture and interface among them. The system's input consists of the following sensors:

- Eight IR (infrared) range measuring sensors;
- Electronic compass;
- Four bump sensors (two front and two rear micro switches)
- Four driving wheel movement measuring resistors (two for each driving wheel in order to achieve reliable enough measurements).

The sensors and robotic system is depicted in figure 4.

Two Basix-X [BasicX] microprocessors are used in order to communicate with PC and to perform input data preprocessing and formatting. Prepared and formatted data as frames (see above) are sent to the PC via RS-232 connection [Strangio 2005]. Few screenshots of the controlling software are shown in the following figures:

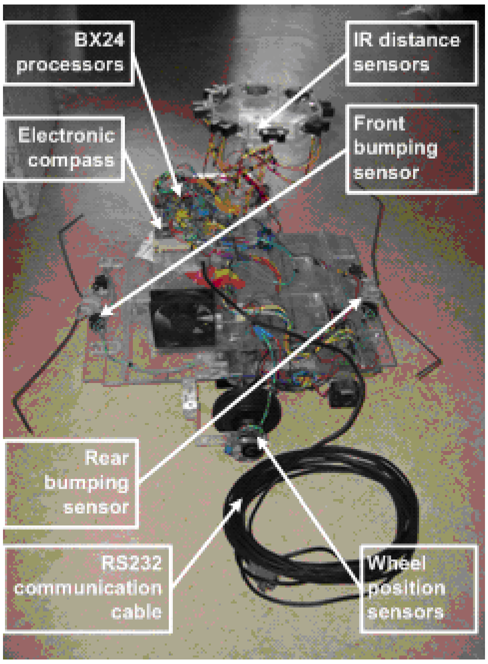


Figure 4. The robotic system.

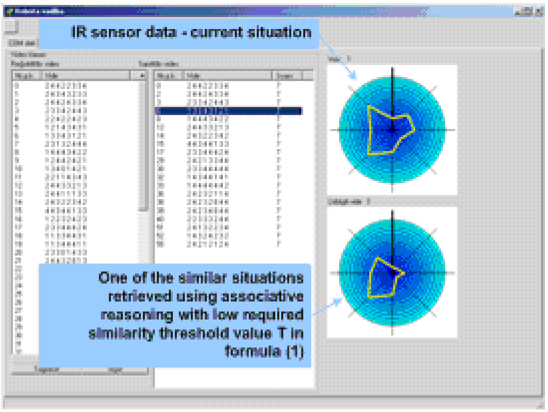


Figure 5. Contr. Software – sensor data section.

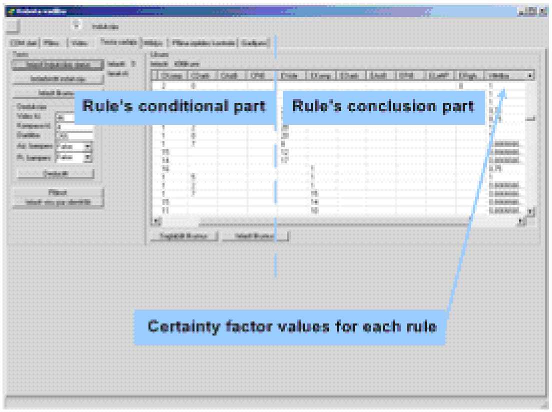


Figure 6. Contr. Software – rule base section.

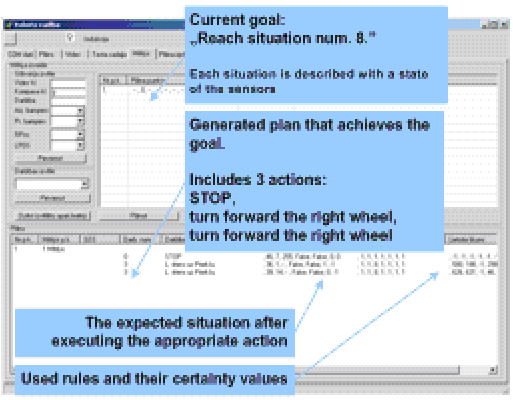


Figure 7. Contr. Software – goal definition

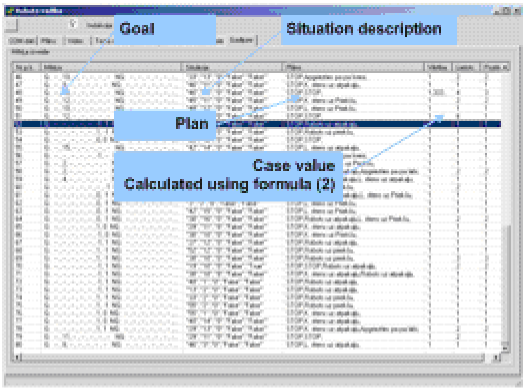


Figure 8. Contr. Software – Case base section.

All other modules of the intelligent system are implemented as a PC-based software that has a user-friendly interface allowing a simple following the system's operation, collection of the research sensitive data, changing system's goals, etc….

The system is built for research purposes only. In other words, it is built for experiments in order to examine and validate the proposed architecture. Therefore the system's user interface is built to be as flexible as possible allowing its user to manipulate with the robot's state, goals and results at the runtime. The most important features of the robotic system are:

- Ability to work with multiple goals with mixed structure that may include – actions, states or both;
- Ability to adapt via using inductive learning algorithm C 4.5 [Quinlan 1996]. Even with the well known disadvantages of the axis parallel classification that is used in the algorithm, the system demonstrates acceptable adaptation abilities. In practical implementations there may be used other methods described in [Centu-Paz 2000, Pappa 2004] that eliminate these problems.
- Case-Based reasoning is used to store information about best-practice cases and to use this information during the planning process.
- Ability to reason using Certainty theory ideas thus allowing addition of new rules that may be conflicting with the existing ones in the rule base.
- Ability to reason using associative links among the situations.
- System's knowledge and the system's state relevant data is stored and processed in explicit and easy to follow manner thus demonstrating the advantages of the used knowledge based techniques.

It is important to stress that at the very beginning of the system's operation it has no information about the consequences of each action – it needs to learn them. Thereby the bottom-up learning is used. But if it is necessary the system's rule base may be filled with rules, cases and other research relevant information thus allowing to model particular state of the system.

Experiments

There are several experiments accomplished by the author in order to examine the system and its behaviour in different working conditions. One of the most important experiments is described in this section.

Experiment's goals

The experiment's primary goal is to prove the architecture's ability to adapt and autonomously achieve the given goals as well as to characterize the system's behavior in uncertain conditions. A secondary goal is a demonstration of the system's operation in uncertain and dynamic conditions.

Experiment conditions

In order to meet the experimentation goals, a 3 × 3 m arena is used. The robotic system can freely move around the arena, but cannot move outside the arena because of special 50 cm high walls around it.

At the beginning of the experiment the system's knowledge base is empty containing only a list of actions that may be executed. It means that the system has to acquire knowledge form the observations (sequence of sensor state snapshots) in order to achieve the given goals. 120 random goals one after another are given to the system. The maximum plan length is limited to 3 actions.

In order to simulate uncertain conditions, the system is randomly turned in unexpected directions thus causing random state transitions. In order to simulate random events, the experimenter from time to time is walking inside the arena in unexpected directions thus causing random state transitions.

Results

The main measure is effectiveness that is calculated using the following formula:

$$E = \frac{M}{P} \times 100\% \qquad (6)$$

In the formula (6):

E – effectiveness expressed in percents; M number of goals achieved by the system; P – number of plans generated during the achievement of the goals;

The results are outlined in the graphics on figures 9, 10, and 11.

The first graphic (see figure 9) shows the number of plans constructed by the system during the run. As it is depicted in the figure 4 at the beginning the number of plans is very high (34), approximately 3.4 plans for one goal. While the system adapts the number of constructed plans is gently decreasing until reaches the minimum – 10 plans for 10 goals. The second graphic (see figure 10) shows the number of learning cycles fired by the system. The shape of the graphic almost repeats the first graphic. The learning mechanism at first collects a set of examples which is used for rule generation. Thereby the less actions are performed the more time is required for collection of the set with appropriate amount of examples. The third graphic shows the number of cases used for plan acquisition (each case includes a ready-to-use plan). The number of cases used is increasing in correspondence with the growth of the system's experience. The last 10 goals are achieved using 7 cases. The planning took almost 6 s during the achievement of the last 10 goals while the retrieval of the appropriate case took only 0.05 s (more than x100 faster). This emphasizes the importance of the case based reasoning in autonomous systems. Also results of the other experiments showed that associative reasoning more the 10 times reduced the amount of knowledge (situation descriptions, rules and cases) used by the system. This lets to decrease planning time and increase flexibility of the system.

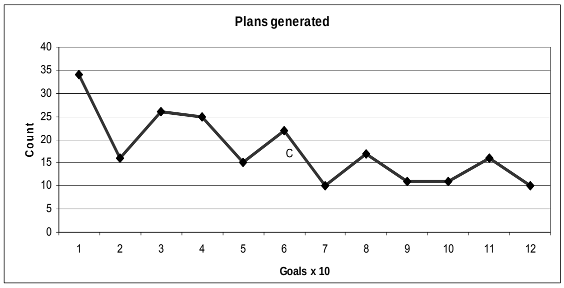


Figure 9. A number of plans constructed by the system

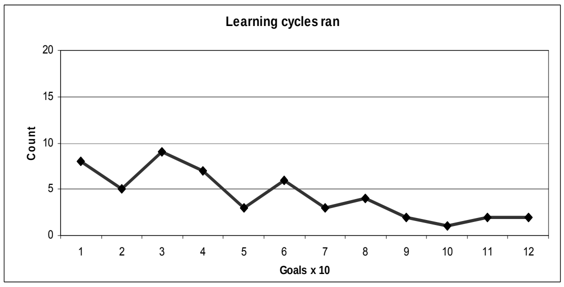


Figure 10. A number of learning cycles

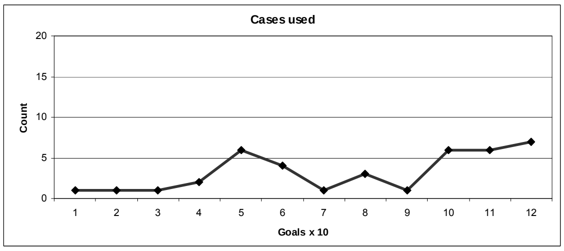


Figure 11. A number of cases used to construct plans

Other experiments

The systems behavior is compared with system's operation with randomly generated plans. During this experiment remarkably simpler goals are given – goals that may be achieved with one action. For example, a certain state when one wheel is turning forward (with empty knowledge base the system does not 'know' consequences of any action, they has to be learned). With each plan length (1, 2, 3, 4, 6) 10 goals are given for achieving in the same arena.

The results are shown in the figure 12.

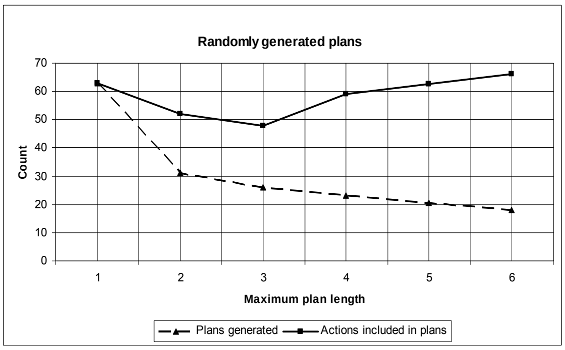


Figure 12. Number of generated plans and included actions

This experiment shows that the maximum effectiveness (aprox. 2 plans per goal) is reached with plan length 6 while it is necessary only one action. The figure 12 shows that at the end of experiment for each goal only 1 plan is required in more complex conditions. Consequentially, the used mechanisms and the knowledge representation schema provide means for certain convergence shown in the previous figures.

Another experiment is conducted in order to examine an importance of the associative reasoning. Like in the previous experiment with random plans, the same conditions are used as well as 10 goals with each example set are given.

In the figure 13 is shown system's performance with different learning sets – 10, 15, 20 and 25 examples per set.

During the next run an associative reasoning is used with different threshold values T (see formula (1) and appropriate descriptions). All other conditions remain the same. Results are shown in the figure 14.

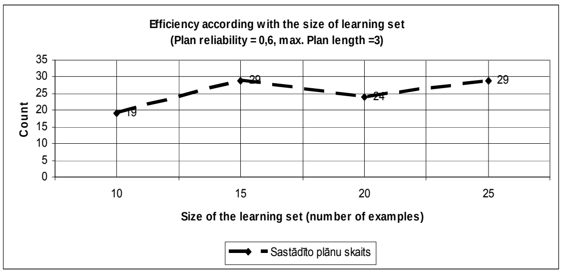


Figure 13. Number of generated plans with different learning sets

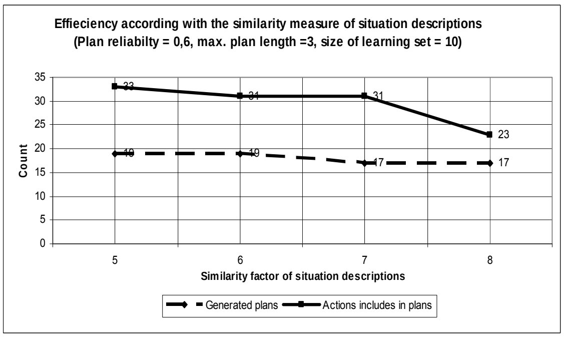


Figure 14. Number of generated plans and included actions with different T values

In both runs a considerable performance growth is achieved comparing with the random plan generation results. When the associative reasoning is used the system's performance is slightly higher. The slight difference is caused by the small number of goals achieved because the associative reasoning is based on the previous experience. Therefore if the system is not experienced enough the associative reasoning cannot provide considerable performance growth. Nevertheless the performance is increased.

Importance of the case based reasoning may be observed in figure 14, where the usage of cases is increasing together with overall experience of the system. For example, the first 10 goals are reached without using cases while the last 10 goals are achieved applying 7 cases.

Other experiments that are not presented in this paper were conducted in order to better describe some specific parameters of the system's behavior under different conditions.

Current State of Further Research

As it is shown above, the implemented architecture as a basic input unit uses frame, which consists of system's and the environment's state snapshots. Each snapshot is described by means of system's sensor states. Thereby the goal also has to be described in means of system's senor states. This results in a significant limitation if the system has to achieve some global goal – goal that is "out of the system's range of sight". For better description let us consider the following simple task:

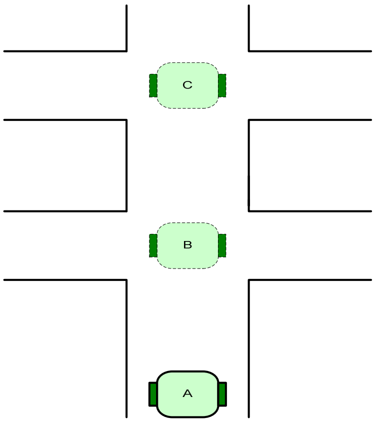


Figure 15. System's task

Where:

A – the systems' current state,

C – system's goal state,

B – unwanted state

If the desired state is C and it is out of the system's sensor range, then it is more likely that the system will stop at the state B because its representation by means of the system's sensor state is identical with the state's C representation. In other words, the system can achieve only local goals. However, having information about states B and C, it is possible to define two separate goals that follow one after another. Obviously, the system might use a geographical or topological map of the environment that could provide the necessary information about states B and C. That is the main direction of the further research that is currently undertaken ie. enhancement of the developed architecture in order to use geographical and topological maps. Map as an additional model would allow to operate with tasks that require: environment investigation and construction of the map, self localization, object lookup and others.

Currently the research is based on a computer model of the robotic system and its environment. The computer model in contradiction with the real system allows much easier to compare different techniques used for implementation of certain modules. The model uses MS Robotic Studio infrastructure and visualization tool while the robot control software is implemented using a general purpose programming language. The used infrastructure provides all basic functionality for modeling a robot and its environment as a visual entity and as a physical entity as well. The following figures show the difference between mentioned two views of the same modeled entity.

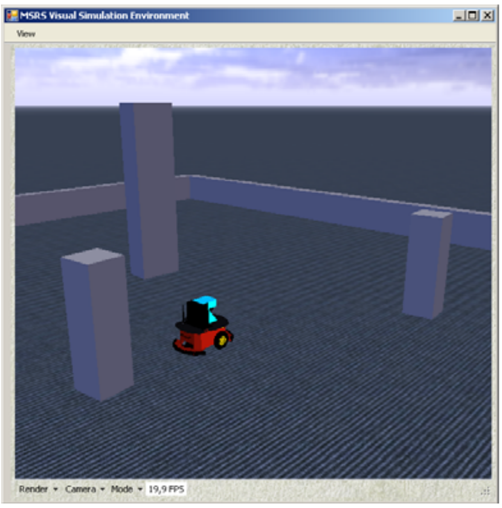


Figure 16. Model's visual view

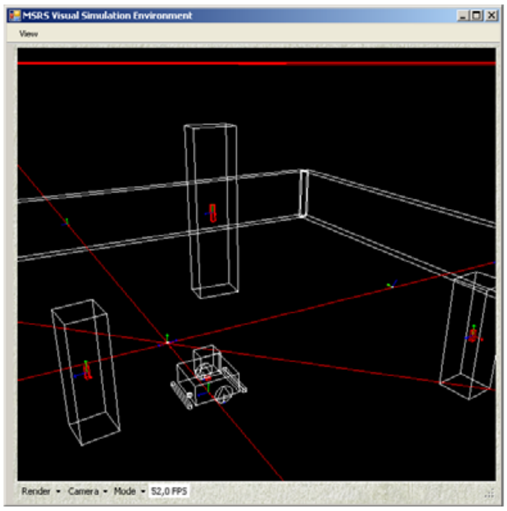


Figure 17. Model's physical view

The visual model is used for convenience of the user, but the physical view is used for calculations. The physical view is less detailed in order to maintain only the essential calculations of the modeled entities.

The modeling tool has been chosen because of the following features:

- Functionality, which is tailored for specific needs of robotics i.e. all of the most popular devices are supported (the others may be added by the user), wheel traction, mass and inertia and other very important physical aspects are supported.
- Enhanced service oriented architecture that supports event-driven control over the modeled robotic system
- Platform independence of the controlling software that allows to switch between model and real robot by adjustment of the service references without any changes in the controlling code.
- Support of serialization of the developed model that allows to "save" the current state and continue modeling from this particular state whenever it is necessary.

The mentioned basic features make the MS Robotic Studio very convenient tool for development and modeling of robotic systems.

Currently all of the basic modules are already implemented using the modeling tool. The link between core modules and map module is in the development stage. Although the system is not developed yet, the test drives of the implemented modules show the ease of use of the mentioned modeling tool and performance of the development that is increased due to the usage of the high level programming language instead of the low level processor specific languages.

As it is stressed in author's previous papers it is necessary to examine the developed architecture under different configurations. For example, the architecture might be implemented using different induction algorithms. The developed model will allow to switch among different possible configurations for more detailed examinations of the developed architecture in the same conditions.

Conclusions

The conducted experiments show that in spite of the known drawbacks of the used algorithms (in particular C4.5) the described architecture provides means for convergence in uncertain and dynamic conditions. However a deeper analysis outlines important limitations that may be avoided via an appropriate goal definition or via using additional models. The further research activities are concentrated on geographical and topological map integration into the developed architecture as a separate module. That would provide the necessary information for achievement of goals that are "out of the direct sight" of the robot's sensors and widen the sphere of application of the architecture.

The conducted experiments do not compare performance of different configurations of the developed architecture that is still an important issue of further examination.

Bibliography

[Antsaklis 1996] P.J. Antsaklis, M. Lemmon, J.A. Stiver, Learning to be autonomous: Intelligent supervisory control. Intelligent Control Systems: Theory and Applications. England, IEEE Press, 1996, pp. 28 – 62.

[BasicX] © NetMedia Inc. www.basicx.com

[Buchanan 1982] B.G.Buchanan, R.O.Duda, Principles of Rule-Based Expert Systems. Stanfrod, USA, Stanford University, 1982, 62 p., STAN-CS-82-926

[Burkhard 2001] H.Burkhard, J.Bach, R.Berger, B.Brunswieck, M. Gollin, Mental models for Robot Control, Proceeding of International seminar "Advances in Plan – based control of Robotic agents", Dagstuhl Castle, Germany, 2001.

[Centu-Paz 2000] E.Cantú-Paz, C.Kamath, Using Evolutionary Algorithms to Induce Oblique Decision Trees. Proceedings of Genetic and Evolutionary Computation Conf. (GECCO), Las Vegas, USA, 2000.

[Davidsson 2000] P.Davidsson, Multi Agent Based Simulation: Beyond social simulation, In Multi Agent Based Simulation, LNC series 1979, Springer Verlag, 2000, pp. 97 - 107.

[Drogoul 2002] A.Drogoul, D.Vanbergue, T.Meurisse, Multi-Agent Based Simulation: Where are the Agents?. Proceedings of Multi-Agent Based Simulation (MABS), Bologna, Italy, 2002.

[Druzinin 1985] V.V.Druzinin, D.S. Kontorov "Systemtechique", Radio and Communication, 1985, 200 p., (В.В. Дружинин, Д.С. Конторов «Системотехника» Радио и Связь 1985, 200 с.)

[Goonatilake 1995] S.Goonatilake, S.Khebbal, Intelligent hybrid systems. Biffind Lane, Chichester, West Sussex, England, John Wiley & sons Ltd. 1995, 325 p., ISBN 0-471-94242-1

[Huang 2003] H.Huang, E.Messina, J.Albus, Autonomy Level Specification for Intelligent Autonomous Vehicles: Interim Progress Report. Proceedings of the 2003 Performance Metrics for Intelligent Systems Workshop, Gaithesburg, USA, August 16 - 18, 2003, pp. 1 – 7.

[Knapik 1997] M.Knapik, J.Johnson, Developing Intelligent Agents for Distributed Systems. USA, Osborne/McGraw-Hill, 1997, 398.p. ISBN 0-0703-5011-6.

[Kokinov 1988] B. Kokinov, Associative memory-based reasoning: How to represent and retrieve cases. In T. O'Shea and V. Sgurev (Eds.), Artificial intelligence III: Methodology, systems, applications. Amsterdam, The Netherlands, Elsevier Science Publishers B.V., 1988, pp. 55 – 58.

[Luger 2002] G.F.Luger, Artificial intelligence – Structures and Strategies for Complex problem Solving. 4th edition, Addison Weslwy, 2002, translation to rus., Искуственный интеллект – Стратегии и методы решения сложных проблем, Москва, Россия, Издательский дом «Вильямс», 2005, 863.с. ISBN 0- 201-64866-0.

[Nikitenko EMS 2006] A.Nikitenko, Autonomous Intelligent Agent Control in Complex Environments, Barcelona, Spain, EMSS 2006 Conference Proceedings, 2006. p. 251 - 260

[Nikitenko KDS 2005] A.Nikitenko, Robot Control Using Inductive, Deductive and Case Based Reasoning. Varna, Bulgaria, KDS 2005 conference proceedings, 2005, Vol. 2, pp. 418 - 427.

[Nikitenko 2005] A.Nikitenko, Intelligent Agent Control Using Inductive, Dedudctive and Case Based Reasoning. Riga, Latvia, ECMS 2005 Conference Proceedings, 2005, pp.486 - 492.

[Pappa 2004] G.L.Pappa, A.A.Freitas, Towards a genetic programming algorithm for automatically evolving rule induction algorithms. Proceedings of ECML/PKDD-2004 Workshop on Advances in Inductive Learning, Pisa, Italy, 2004, pp. 93 – 108.

[Quinlan 1996] J.R.Quinlan, Improved Use of Continuous Attributes in C4.5. In Journal of Artificial Intelligence Research, 1996, Vol. 4, pp. 77 – 90

[Russell 2003] S.Russell, P.Norvig, Artificieal Intelligence – A Modern Approach. 2nd edition, Upper Saddle River, New Jersey, USA, Pearson Educations Inc., 2003, 1080 p., ISBN 0-13-080392-2.

[Strangio 2005] C.E.Strangio, The RS232 standard [online], CAMI Research Inc., Lexington, Massachusetts, 1993, [cited: 20.12.2005]. Available from World Wide Web: <http://www.camiresearch.com/Data_Com_Basics/RS232_standard.html>.

[Wichert 2000] A.M.Wichert, Associative Computation, Ph.D. thesis, Ulm, Germany, Univeristy of Ulm, July 2000.

[Wooldrige 1995] M.Wooldrige, N.Jennings, Intelligent Agents: Theory and practice. Knowledge Engineering Review, Cambridge, England, Cambridge University Press, 1995, Vol. 10, No 2, pp. 115 – 152.

Author's Information

Agris Nikitenko - Riga Technical University, Department of System Theory, Meza street 1 k.4, Riga LV-1048, Lavtia; e-mail: agris@cs.rtu.lv

CLINICAL DECISION SUPPORT SYSTEM SONARES

Svetlana Cojocaru, Constantin Gaindric

Abstract:. A decision support system SonaRes destined to guide and help the ultrasound operators is proposed and compared with the existing ones. The system is based on rules and images and can be used as a second opinion in the process of ultrasound examination.

Keywords: decision support systems, knowledge acquisition, image processing, ultrasound examination.

Introduction

We will describe a decision support system SonaRes, destined to support ultrasound diagnostics. The system plays a consultative role and offers to users its variants of diagnosis. System’s solutions are motivated by presenting method(s) of its obtaining and the corresponding images that help to understand its reasons.

The problem of assuring an adequate medical assistance to the population depends both on training and qualification levels of medical personnel and on performance of the used diagnostic equipment. Nowadays it is impossible to offer medical services, even at the most modest level, without using of medical equipment, apparatuses, devices and technical complexes. Applications of medical equipment are very diverse and include diagnostics, treatment, supervision, compensation of a lesion or a handicap, etc.

There is no doubt that technical assistance in medicine is as important as in pharmaceutical assistance. The technical assistance in medical examinations is impossible without strong qualification of staff that exploits this equipment and interprets the obtained information. The quality of medical services also directly depends on how correctly and efficiently the medical diagnostic and treatment/recuperation equipment is used.

Ultrasound equipment is much cheaper than the MRI - Magnetic Resonance Imaging, CT – Computer Tomography, digital radiography etc, and its rational and efficient use could fill in many gaps in medical diagnosis. Technical progress of the last years in the field of ultrasound diagnostics allowed these methods to come to a leading position among imagistic procedures.

Echography has proved to be one of the most usable and beneficial paraclinic investigations as it is non-invasive and extremely efficient, has a great accuracy in its area of application, and is executed easily by a well-trained specialist. Despite the enthusiasm it meets, the echography has its real limitations like any other procedure.

They are expressed sometimes through false-negative or false-positive images; sometimes these limitations are imposed by an examining physician's ability to obtain qualitative images or to interpret them.

In order to obtain quickly correct information on the specific case based just on images and their descriptions, it is important to create a unified system that will allow storing of the images and their annotations. Special techniques should be developed to annotate images. This collection of images and annotations will help ultrasound technicians to justify their final conclusions.

The primary use of the system might be as a 'second opinion' in difficult cases and in emergency; it does not replace physician who interprets echograms. Thus, SonaRes is destined to improve health care by providing a highly efficient diagnostics tool. The tool is well-suited to needs and current state of the medical equipment in hospitals and clinics.

A comparison with existing systems

Presently, there are only a few decision support systems for ultrasound diagnostics in the world, even though the attempts to elaborate systems of the kind have been started since 1970's. We can refer to publications in the mentioned field [Diez, 1997-Cabinet].

Some of the most known systems created for ultrasound investigation we can mention are DIAVAL [Diez, 1997] designated for echocardiography, ProtoISIS [Anderson, 1994, Kahn] (ultrasound and computerized tomography), SonoConsult [Huettig, 2004] et al.

Till now two basic approaches were applied to development of computer assisted ultrasound diagnostic systems:

a) Systems based on image analysis and classification.

Such systems pursue the purpose to make the decision on the basis of comparison of the initial images with those from a database (DB). The comparison helps to classify the available image according to classification existing in a database and/or give the user an opportunity to define a degree of its similarity with images available in a database. In the case of detection of similarity to any image (precedent) the decision based on already known decision for the existing precedent is given out.

b) Systems based on rules.

Such systems pursue the purpose to make the decision on the basis of the description obtained from the user and the data available in system and rules. More often such systems serve for information or training purpose with or without an additional diagnostic component.

In ProtoISIS the classification of images is made on the basis of 4 sets consisting from 25 precedents each. The probability of correct classification makes 72-84%.

Better results have been received in the system described in [Huo, 2001 - Drukker, 2004] concerning the domain of computerized detection and classification of cancer on breast ultrasound. A two-stages computerized method has been developed: the detection stage and classification stage. At the first (detection) stage the suspicious regions in ultrasound images are detected and subsequently distinguished among different lesion types. After the detection stage all candidate lesions are classified by a Bayesian neural net, based on computer-extracted lesion features. Two separate tasks are performed and evaluated at the classification stage. The first classification task is the distinction between all actual lesions and false-positive detection; the second classification task is the distinction between actual cancer and all other detected lesion candidates (including false-positive detection's).

First stage, gives the performance values of 94% and 91% for training and testing data sets respectively. Second stage, based on candidates lesion classification, gives the performance values of 87% and 81% for training and testing data sets.

SonoConsult is a knowledge-based system which uses simple and complex rules to make a required decision, promotes completeness and carefulness of input of patient's state, and thus helps to minimize an opportunity of reaching the erroneous judgment.

The following table contains the main features to be incorporated in the developing system and points those features which reside in the existing systems.

	SonoConsult	Bayesian Network	ProtoISIS	LookInside	tCaUD	Syngo US WS	IMAGE-IT	ULTRA 32(64)	Makhaon PACS	Roentgenprom WS	SonaRes
Use of both the images and their descriptions	_	X	–	X	–	X	X	X	X	X	X
The interactive interface for knowledge acquisition	–	–	–	–	–	–	–	–	–	–	X
The intelligent interface	–	–	–	–	–	X	X	X	X	X	X
Expertise reporting	X	–	–	–	X	X	X	X	X	X	X
Explanation of the decision	–	–	–	–	–	X	X	X	–	–	X
Possibility of adding to knowledge base on the basis of precedents	X	X	X	X	–	–	X	X	–	–	X
Examination of the organs interaction	–	–	–	–	–	–	–	–	–	–	X
Image processing	–	–	X	–	–	X	X	X	X	X	X
The standardized descriptions and decisions	X	–	X	X	X	X	X	X	X	X	X
Possibility to use the system in automated learning	X	X	X	–	–	X	X	X	X	X	X
Treating of patient's state in dynamics	–	–	–	–	X	X	X	X	X	X	_

One can compare ten best and well-known existing systems using the data from the table:

- SonoConsult [Huettig, 2004], which is an expert system for structured and case-adequate documentation of sonographic findings with an additional diagnostic component;
- Bayesian Network [Haddawy, 1994] is a technique for reasoning under uncertainty, currently is being developed for application to medical decision making;
- ProtoISIS [Anderson, 1994, Kahn] is a decision support system, based on comparison of initial researched image with the images from the images' database for the purpose of similarity determination;
- "LookInside" [capabilities] is designed to operate with database of patients amenable to ultrasound examination;
- "The Cabinet of Ultrasound Diagnosis" (tCaUD) [cabinet] is also designed to operate with database of patients amenable to ultrasound examination.
- Syngo US Workplace [Syngo] is developed by Siemens Medical Solutions. It offers tools for archive, reporting and image processing.
- IMAGE-iT [IMAGE-iT] is developed by ASHVA TECHNOLOGIES and presents a software oriented to image processing, storage, reporting.
- ULTRA 32 (ULTRA 64) [ULTRA] is developed by SONULTRA CORPORATION. It is a system to support the reports construction, acting directly from data obtained by ultrasonographic scanners.
- Makhaon PACS [Makhaon] is a workstation supporting DICOM standard, contains a reporting module based on patterns international pathologies classifier.
- Roentgenprom [roentgenprom] is a system to support the investigation process. It offers some patterns for organs description according to a standard methodology.

From the table one can determine good characteristics and weak points of analyzed systems. We can see that no one system is taking into account interaction between organs. The system SonoConsult, which shows some performance results during its utilization, and is very appreciated by the specialists-physicians, has the most similar aim.

Main goal and system structure

Our goal is to develop an approach which includes interaction between organs and uses current and precedent similar images in decision making process. Special attention is paid to ergonomic user interface, which is generated dynamically by system according to the DB content and is adaptable to preferences and objectives (of investigation type) of the physician-echographist.

We will offer to specialist, even without wide experience, an access to a resource where the process of ultrasound examination is detailed and formalized and includes an enormous amount of useful information on anatomy, ultrasonic semeiology, differential diagnostics as well as condensed presentation of the main nosologic entities that should appear in the physician's mind at the moment of examination of each organ.

The system SonaRes helps the specialist in ultrasonic analysis to draw the conclusion more correctly, especially, in emergency cases or in unspecific clinic/paraclinic cases, which do not seem to be included in any classical presentation; in cases where they obtained ultrasonic semeiology can provide a correct diagnosis without complicated and, often difficult of access, medical investigations.

SonaRes offers to a user a second opinion with necessary explanations and images that are similar to the examined case. Images can be processed and problem zone, if it is necessary for the user-physicist, can be marked out.

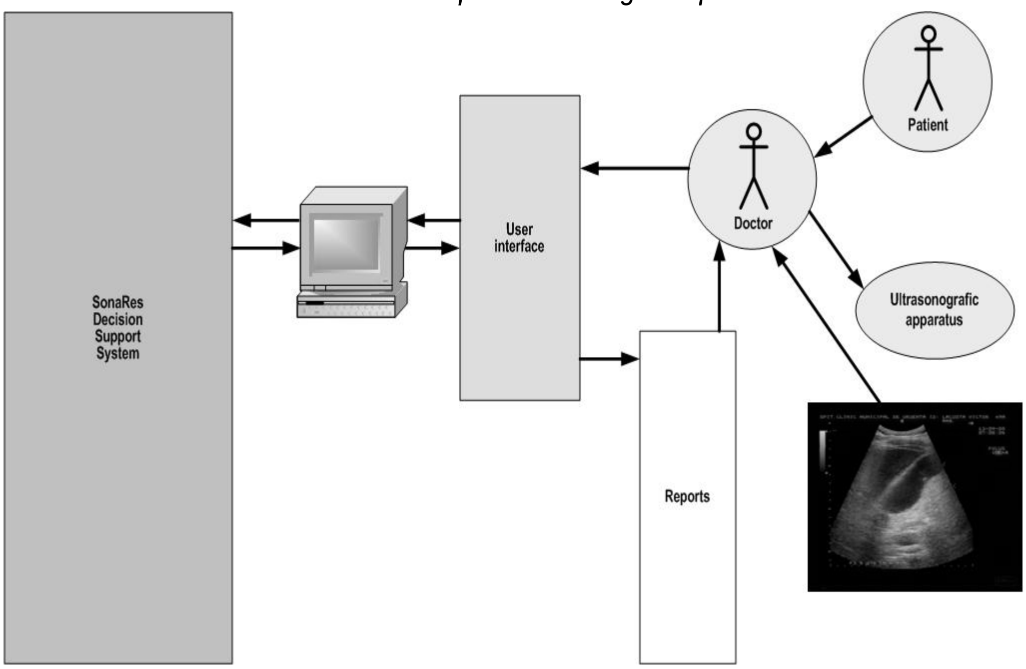
The main components of the system are the following:

- Knowledge acquisition
- Examination support
- Unified database (knowledge, images, annotations etc.)
- Image processing algorithms
- Reports generator

In order to develop these components we are elaborating and adapting:

- formalized descriptions of the abdominal organs, pathologies, anomalies;
- formalized descriptions of the ultrasound investigations methodology;
- unified, standardized disease descriptions;
- knowledge acquisitions methods based on ultrasound investigations characteristics;
- a diagnostics validation tool;
- a database model for medical images, their annotations and fuzzy information storage;
- images clusterization and quick database searching algorithms;
- an ergonomic, dynamically generated and user friendly interface;
- reports' prototypes and their generator.

The scheme of patient investigation process



At the first stage we deal with abdominal zone investigation. The investigation process of this zone is especially difficult (more organs with additional interactions, higher level of confusion, etc.). We have approved our technique on gall bladder and extend it on other organs.

Knowledge structure modeling

As models of knowledge representation in the medicine domain a model based on rules or a semantic network usually are chosen. In both cases the problem is reduced to [Secrieru, 2007, Popcova, 2006] :

- determination of objects, concepts and their attributes which are used in the given problem area;
- definition of links between concepts;
- determination of metaconcepts and detailed elaboration of concepts;
- construction of the knowledge pyramid, being scale of metaconcepts ranks, rising on which means the deepening in understanding and increasing the level of metaconcepts generalization.
- validation of rules representing knowledge.

Knowledge validation

A set of trees representing decisions rules are constructed based on data tree structure and using knowledge about organ’s pathologies and anomalies. These trees contain all the necessary factors which can help to produce a conclusion. The validation goal is to evaluate the knowledge base correctness and completeness. One can obtain this information by testing the existing rules. The testing is performed by physicians, in order to do this work in more efficient way a validation tool was elaborated, which permits to simulate the examination process and evaluate the obtained conclusions.

The tool offers the possibility to obtain various conclusions modifying the attributes values. During the validation process the current session containing all the selected attributes values is saved. Thus to simulate a new examination which differs from the previous one just by one attribute it is sufficient to modify only one value. Also it is easy to establish the list of rules containing a specific combination of attribute-value [Jantuan, 2007].

Image processing

Non-using of images during decision-making process can lead to the loss of the valuable information and does not correspond to daily practice of the doctor. Therefore we use a collection of annotated images stored in Image DB, which can be used as illustrations in similar disease cases. Ultrasound images from DB will be preprocessed in interesting zones to facilitate their clusterization (this is necessary for quick and relevant search).

Image preprocessing will be performed using existing algorithms (image processing algorithms, adapted to ultrasound diagnostic methodology), as well as by using the new ones (based on heuristics, fuzzy, pattern-oriented filters etc.). The following image processing methods [Dzung, 1998 - Popcova, 2004] were applied: statistical treatment, noise reduction, contrast adjustment, borders and organ contours determination etc. and their combination to obtain more efficient result.

Examination support

The proposed method of acquisition (by means of expert shell) and storage of expert knowledge in Unified DB permits to effectuate a quick search of necessary information in two directions or modes. The first direction is from the concrete case description to determination of pathology and/or an anomaly; and the second one – from formulation of a hypothesis to its confirming or denying.

Following the first direction the user gives the necessary information describing a concrete case, and the system tries to determine if it is a pathology and/or an anomaly. To exclude at the early stage the input of inconsistent, erroneous or excessive information, this direction is followed step-by-step. If at any step the system can determine, on the basis of the entered information, a pathology and/or an anomaly, it informs the user.

Following the second direction, the user forms a hypothesis about presence in the concrete case of pathology and/or an anomaly. Then the system by means of additional questions tries to confirm or to deny this hypothesis.

Realization of both modes within the framework of unified support system of ultrasonic investigation process corresponds to the daily work of physicians. The first operating direction satisfies the requirements of the detailed patient examination; and the second direction corresponds to a simplified one, when it is necessary to confirm or to deny any diagnosis

A convenient dialog with user-physician (due to dynamic intelligent interface which includes a standardized explanation of the decision proposed by system) involving images in decision making process (based on visualization and comparison of ultrasound examined image with similar images from Image DB) will permit to create a comfortable environment for physicians and will help him to prepare a standardized report, containing the examination results and, if necessary, the recommendations for additional investigation.

Conclusion

The proposed system does not intend to replace completely the physician; it just will offer him a second opinion. In all cases user will receive all rules and judgments on the basis of which the decision was made. If the user doesn't agree with the decision, proposed by the system, his opinion will be sent to expert group for examination.

Acknowledgments

This work is supported by the STCU, project ref. 4035.

Bibliography

[Diez, 1997] F.J.Diez, J.Mira, E.Iturraalde, S.Zubilaga. DIAVAL, A bayesian expert system for echocardiography. – Artif. Intell. Med. n.10, 1997, pp.59-73.

[Anderson, 1994] Kahn C.E. Jr., Anderson G..M.Case-based reasoning and imaging procedure selection. Investigative Radiology 1994; 29:643-647., http://www.mcw.edu/midas/old-papers.html)

[Kahn] Charles E.Kahn, Jr. Clinical Trial and Evaluation of a Prototype Case-Based System for Planning Medical Imaging Work-up Strategies. www.mcw.edu/midas/papers/ISIS.CBR94.ps

[Huettig, 2004] M.Huettig, G.Buscher, T.Menzel, W.Scheppach, F.Puppe, H.P.Buscher. A Diagnostic Expert System for Structured Reports, Quality Assessment, and Training of Residents in Sonography. Medizinische Klinik, 2004, 99, nr.3, p117-122.

[Huo, 2001] Huo Z, Giger M.L., Vyborny C.J.: Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis. IEEE Transactions on Medical Imaging 20: 1285-1292, 2001.

[Horsch, 2004] Horsch K, Giger ML, Vyborny CJ, Luz V: Performance of CAD in the interpretation of lesions on breast sonography. Academic Radiology 11: 272-280, 2004.

[Drukker, 2004] Drukker K, Giger M.L., Vyborny CJ, Mendelson EB: Computerized detection and classification on cancer on breast ultrasound. Academic Radiology 11: 526-535, 2004.

[Haddawy, 1994] Haddawy P, Kahn CE Jr, Butarbutar M. Bayesian network model for radiological diagnosis. Medical Physics 1994; 21:1185-1192.

[capabilities] http://www.lins.ru/capabilities.shtml

[cabinet] http://cabinet.fromru.com/

[Syngo]www.medical.siemens.com/webapp/wcs/stores/servlet/ProductDisplay~q_catalogId~e_-11~a_catTree~e_100001~a_langId~e_-11~a_level~e_0~a_productId~e_172044~a_storeId~e_10001.htm

[IMAGE-iT] www.ashvatech.com/imageit.htm

[ULTRA] www.sonultra.com/products/reporting.htm

[Makhaon] www.makhaon.com/index.php?lng=ru&p=products

[roentgenprom] www.roentgenprom.ru/products/software/arm/

[Secrieru, 2007] Iu.Secrieru, O.Popcova, S.Puiu, D.Sologub. Knowledge Structure Modeling in Ultrasound Investigation Domain. Bit+. The 7th International Conference "Information Technologies 2007", 24 -28 April, 2007, Chisinau, Republic of Moldova (to appear).

[Popcova, 2006] O.Popcova, I. Secrieru, D.Sologub, E.Jantuan, V.Papanaga. Decision Support System for Ultrasound Diagnostics // Proc. of the First International Conference of Yong Scientists Computer Science&Engeenering (CSE 2006) – Lviv, 2006 – p.30-31.

[Jantuan, 2007] E.Jantuan. Knowledge validation tools. Bit+. The 7th International Conference "Information Technologies 2007", 24 -28 April, 2007, Chisinau, Republic of Moldova (to appear).

[Dzung, 1998] Palm Dzung L., Xu Chenyang, Prince Jerry L. A survey of current methods in medical image segmentation // Technical report. – Johns Hopkins University, Baltimore, 1998. http://www.csem.duke.edu/seminars/segmenation.pdf

[Jardim, 2002] Jardim Sandra Vilas Boas, Figueiredo Mário A. T. Automatic analysis of fetal echographic images // Proc. Portuguese Conf. on Pattern Recognition. – Aveiro: RecPad, 2002.

[Dinggang, 2003] Dinggang Shen, Yiqiang Zhan, Davatzikos C. Segmentation of prostate boundaries from ultrasound images using statistical shape model // IEEE Transactions on Image Processing. – 2003. – vol. 22, № 4. – P. 539-551.

[Popcova, 2004] O.Popcova, S.Cojocaru, C.Gaindric. Image Processing in Ultrasound Diagnostic System. 8th International Symposium on Automatic Control and Computer Science, Iasi, Romania, October 22-23, 2004.

Authors' Information

Svetlana Cojocaru – Dr.hab., Institute of Mathematics and Computer Science, Academy of Sciences of Moldova, Academiei str., 5, Chisinau, Republic of Moldova, e-mail: sveta@math.md

Constantin Gaindric - Dr.hab., Prof.,corr.member Academy of Sciences of Moldova, Institute of Mathematics and Computer Science, Academy of Sciences of Moldova, Academiei str., 5, Chisinau, Republic of Moldova, e-mail: gaindric@math.md

AUTOMATED RESPONSE TO QUERY SYSTEM

Vladimir Lovitskii, Michael Thrasher, David Traynor

Abstract: SMS (Short Message Service) is now a hugely popular and a very powerful business communication technology for mobile phones. In order to respond correctly to a free form factual question given a large collection of texts, one needs to understand the question at a level that allows determining some of constraints the question imposes on a possible answer. These constraints may include a semantic classification of the sought after answer and may even suggest using different strategies when looking for and verifying a candidate answer. In this paper we focus on various attempts to overcome the major contradiction: the technical limitations of the SMS standard, and the huge number of found information for a possible answer.

Keywords: mobile text messages, text message analysis and question-answering system

ACM Classification Keywords: I.2 Artificial intelligence: I.2.7 Natural Language Processing: Text analysis.

Introduction

This paper represents results of our further research in the text data mining and the natural language processing areas [1-5] restricted by mobile's text-based SMS messaging. SMS is now a very powerful business communication technology widely used from small businesses and home users through to large corporations, governmental and non-governmental organisations. However, many of these users have little or no experience of SMS technology and only a vague idea of how successful they could be when properly harnessing the power of SMS communication.

SMS text messaging is currently being evaluated in many different areas:

- **Mobile Banking.** Mobile Banking Services including: Account Balance Enquiry, Account Statement Enquiries, Cheque Status Enquiry, Cheque Book Requests, Fund Transfer between Accounts Credit/Debit Alerts, Minimum Balance Alerts, Bill Payment Alerts, Bill Payment, Recent Transaction History Requests, Information Requests i.e. Interest Rates/Exchange Rates. (e.g. the HSBC's SMS Enquiry Service [6]). Although these services are appearing they do raise specific issues concerning security, especially when the data is extremely sensitive and confidential. 2ergo have recognised this and have developed and patented a secure messaging protocol [7] to address such issues through encryption technology.

- **Mobile Government** services [8]. For example, Ireland's tax collection agency, Office of the Revenue Commissioners, as of 2005 now receives at least as many enquiries by text message as by telephone. The SMS enquiry service allows citizens to claim tax credits and request a number of tax forms and information leaflets by sending text messages from their mobile phones.
- **Mobile Learning** (SMS in education [9]). Students find SMS messages useful and are keen to use it in a number of ways: announcements, assessment marks, assessment feedback, appointments, revision tips and also quick easy access to some library service. Thus, it is believed that SMS in education can facilitate collaboration, strengthen community spirit, assist timely completion of coursework and other assessments, and ultimately reduce the attrition rate amongst students. 2ergo have released a new product targeted at the education market [7]. It is aimed at anti-truancy, anti-bullying and group messaging to provide school administrators with the means to manage and control these very important issues.
- **Mobile Airport** services [10]. The SMS service gives passengers fast responses to queries about the airport when they are possibly in transit and scheduled to connect with a flight. It reduces the need for people to phone the airport information number, which can often be engaged and thereby increasing anxiety levels amongst passengers, or go to the information desk inside the terminal.

Given the recent experience the list of areas where mobile messaging will prove useful is likely to be extended. What these applications have in common is that they represent a **W**ell-**D**efined **A**pplication **D**omain (WDAD). For each of them a 5-digit short code can be predefined, e.g. when the message is sent to the 64222 short code WDAD "Edinburgh Airport" will be selected automatically and therefore the answer to the question *"Where is the nearest hotel?"* will be synthesized relatively simply by the Question-Answering System (QAS).

The central question to be addressed by this paper, however, is how to provide the response to a question if the application domain is unknown i.e. **U**nknown **A**pplication **D**omain (UAD). The attempt to find a response for UAD has been undertaken already by Google [11]. The goal of **the Google SMS service** is to provide the large existing base of users with access to the types of information they are most likely to need when mobile. Users simply send their query as a text message and receive results in the reply. Simple conventions are used to express queries for phone book listings, dictionary definitions, product prices, etc. Google touts this service as ***"Just text. No links. No web pages"*** [10]. Simply the answers one is looking to find. Examples of queries are *"pizza 21228"* to find pizza places located near to the University of Maryland, Baltimore County (UMBC), or *"george bush, washington dc"* to find the address and phone number of the US president or *"Price ipod 20gb"* to get a list of prices (and sellers) for an ipod. The **M**obile **Q**uery (MQ) is sent to the 5-digit US short code 46645, which corresponds to Google on most phones. One could, of course, leverage existing mobile technology through the WAP browser on mobile phones as an alternative to using SMS. For example, if you want to find out the dictionary definition of the word *"spring"* simply enter message ***"define:spring"*** to search in your mobile "Google" on WAP enabled phones. The following response will be displayed: *"the season of growth; "the emerging buds were a sure sign of spring"; "he will hold office until the spring of next year""*. If the same query is repeated but on a personal computer rather more information regarding the definition will be displayed (the most compact part of those information is shown below).

Definitions of **spring** on the Web:

- the season of growth; "the emerging buds were a sure sign of spring"; "he will hold office until the spring of next year";
- a natural flow of ground water;
- jump: move forward by leaps and bounds; "The horse bounded across the meadow"; "The child leapt across the puddle"; "Can you jump over the fence?";
- form: develop into a distinctive entity; "our plans began to take shape";
- a metal elastic device that returns to its shape or position when pushed or pulled or pressed; "the spring was broken";

- leap: a light, self-propelled movement upwards or forwards;
- bounce: spring back; spring away from an impact; "The rubber ball bounced"; "These particles do not resile but they unite after they collide";
- give: the elasticity of something that can be stretched and returns to its original length;
- develop suddenly; "The tire sprang a leak";
- a point at which water issues forth;
- produce or disclose suddenly or unexpectedly; "He sprang these news on me just as I was leaving".

Despite the quite impressive results using MQ there remain some significant problems:

- **Quality not quantity**. The best conformity between the returned response and the MQ is more important than the quantity of information found. How can we define the **quality of response**? When a user asks for the definition *of spring* Google sent the first line of those definitions to the user's mobile. Is it what the user anticipated, or would it be more important for him to know that *"spring is a metal elastic device…"*? Using a single SMS restricts the answer to 160 characters, and depending upon the type of phone used the display area is very likely to be much less than this. Although *"no answer is worse than a bad answer"*. A disappointing response could deter the customer from using the SMS service again.
- **No dialogue**. Let us distinguish between QAS and Mobile QAS (MQAS). The principal difference is that QAS allows for a clarification dialogue [3,4] with the user to provide the best possible response through interaction but MQAS does not. As for MQAS only a single short user MQ must be used to create the response.
- **Ungrammatical MQ.** The dream about MQ: *"Be precise and informative about your problem. Write in clear, grammatical, correctly-spelt language"* – is far from being realised. As a rule an MQ will be ungrammatical, not because users are illiterate but because <u>most users are lazy</u> i.e. they want to achieve the desired result by using the minimum effort. For example, they do not want to use upper case to type MQ *"george bush, washington dc"* [9] or use dots to separate *"d"* and *"c"*. Other examples of typical grammatical errors are: typing *"its"* instead of *"it's"*, *"lose"* instead of *"loose"*, or *"discrete"* instead of *"discreet"*. The fact is that *MQ simply cannot be spelt, punctuated, and capitalised correctly* but the main requirement for MQAS is - to handle non-standard or poorly formed/structured (but, nevertheless, **meaningful**) user's MQ.
- **Fuzzy MQ.** The users are generally unable to describe completely and unambiguously what it is they are looking for [4]. For example, the MQ *"Who won the 2006 World Cup?"* is absolutely clear from the user's point of view because it goes without saying that *football* is implicit in the question. Fuzzy MQ requires **fuzzy searching**. Fuzzy response searching is a technique for finding a reply that approximately matches the search MQ.
- **Facts Ambiguity.** The problems of answering the MQ depends not only on an incomplete and/or ambiguous MQ but also on *facts ambiguity*, especially when MQAS, as a result of searching information on the internet, expands its local KB by adding appropriate facts. It is not easy to recognise ***semantic***, ***lexical*** or ***structural*** ambiguity of facts. For example, the fact *"Stolen painting found by tree"* has structural ambiguity and can be interpreted as: (1) *A tree found a stolen painting*; (2) *A person found a stolen painting near a tree*. The MQ *"Who stole a painting?"* requires some effort from MQAS to reply correctly (we will discuss this in the *Mobile Queries Classification* session).

The meaning of an MQ depends not only on the things it describes, explicitly and implicitly, but also on both aspects of its causality: what caused it to be said and what result is intended by saying it. In other words, the meaning of an MQ depends not only on the MQ itself, but on **who** (age, sex, nationality) sends it and **when**, **where**, **why** and to **whom** it is sent. Such information helps to understand *what* the user meant in his/her MQ. In this paper we shall discuss **why** the user sent the MQ and the determination of some of the constraints the MQ imposes on a possible answer.

Why was the Mobile Query Sent?

Why do people not ask questions? It is very important to know why, especially if you want to learn and improve performance in an organisation. There are, as minimum, some 10 reasons why people never ask questions [12]. But here we want to find out the answer to the opposite question: *why do people ask questions?* While it is possible to question without thinking, it is impossible to think without questioning. A good question might just provide the means for overcoming a particular obstacle and achieving a stated goal or objective. There are some 12 reasons identified for asking questions [13]. This list is based on three assumptions: (1) a genuine question is one to which the answer is not known; (2) curiosity is the driver; and (3) the questions are intended in a constructive manner, i.e. to intimidate, dominate, demonstrate how smart one is, or to prove that one is right. These reasons are:

1. Gather information.
2. Build and maintain relationships, participate in communication.
3. Learn, teach, and reflect.
4. Think clearly, critically, and strategically.
5. Challenge assumptions.
6. Solve problems and make decisions.
7. Clarify and confirm information heard.
8. Negotiate and resolve conflicts.
9. Set and accomplish goals.
10. Take charge and focus attention on yourself.
11. Create and innovate – open new possibilities, provoke thought in others.
12. Catalyse productive and accountable thinking, conversation, and action.

The reason behind asking questions depends on the situation when the question has been asked. Let us enumerate some obvious situations:

- Ask the audience a verbal question, e.g. a teacher questioning pupils.
- Ask a person a verbal question but in front of an audience, e.g. at a scientific conference.
- Ask a person a written question but in front of an audience, e.g. at a scientific conference.
- Ask a person an anonymous written question but in front of an audience, e.g. at a company's meeting.
- Ask a "face to face" question.
- Ask a "face to face" question on the phone.
- Ask an audience a written question, e.g. through the internet.
- Ask a personal written question, e.g. through email.
- Ask an SMS question. It is completely different in comparison to the previous point because of size, convenience and money.
- Ask the Artificial Intelligent Question-Answering System (AIQAS) a verbal question.
- **Ask the AIQAS a MQ.** This is exactly our case and just **two** reasons (1 and 6) among the 12 listed earlier are appropriate for this particular situation.

On the one hand two reasons to send an MQ have been selected, on the other hand there should be some external, implicit reason that causes the user to send an MQ. Let us call such a reason a **meta-reason**.

Meta-reason for Mobile Query Sending

It is very important to underline that in this paper we consider the precise situation when an MQ is sent to the AIQAS but not to another person. Initially, we decided to find out about the SMS-activities during the year i.e. it to discover whether and how SMS-activities depend on seasons, weekdays, holidays, events etc. (see Figure 1). Red (top) charts represent quantity of daily SMS. The set of events is represented as **events hierarchical structure**, nodes of which, in essence, are clusters and have links to corresponding nodes of **clusters structure**. Suppose an MQ was sent on Saturday, 08.07.06. MQAS will extract a list of events taking place on or around this day from the KB (see Figure 1) and create a list of potentially active clusters: for example, *football, tennis, motor racing, music,* and *show*. Thus, MQAS has selected five WDAD ahead of the MQ analysis. Of course, for MQ: *"What tablets can help me with headache?"* these WDAD would not be used but they might help to disambiguate MQ related to one of these enumerated clusters.

Figure 1. SMS-activities and Events

Mobile Queries Classification

As part of its MQ processing component, the MQAS first attempt to classify each MQ by their type and represented as a pair: **<Question Type>** ↦ **<Answer Type>** [15]**.** Only two kinds of question are allowed for MQ: (1) **Specific question** (*what, who, why, how, where, when*, etc.) and (2) **Yes-No question.** For example, instead of asking: *"Name all directors of 2ergo"*, the specific question *Who* should be used: *"Who are directors of 2ergo?"*. Examples of MQ classification are shown below:

- **What ↦ (money, definition, name, etc): "What was the monetary value of the Nobel Peace Price in 1989?"**
- **What ↦ (person, organisation): "What costume designer decided that M. Jackson should wear only one glove?"**
- **What ↦ (date): "In what year did Ireland elect its first woman president?"**
- **What ↦ (location): "What is the capital of the Ukraine?"**
- **Which ↦ (person, organisation): "Which former Klu Klux Klan member won an elected office in the U.S.?"**
- **Which ↦ (date): "In which year was New Zealand excluded from the ANZUS alliance?"**
- **Which ↦ (location): "Which city has the oldest relationship as sister-city with Los Angeles?"**
- **Who ↦ (person, organisation): "Who is the author of the book "The Iron Lady: A Biography of Margaret Thatcher"?"**
- **Whom ↦ (person, organisation): "Whom did Italy beat in the final of 2006 Football World Cup?"**
- **Where ↦ (location): "Where is Kharkov?"**
- **When ↦ (date): "When did the Jurassic Period end?"**
- **Why ↦ (reason): "Why do people shake hands to show friendliness?"**
- **How ↦ (action): "How did Socrates die?"**

- **How many ↦ (number): "How many James Bond novels are there?"**
- **How much ↦ (money, price, time, etc.): "How much pizza do Americans eat in a day?"**
- **How long ↦ (time, distance): "How long does it take to travel from Plymouth to Rawtenstall?"**

As for a Yes-No MQ, in general, it is better to avoid asking a Yes-No MQ unless one wants to receive a yes or no answers. For example, instead of asking *"Are there any tablets to relieve my headache?"* it is better to ask *"What tablets can relieve my headache?"*. A Yes-No MQ highlights a very important problem of communication with both a "human" QAS or with AIQAS. By default, when a user sends an MQ he/she does not doubt that the recipient is a human being. That is why the Yes-No MQ like *"Does anyone know ... ?", "Can someone ... ?", "Is it possible to ... ?"*, etc is very popular. In a call centre the recipients appear to find such MQ annoying [16] — and are likely to return logically impeccable but dismissive answers like *"Yes, you can be helped"* in response to the MQ *"Can you help me to find an address of 2ergo?"*. As for our MQAS, on the one hand, we decided to give the impression that our system is not artificial, on the other hand, MQAS is always "well behaved" (i.e. it is simply impossible to annoy it – since it is a machine) and that is why MQAS would reply to user *"Yes, you can be helped but it is better to ask a specific question"*.

The MQ classifier is critical during the stage of an MQ's type recognition (see Figure 2). A list of supplementary key words for each *question word* allows the specification of the type of expected response more precisely (see an example of *When* description in **Knowledge Base Structure** session).

Derivation of implicit information for MQ Response

The primary focus of MQAS is for the creation of a coherent, understandable answer that is responsive to the originally posed MQ. The factually explicit MQ e.g. *"What is the Taj Mahal?"* does not require a great deal of effort to create the answer if the MQAS KB contains the fact: *"Taj Mahal is a beautiful mausoleum at Agra built by the Mogul emperor Shah Jahan (completed in 1649) in memory of his favourite wife"*. But the problems are, firstly, it is impossible beforehand to classify MQ as **explicit** or **implicit**, and secondly, in reality most of the MQs are implicit. For example, if the MQAS KB contains the fact: *"Aleksandr Sergeyevich Pushkin is a Russian poet (1799 -1837)"*. The obvious implicit MQs related to this fact are *"When was Pushkin born?"* or *"When did Pushkin die?"*. Below we discuss the possible ways for the derivation of implicit information from the KB.

In this paper we continue to apply a ***psycholinguistic approach*** to natural language (NL) processing [1]. The only system truly capable of adequate understanding of an MQ is human. What is more, children seem to use NL effortlessly in spite of *not knowing the grammar*. Parsing was taught in school as an algorithmic task. For derivation of implicit information a ***natural inference engine*** based on *human reasoning* is used. *Human reasoning* might be described using fuzzy attributes: *approximate, common sense, default, enumerative, evidential, hypothetical, inexact, integrating, plausible, procedural, taxonomic.* In other words, instead of logically making some conclusion a human would very often ***justify*** the decision based on maximum argumentation about problem solving <u>within his/her knowledge</u>. At present it is understandable that the real knowledge base (KB) is ***incomplete***, ***inconsistent, inaccurate*** and ***open***. These facts do not necessarily permit the use of a traditional logical approach. For example, assume that the following two statements are in the KB: (1) *"A student likes to read a detective story"*, (2) *"A student does not like to read mathematical books"* and the MQ simply states *"Does a student like to read?"*.

The derivation of implicit information is not always based on the logical operations:

- From the conditional judgement *"If 'A' then 'B'"* it is not possible to draw the conclusion *"If not 'A' then not 'B'"*. As for the sentence, *"If you smoke you will fall ill"* the conclusion *"If you do not smoke you will not fall ill"* is not true since non-smokers also become ill. Representation of the conditional judgements stated in the linguistic form and the conclusions drawn from them depend widely on their contents.
- The examinees were given the sentences: *"Acorns always grow on an oak"* and *"Acorns grow on a tree"*, they concluded on this basis: *"This tree is an oak"* [14]. The potential fallacy of this conclusion becomes

evident when considering another example apparently having the same structure. From the sentences, *"All football-players are good runners"* and *"Peter is a good runner,"* one can hardly draw the conclusion that *"Peter is a football-player"*. In the first instance, at first glance, the rule of inadmissibility of the general use judgement was violated. It is not possible to conclude that *"All B's are A's"* from the judgement *"All A's are B's"*. It should be specified that the simple transformation of the general judgement is not always incorrect but only in those cases when the participants of the judgement have dissimilar size. In the case of the identical sizes the simple transformation of a general judgement is possible. For instance, the judgement *"All squares are equilateral rectangles"* transforms into the judgement *"All equilateral rectangles are squares"* as the "participants": *"square"* and *"equilateral rectangle"* are of equal size. Really, there are no rectangles other than equilateral ones being squares. As to the judgement *"All football-players are good runners"* this is not the case. Here the "participants" sizes: *"football-players"* and *"good runners"* are different and the simple transformation of this judgement is inadmissible.

Cognitive transformations as the natural inference source for the derivation of implicit information:

- Taking into account the cognitive transformations based on the semantic relations featuring the properties of **symmetry**, **transitivity** and **additivity** allows MQAS to answer a Yes-No type MQ. For example:
  (1) **Fact:** *"'A' is a colleague of 'B'". MQ: "Is 'B' a colleague of 'A'?".*
  (2) **Fact:** *"'A' buys a doll".* **MQ:** *"Does 'A' buy a toy?".*
  (3) **Fact:** *"'A' is a teacher of 'B'".* **MQ:** *"Is 'B' a pupil of 'A'?".*
- It is important to process facts on the things belonging to transformation. For example, from the fact: *"Mary has given a book to John after New Year"* three different MQ might be answered:
  (1) *"Did John have this book before New Year?"*;
  (2) *"Does John has this book now?"*;
  (3) *"Does Mary have it now?".*

The derivation of implicit information based on *human reasoning* is a part of Reply Search Engine (RSE). The central question to be addressed by any QAS is how the storage of information is organised in KB and we now turn to consider this.

Knowledge Base Structure

In the general case, under the *knowledge base structure* one should understand the regularity of data distribution in memory assuring the storage of various links between separate elements of stored information. At every moment KB deals only with relatively *small fragments* of the external world. So, the corresponding structures are needed to integrate these fragments separated in time into the integral picture. The structures obtained as a result of integration should contain more information than had been used for its creation. The organisation of knowledge storage should make allowance for such features of the human memory as [17]:

- associativity;
- ability to reflect similar features for different objects and different features for similar objects;
- hierarchical and heterarchical organisation of information. The idea of heterarchical approach correlates naturally with the human ability to use all kinds of information in the process of natural language understanding. As Quillain remarks "*... a full association structure... forms simply large, very complicated net of nodes and unidirectional memory links between them... The predetermined hierarchy of "super-" and "subclasses" is absent; every word is a "patriarch" in its own hierarchy if some process of search initiates with it. Analogously every word is in different places within hierarchies of large diversity of wordy concepts if the search process starts with them*" [18, p.5];
- associative relations weight variable;
- representation of the environment statistical properties;

- independence of the knowledge extraction time from the volume of knowledge being stored in this memory;
- the knowledge cannot realistically be regarded as a static resource, to be accumulated and stored within a system. It is a formative, *self-organising* character, with the ability to change the organisation within which it is held.

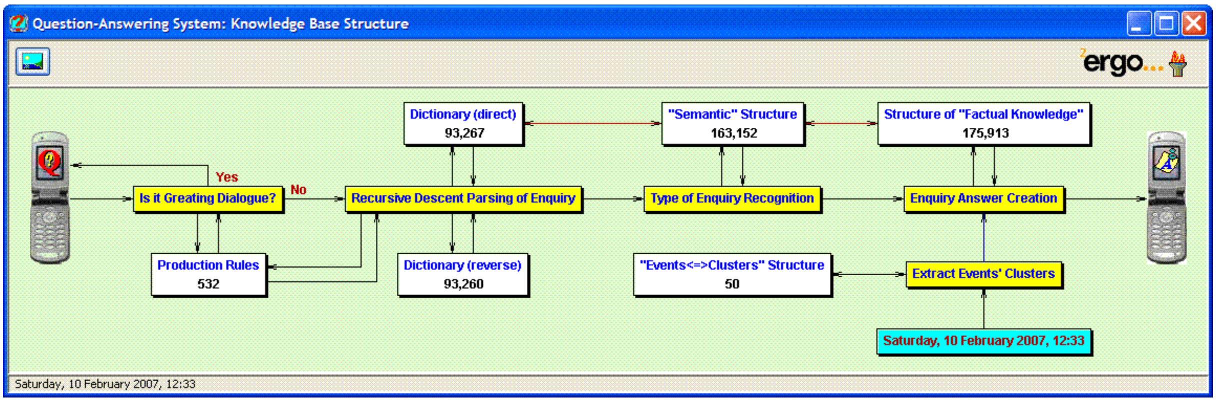


Figure 2. Knowledge Base Structure

The attempt to take into account these features of human memory in MQAS KB has been undertaken. The structure of KB is shown in Figure 2. MQAS's consists of six different structures:

- **Direct** and **Reverse Dictionary** (DD and RD respectively) are represented by an L-Tree structure. DD is an initial structure which provides the recognition of *new* words, the normalisation of *known* words and the direct links with the corresponding nodes of **Semantic** structure and **Factual Knowledge** structure. RD together with DD is used to correct wrong words, spelling them automatically.
- **Semantics** structure (or Sequential-Simultaneous Structure (SSS)) provides the *sequential* and *simultaneous* analysis of string information and handling of *new* and/or *known* sentences or combinations of words. SSS is a special combination of hierarchical and network structures. Each of the elements in SSS are associated with several other elements logically including itself or included by it. With the successive presentation of facts as a sequence of words the strongest relation in it is the relation between the nearest neighbouring words. *"Their succeeding one after another presents evidently an important condition of structuring"* [19, p.231]. The ***sequential*** part of SSS provides ***hierarchical*** organisation of information and the ***simultaneous*** - ***heterarchical*** organisation of information
- **Events-Clusters** Structure (ECS)
- **Factual Knowledge** Structure (FKS)
- **Production Rules** Structure (PRS) [4]. New class of PR **Question** has been added. The example of PR in format: **<Class>** $\mapsto$ **<Antecedent>** $\Rightarrow$ **<Consequent>** is shown below.

Question $\mapsto$ **When** $\Rightarrow$ **0:**born,start,begin,commence,come,become;**1:**live,interval,period;

2:die,end,finish,stop;**3:**occur,happen,find.

Let us come back to a fact: *"Aleksandr Sergeyevich Pushkin is a Russian poet (1799 -1837)"* in FKS. Assume the MQ *"When was Pushkin born?"* has been asked. In the consequent of PR *Question* $\mapsto$ *When* subclass **0** has been found, which means that the left-hand number from the interval should be selected as a birth year. First of all MQAS tries to find in FKS the direct fact for an answer like *"Pushkin was born in 1799"*. If such a fact does not exist MQAS starts to analyse facts with date's interval and then from the appropriate fact select left number as a birth year i.e. *"In 1799"*.It is worth noting that, MQAS always try to minimise the length of the response because of the display constraints of the mobile device and the SMS message.

At the end of this session it is important to emphasise:

- Local KB of any MQAS cannot be **complete**, in other words, we shall never be able to establish **information completeness** of KB.
- MQAS should be **self-learning** i.e. when MQAS cannot find a suitable response in the local KB it should search using the **Internet** and then not only send the reply to the questioner but also **automatically extend the local KB**. The searching of information across the Internet represents the greatest problem because the result may identify a huge set of documents among which the appropriate response will need to be found, We shall discuss this problem in the next paper.

Mobile Query Parsing

The main requirement for MQAS is to reply to any (even non-standard or poorly formed but, nevertheless, **meaningful**) user's MQ. But it is not easy to find the answer even for an ideal MQ because for an artificial intelligent system like MQAS the power of natural language to describe the same events but in quite different ways is a great problem. For example, the primitive action: *"take by theft"* might be described as: *"hook", "snitch", "thieve", "cop", "knock off", or "glom".* The main purpose of MQ processing is to understand *what was meant* rather than *what was said*. The mechanism of query parsing is very simple: *"eliminating the unnecessary until only the necessary remain"* and has been discussed elsewhere [2]. Here we just remind ourselves of the main steps involved in MQ processing. This MQAS takes the MQ as a character sequence, locates the MQ boundaries, and converts the original MQ to a *skeleton*. Such conversion will require several steps:

- MQ related synonyms' dictionary creation to equate synonymous words and phrases, such as "NLP" and "Natural Language Processing";
- Irregular verb normalisation. Once the word has been identified then it should be changed back to its simplest form for efficient word recognition. For example, *writes, writing, wrote, written* will be changed to *write* and the corresponding attributes of the original form will be saved;
- Noisy (non-searchable) word elimination;
- Plural to singular conversion.

The skeleton of the MQ is matched against all relevant data in SSS to find the appropriate links to FKS. The result of such searching is shown on Figure 3.

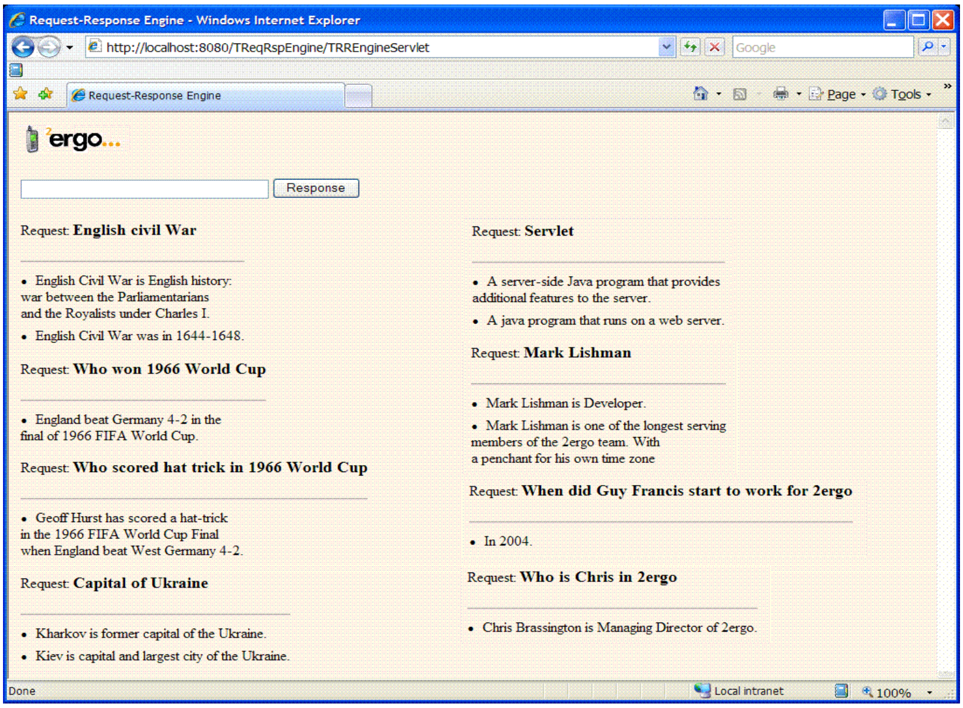


Figure 3. The examples of MQAS responses

Conclusion

The possible solutions of how to provide the response to a question if the application domain is unknown have been considered. In the result of that, the MQAS effectively places information directly into the hands of any users - eliminating the need for technical support specialists continually to address *ad hoc* requests from end users.

The MQAS are addressed from both scientific and industrial perspectives. Whether MQAS is searching the local KB, or the worldwide web, MQAS understands the relationships between words, enabling it to extract all key concepts and automatically build a semantic index organised in a problem-solution format. Because MQAS extracts and organises the content, the user receives specific and relevant answers to his/her MQ — not a list of documents.

Bibliography

[1] G.Coles, T.Coles, V.A.Lovitskii, “Natural Interface Language”, Proc. of the VIII-th International Conference on Knowledge-Dialogue-Solution: KDS-99, Kacivelli (Ukraine), 104 -109, 1999.

[2] T.Coles, V.A.Lovitskii, “Text Searching and Mining”, J. of Artificial Intelligence, National Academy of Sciences of Ukraine, Vol 3, 488-496, 2000.

[3] D.Burns, R.Fallon, P.Lewis, V.Lovitskii, S.Owen, “Verbal Dialogue Versus Written Dialogue”, International Journal “Information Theories and Application”, Vol 12(4), 369-377, 2005.

[4] G.Francis, M.Lishman, V.Lovitskii, D.Traynor, “Instantaneous Database Access”, Proc. of the XII-th International Conference on Knowledge-Dialogue-Solution: KDS-2006, Varna (Bulgaria), 314 - 322, 2006.

[5] K.Braithwaite, M.Lishman, V.Lovitskii, D.Traynor, “Distinctive Features of Mobile Messages Processing”, Proc. of the XII-th International Conference on Knowledge-Dialogue-Solution: KDS-2006, Varna (Bulgaria),322 - 329, 2006.

[6] www.hsbc.com.sg/1/2/personal/services/sms-from-hsbc

[7] www.2ergo.com (2safeguard™, 2NC™)

[8] www.revenue.ie

[9] www.smseducation.org

[10] www.edinburghairport.com

[11] Google SMS query service

[12] www.Think6Results.com

[13] M.Adams, Inquiry Institute, www.inquiryinstitute.com

[14] R.J.Harris, G.E.Monaco, "Psychology of Pragmatic Implication: Information Processing between the Lines", J. Exp. Psychol. General, 107, 1-22, 1978.

[15] D.Moldovan, S.Harabagiu, et al. “Lasso: A Tool for Surfing the Answer Net.”, TREC-8 Draft Proceedings, NIST, 65-73, 1999.

[16] www.82ask.com.

[17] V.A.Lovitskii and K.Wittamore, “AMONIL: Achievement of Multiple Objectives using a Natural Interface Language”, Proc. of the International Joint Conference on Knowledge-Dialogue-Solution: KDS-97, Yalta (Ukraine), 270-281, 1997.

[18] M.R.Quillian, "Word concepts: A theory and simulation of some basic semantic capabilities", C.I.P. working paper 79, Carnegie Inst. of Technology, Pittsburgh, 1965.

[19] J.Hoffmann, “Das Aktive Gedachtnis. Psycologische Experimente und Theorien zur Menschlichen Gedachtnistatigkeit”, VEB Deutscher Vergal der Wissenschaften, Berlin, 1982.

Authors' Information

Vladimir Lovitskii – 2 Ergo Ltd, St. Mary’s Chambers, Haslingden Road, Rawtenstall, Lancashire, BB4 6QX, UK, vladimir@2ergo.com

Michael Thrasher – University of Plymouth, Plymouth, Devon, PL4 6DX, UK, mthrasher@plymouth.ac.uk

David Traynor – 2 Ergo Ltd, St. Mary’s Chambers, Haslingden Road, Rawtenstall, Lancashire, BB4 6QX, UK, david.traynor@2ergo.com

CONCEPTUAL FOUNDATIONS OF CONSTRUCTION OF THE MODELS AND PROCEDURES FOR PREDICTION OF THE AVALANCHE-DANGEROUS SITUATIONS INITIATION

Alexander Kuzemin, Vyacheslav Lyashenko

Abstract: The conceptual foundations of the models and procedures for prediction of the avalanche-dangerous situations initiation are considered. The interpretation model for analysis of the avalanche-dangerous situations initiation based on the definition of probabilities of correspondence of studied parameters to the probabilistic distributions of avalanche-dangerous or avalanche non-dangerous situations is offered. The possibility to apply such a model to the real data is considered. The main approaches to the use of multiple representations for the avalanche dangerous situations initiation analysis are generalized.

Key words: avalanches, probability, set, situation, model, avalanche danger, avalanche slide.

Introduction

Creation of an adequate system for analyzing forming conditions and taking measures for improvements in prevention of such situations becomes an important problem in connection with the increasing number of arising emergency situations of natural and technogene nature, increase in their scales. Transition to the management methods based on the analysis and estimation of the risk of the danger quantitative characteristics in the case of initiation of hazards for population and environment is possible as one of such measures.

Avalanche-dangerous regions occupy 6% of the land area. But despite this the problem of such phenomena investigations is rather important and urgent as analogous phenomena can become the cause of people's death and considerable destructions [1, 2].

As a whole, one of the directions, i.e. the risk of emergency situations management and, in particular, situations initiated by the snow-slip, should be considered constant monitoring and building of interpretation models for prediction of such situations initiation. Hereinafter, such models form the basis for the system of decision–making support; this is favorable to development recommendations on modern performance of maintenance measures directed to natural calamities prevention.

Among the most essential and important problems in the given aspect one should note substantiation of the utility to use the corresponding mathematical apparatus intended both for investigation of the avalanche dangerous situations development dynamics and for development of methods for estimation of the potential avalanche cells, prediction of avalanches volumes and descent frequency. This concerns the fact that every avalanche can be regarded as a unique phenomenon of nature with its specific peculiarities. At the same time despite its uniqueness it is possible to single out the climatic conditions variations characteristic ranges which are prerequisites to prediction of the feasible avalanche descent. Eventually, the totality of these two factors defines the presence those approaches to prediction and warning of avalanches descent, at present these approaches are used in geoinformation systems (GIS) which make it possible to accumulate continuously meteorological information, carry out various calculations, reveal regularities and realize spatial tie of the obtained results [3, 4].

Available methods of the avalanche dangerous situations initiation prediction and shortages of the traditional approaches application. Considering methods and models of avalanches descent prediction the method images of similarity and regression analysis are singled out the most often [3, 5]. At the same time, there is no doubt that the foundation of avalanche dangerous situations initiation prediction consists in the procedure of the preliminary analysis of such events. In this case, as a rule, the solution of the formulated problem is based on the statistical analysis methods. In particular, the approaches of such analysis make it possible to substantiate the most significant system of the facts which is expedient to use in the avalanche-dangerous situations prediction procedures. Such approaches found their development in the predictions of snow avalanches descent based on application of the nearest neighbors' method or through the application of the regression equations [5, 6]. But results of the prediction obtained with such methods are not always applicable and demonstrate a number of shortages: they require significant computational resources; they don't embrace existing variety of causes resulting in avalanches formation. The shortage is also the impossibility to define the degree of the avalanche hazard, number and dimensions of avalanches [7].

The data of nomograms which in a general case extend the interconnection of such indices as temperature, value of snow cover and precipitations are also used for estimation of the avalanches descent probability. But in spite of this the remaining non-predictive nature of the avalanche dangerous situation doesn't always allow to prevent negative consequences of emergencies caused by their descent. This is associated with that the available procedures of the avalanche dangerous situation initiation prediction are not sufficiently precise. At the same time the severity of the problem and variety of the ways to solve it motivate the necessity to search alternative methods which can give more argued answers.

Thus, as the main aim of the given investigation one can single out consideration of the approaches, alternative to the available methods, to the avalanche-dangerous situations initiation prediction. First of all, in this case the consideration of the conceptual foundations of the models and procedures of such a prediction is significant in our opinion.

The probabilistic aspects of the avalanche climate initiation medium analysis. Analysis of different characteristics of the avalanche climate initiation medium makes generally the foundation of the avalanches descent prediction. Among such characteristics the most abundant ones are: the air temperature, humidity, atmospheric precipitations volume, wind velocity, angle of the slope of surface (descent angle) where the avalanche descent is possible. In general, the variation dynamics of both individual of the above characteristics the avalanche climate initiation and their totality can, with some probability, characterize initiation either avalanche-dangerous or avalanche non-dangerous event. As this takes place, a feasible range of the studied avalanche climate initiation characteristics variations describes a definite region of avalanche-dangerous and avalanche-non-dangerous situation. In the conceptual plan the essence of the probabilistic aspect of the avalanche climate initiation analysis can be reduced to the definition of the probability to assign some point as the considered medium current characteristics either to the region of the avalanche initiation or to the initiation of avalanche non-dangerous situation. Otherwise the given approach can be treated as a correspondence of the current characteristics of the avalanche climate initiation medium; parameters of these characteristics define some region using probabilistic distribution of avalanche dangerous or avalanche non-dangerous situations preceding this. Consequently, it is possible to speak about so-called probable conformity of the researched characteristics of the avalanche-dangerous climate environment to probabilistic distribution of the avalanche-dangerous or avalanche non-dangerous situations.

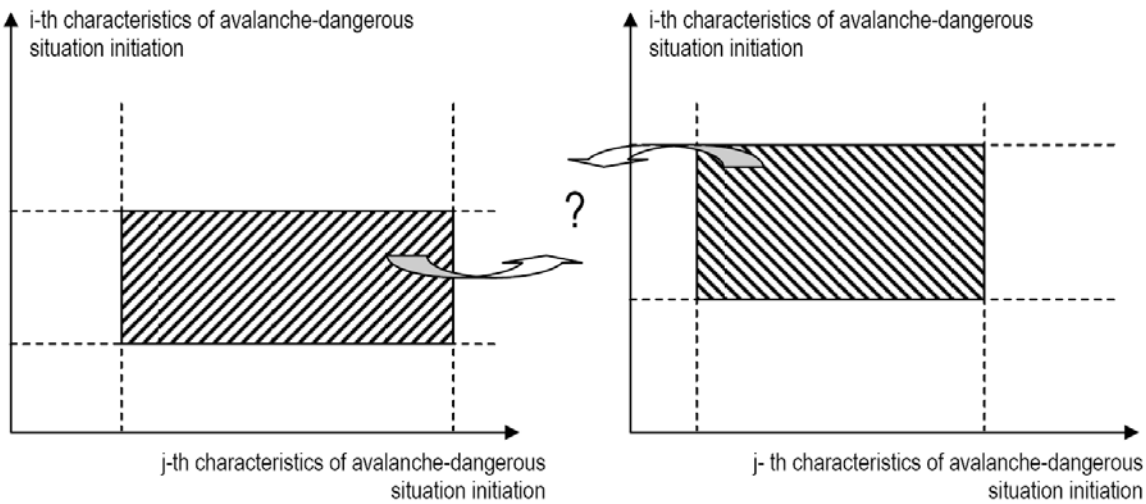


Fig.1 On the explanation of the probabilistic aspect of the avalanche climate initiation medium analysis

In particular, procedure of such analysis can be considered proceeding from the pairwise analysis of various characteristics of the avalanche climate initiation. The advisability of such a transition is related to the fact that at the stage of the preliminary analysis it is possible to omit less significant factors of impact on the avalanche-dangerous situation initiation. Thus, the base element of the analysis procedure being considered is estimation of the probability of the avalanche climate initiation current parameters to fall within the regions typical and atypical for the avalanche climate initiation. The given regions can be presented in the plane in the form of the rectangle; its metric values correspond to definite parameters of variation of the avalanche- dangerous and avalanche non-dangerous situations initiation medium characteristics (Fig.1) [8].

To explain the offered aspect and substantiate its significance let us analyze the probabilistic aspects of some characteristics of the avalanche climate initiation medium using the real data of the avalanches descent in ITAGAR Chychkan region Kyrgyzstan Republic within 2001–2006, obtained in the frameworks of carrying out joint scientific and research work. The essence of such analysis is reduced to the estimation of the feasible assignment of the avalanche non-dangerous situations to the avalanche-dangerous ones and vice versa in terms of different characteristics of their initiation.

First and foremost, it should be noted that the considered characteristics of the avalanche climate initiation medium follow the normal distribution law. This makes it possible to use this law for estimation of the corresponding probabilities. The corresponding probabilities with regard to the avalanche dangerous and avalanche non-dangerous situations are presented in Table 1.

Table 1.

Probabilities of correspondence of the current parameters of avalanche climate initiation medium to the avalanche dangerous and avalanche non-dangerous situations

Characteristics of the avalanche climate initiation medium analysis	Feasibility of correspondence
under condition of considering the avalanche dangerous situation and avalanche dangerous current parameters	
air temperature – wind velocity	0,854
air temperature – wind velocity	0,823
wind velocity – precipitations quantity	0,707
precipitations quantity – descent angle	0,809
under condition of considering the avalanche-dangerous situation and avalanche- non-dangerous current parameters	
air temperature – wind velocity	0,488
humidity – precipitations quantity	0,582
wind velocity – precipitations quantity	0,317
precipitations quantity – descent angle	0,341
under condition of considering the avalanche-non-dangerous situation and avalanche- non-dangerous current parameters	
air temperature – wind velocity	0,798
humidity – wind velocity	0,878
wind velocity – precipitations quantity	0,866
precipitations quantity – descent angle	0,939
under condition of considering the avalanche-non-dangerous situation and avalanche- dangerous current parameters	
air temperature – wind velocity	0,555
humidity – wind velocity	0,482
wind velocity – precipitations quantity	0,403
precipitations quantity – descent angle	0,591

As can be seen from the data in Table 1 the assumptions made above are reasonably justified i.e. the probability of correspondence of the like situations and parameters is essentially significant. This allows making generalization even for estimation of probable initiation of the avalanche dangerous as a whole. To do this one should consider:

- either generalization of the obtained probabilities reasoning from the significance of different groups of characteristics of the avalanche climate initiation characteristics analysis in the assumption that the probabilities of correspondence can be considered as conditional probabilities of the concrete situations analysis;
- or a separate group of characteristics of the avalanche climate initiation medium analysis based on the greatest/least values of the correspondence probabilities.

Microsituations classes of avalanche danger initiation. Representation of the avalanches descent factors variety in the form of a set of microsituations helps to increase reliability of their analysis and prediction. Every such a microsituation corresponds to a definite combination of factors of the avalanche initiation environment. At the same time, such representation makes it possible to brake up the whole set of causes affecting the avalanches initiation into two subclasses. One of subclasses characterizes a set of microsituations reflecting the avalanche initiation and the other subclass is typical for non-avalanche situation as a whole. Then the emergency avalanche situations risks management can be presented as a generalized description of the system with the help of a totality of different microsituations. Based on such interpretation the logical rules of the analyzed data set generalization for further their subdivision into classes of avalanche dangerous and non-avalanche dangerous situations:

$$\text{Avalanche dangerous} = \left(\{F_L^L(X)\}/\{F_N^N(X)\}\right) \cup \left(\{F_N^N(X)\}/\{F_L^L(X)\} \cup\right)$$
$$\cup \left(\{F_L^N(X)\}/\{F_N^L(X)\}\right) \cup \left(\{F_N^L(X)\}/\{F_L^N(X)\}\right) \cup \left(\{F_L^N(X)\}/\{F_N^N(X)\}\right),$$
$$\text{Non avalanche dangerous} = \left(\{F_N^L(X)\}/\{F_L^L(X)\}\right) \cup \left(\{F_L^L(X)\} \cap \{F_N^N(X)\}\right) \cup$$
$$\cup \left(\{F_L^N(X)\} \cap \{F_N^L(X)\}\right) \cup \left(\{F_N^N(X)\}/\{F_L^L(X)\}\right),$$

where $F_L^L(X)$ ($F_L^N(X)$), $F_N^L(X)$ ($F_N^N(X)$)– probability function of referring avalanche dangerous (non-avalanche dangerous) microsituation to the avalanche dangerous (non-avalanche dangerous) class, respectively, on the set of factors of the avalanche danger initiation X [9].

In this case performance of the preliminary statistical analysis of the avalanche dangerous situations initiations warning makes it possible to define discriminant functions of such situations prediction. Results of estimating the probability of referring of the investigated data to the classes describing avalanche dangerous or non-avalanche dangerous situations of the avalanches descent obtained with such functions also allow to single out different types of situations on the whole set of the analyzed data of the avalanche climate environments. Respectively, this may be the situation of the avalanche initiation (Ω_L) and non-avalanche dangerous situation (Ω_N).

In its turn the considered situations Ω comprise a definite set of microsituations $\Omega = \{\omega_i\},\ i = \overline{i,n}$, each of them corresponds to a definite group of the considered types of data of the avalanche initiation environment and representing on the one hand, certain probability of avalanche situation initiation, and on the other hand — characterizing the probability of non-initiation of the avalanche dangerous situation.

To put it differently, avalanche dangerous situations Ω_L represent the combination of microsituations $\Omega_L = \bigcup_i \Omega_{L_i}$ uniform by the type of data, and avalanche non-dangerous situations Ω_N represent the combination. Each specific microsituation represents, within certain limits, the probability of initiation either of the avalanche dangerous, or avalanche non-dangerous situation as a whole as regards a definite type (group) of data with their different values which characterize the considered situation.

Consequently, as a whole the results of estimating the probability of referring the investigated data to the classes describing avalanche dangerous or avalanche non-dangerous situations of the avalanches descent can be regarded as a basis for transition to formation of the corresponding microsituation systems, the data relative to which the knowledge base of crises situation management information-analytical system is created.

As this takes place, in any case the obtained probabilities of correspondence to the classes of the avalanche dangerous and avalanche non-dangerous situations initiations for some interval of time allow passing to the consideration and analysis of the avalanche dangerous situations initiations. As the boundaries of such sets one can consider the corresponding boundaries of the distribution polygon built using the correspondence probabilities in some temporal interval. Analysis of such sets makes it possible to formalize the totality of the avalanche dangerous and avalanche non-dangerous situations and to build the procedure of the corresponding prediction.

At the same time presentation of the avalanche danger factors in the form of the microsituations classes allowed to get an objective correspondence between the probabilistic estimates of the avalanches descent and the avalanche danger scale degrees; eventually this makes it possible to correct time of the prediction system response to the possible avalanche descent. Essence of such estimate consists in construction and analysis based on the theory of fuzzy sets, corresponding functions of prediction time correction $\mu(X)$ (Fig. 2) [10].

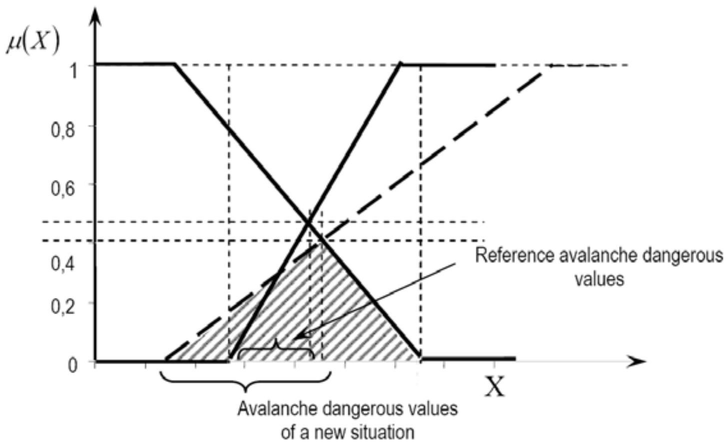


Fig. 2. Methods of the fuzzy sets theory as the basis for correction of the prediction system time response to the possible avalanche descent

As a whole, the essence of such description reduces to construction of the fuzzy model for estimating temporal characteristics of the avalanche dangerous situations. Thus, for example, it is possible to suppose that in the case of the avalanche non-dangerous situation analysis the time till the assumed avalanche descent will be the more the greater is the probability of referring current parameters of analysis of the avalanche climate initiation environment to such situation. Respectively, in the case of insufficient probabilities of referring of the avalanche climate initiation environment analysis current parameters to such situation can testify to insignificant reserve of time till the avalanche descent moment. When considering the avalanche dangerous situations the corresponding characteristics are the opposite. This allows to introduce into definition the functions of the avalanche descent time fuzzy set; their generalization makes it possible to predict the avalanche descent time.

It is possible to pass to the distribution functions in the estimate of the avalanche descent time on the basis of the corresponding probabilities analysis of the fuzzy set available data separation into avalanche dangerous and avalanche non-dangerous situations.

Thus, the model construction generalized scheme and construction of the procedure for prediction of the avalanche-dangerous situations initiation is reduced to:

- sequential obtaining of the probabilistic characteristics of the avalanche climate initiation medium;
- construction of the corresponding sets of subdivision into avalanche-dangerous and avalanche-non-dangerous situations;
- analysis of the avalanche descent initiation time using fuzzy models of its interpretation.

Conclusion

This work presents the general concept of building models and procedures for prediction of the avalanche-dangerous situations based on the probabilistic and multiple approaches to its interpretation. Such an approach allows, first and foremost, taking into account the range of variations of the factors acting on the avalanche-dangerous situations initiation, to build adequate procedures for their prediction. Moreover, the essential characteristics of the problem under consideration which is confirmed, in particular, by the feasibility of the probabilistic model of the general approach based on the real data.

Bibliography

1. Federici P., Tamburin A.i, Luzi G., Rott H, Schaffhauser A., Strozzi T., Bernardini G. Galahad: An EU Project for the Remote Monitoring of Glaciers, Avalanches and Landslides // IDRC. – Davos, 2006. – Vol. 2. – P.P. 177–180.
2. Voitkovsky K.F. Science of Avalanches. – M.: MSU, 1989. – 158 p.
3. Durand Y., Brun E., Merindol L., Guyomarc'h, Lesaffre B., Martin E. A meteorological estimation of relevant parameters for snow models. Ann. Glaciol.,18, 1993. – P.P. 65–71.
4. Kuzemin A., Toroev A. Mobile Means of Control and Prediction of Avalanche Climate Using Information Conversion in Acoustic Range 291 // IDRC. – Davos, 2006. – Vol. 2. – P.P. 291–294.
5. Buser, O., Butler, M. and Good, W. 1987. Avalanche forecast by the nearest neighbors method // IAHS. – Publ. 162. – P.P. 557–569.
6. Izhboldina V.A. Aerosinoptical conditions of formation and descent of snow-storm avalanches in the Kola Peninsula// Collection of articles: Studies of snow and avalanches in the Khibini Mountains. – L.: Hidrometeoizdat, 1975. –P.P.51-63.
7. Fuhn P. An overview of avalanche forecasting models and methods. Oslo, NGI, Pub. N 203, 1998, P.P. 19–27.
8. Kuzemin A., Lyashenko V. Probabilistic and multivariate aspects of construction of the models and procedures for prediction of the avalanche-dangerous situations initiation // Fifth International Conference «Information Research and Applications» i.TECH 2007, 26-30 June, Varna, Bulgaria – Sofia: FOI ITHEA – 2007. – Vol. 2. – P.P. 284–288.
9. Kuzemin A.Ya., Lyashenko V.V., Fastova D.V. Interpretation model for analyzing the environment the avalanche initiation climate // Data Recording, security and processing. – V. 9. – №1. – 2007. – P.P. 27–34.
10. Kuzemin A., Lyashenko V., Toroyev A., Klymov I. Developing an expert system for situational analysis of avalanche danger // Fifth International Conference «Information Research and Applications» i.TECH 2007, 26-30 June, Varna, Bulgaria – Sofia: FOI ITHEA – 2007. – Vol. 2. – P.P. 294–297.

Authors' Information

Kuzemin A.Ya. – Prof. of Information Department, Kharkov National University of Radio Electronics, Head of IMD, Ukraine, e-mail: kuzy@kture.kharkov.ua

Liashenko V.V. – senior scientific employee, Kharkov National University of Radio Electron, Ukraine

THE APPROACH TO DEVELOPMENT THE HUMAN RESOURCES INTELLECTUAL MANAGEMENT SYSTEM. MANAGEMENT PROCEDURES.

Roman Benger, Elena Antonova

Abstract: The Article suggests an approach to designing the Human Resources Intellectual Management System in order to increase Human Resources reliability, using the management methods known from the Theory of Management.. The Article examines the realization of the Subsystem of implementing management methods by the number of management procedures, executing the corresponding management method.

Keywords: The theory of unreliable elements, The knowledge system, The intelligent control, KDS.

Introduction

Management is a complex and multi-level process, covering all the spheres of human activity, and the most complex and unreliable elements of this process are people (human resources). In any managing system, the person's actions depend on numerous factors, and many actions are carried out spontaneously or thoughtlessly.

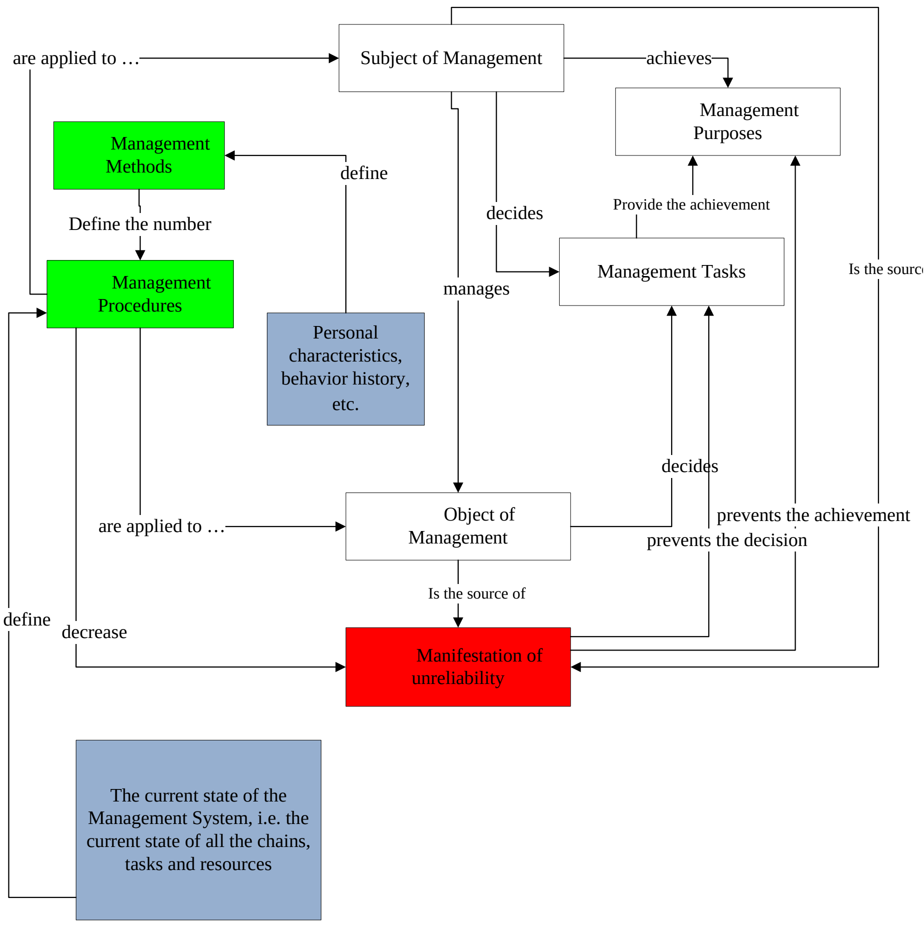
In order to control the people as the Management Process element, the Science of Management develops the methods of influence and motivation, basing on the number of human psychological peculiarities.

The informational management systems, existing now, do not fully consider all the aspects of the human resources as the “unreliable” element. To realize the system that considers all the aspects of unreliability, working out the Intellectual Management System of “unreliable elements” is needed.

The Article suggests an approach to designing the Human Resources Intellectual Management System in order to increase Human Resources reliability, using the management methods known from the Theory of Management. The Article examines the realization of the Subsystem of implementing management methods by the number of management procedures, executing the corresponding management method.

The Management System Model

Within the framework of the set task, the Management System is schematically presented on Scheme 1.



Scheme 1. The Management System Model

The Elements of the Model (according to Scheme 1)

The main element of the examined Management System Model are the Management Procedures, their interrelation and functioning provides the definite types of influence on the “unreliable elements” (human resources) in order to achieve the global purpose – increasing the human resources reliability (during their professional tasks solving). Let us examine the Elements of the Model in details:

Element: Object of Management

Description: the Object, at which the management influence is directed, in order to provide management system functioning; it provides the management tasks fulfillment, which results in achievement of the management purposes.

Element: Subject of Management

Description: provides the tasks planning, resources distribution, decision-making and control over the achievement of the set purposes.

Element: Management Tasks

Description: a number of actions that should be performed by the Subject or the Object, directed at achievement of the Management Purposes.

Element: Management Purposes

Description: Management Purpose is the predetermined state of the Management System.

Element: Manifestations of Unreliability

Description: a set of attributes of the Subject or the Object of Management, preventing from solving the Management Tasks, and consequently, from achieving the Management Purposes.

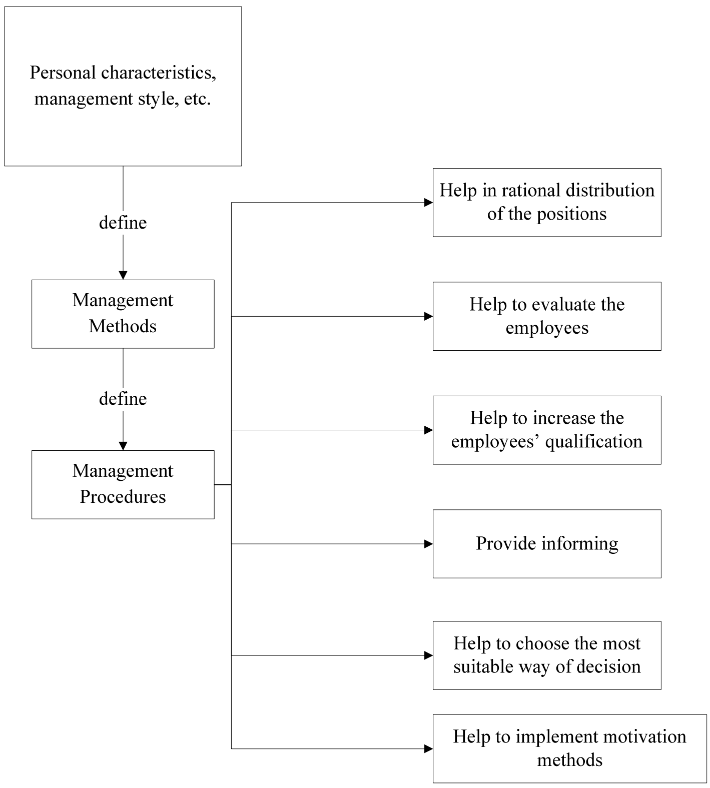
Element: Reliability Increasing Methods (Management Methods)

Description: Management Method is the definite way (style) of Management, basing on the number of characteristics of the Subject and Object of Management, i.e. it is the function, defined on the tuple set of the values of the Subject and Object of Management. Function range is the set of the finite subsets of the Management Procedures set.

Element: Management Procedures

Description: Management Procedure is the set of possible actions, directed at the Object or Subject, depending on the state of the Management System, which could prevent the Manifestation of Unreliability or minimize its negative result.

The main actions performed at the expense of Management Procedures, are shown on the Scheme 2.



Scheme 2. Management Tasks solved by the Management Procedures

Thus, the Management Procedures represent the management process as the management of the activity according to the following tasks: planning and execution of works; coordination of actions of the specialists, participating in the task solving; control over the executing the tasks of the professional activity; influence on the executors, etc.

A number of Management Methods is listed in the literature on management. Let us examine some set of the Management Methods:

1. Management Method on the basis of job descriptions requirements. According to job descriptions, the Management System participants' roles are defined, as well as their responsibilities, possibilities, rules of professional tasks solving. The job descriptions requirements, regulating the employees actions in accordance with the defined responsibilities in the collective, are the basis of assigning the human resources for sub-tasks solving.

2. Management Method on the basis of incentive and punitive measures. In order to stimulate the correct fulfillment of the tasks within the short terms, the incentive measures are applied to the Management System participants. In case of tasks non-fulfillment, the punitive measures are applied to the participants.

3. Management Method, in which the interests and psychological peculiarities of the person are considered. To fulfill the task, the Management System chooses such a participant that his interests coincide with the task subject, and his psychological peculiarities maximally meet the task requirements.

4. Management Method, in which the personal participation of employee in decision-making is considered. On the definite stage of work fulfillment, the Management System gives participant the possibility to choose one of the further ways of task solving, provided by the decisions plan. Thus, the participant of the Management System is given the “freedom” of choosing the further task solving, but within the limits of defined decisions plan.

5. Management Method, in which the record keeping and control over the employee is provided. At every stage of task fulfillment, the system arranges the time points of reference, in which the checking of task fulfillment by employee is provided – whether he meets the defined schedule or not. The advancement level charts are made for all participants of the decision.

6. Management Method, in which the informing of the system participants is provided. The Subject is reported about the current state of activities of the Object, the progress of task fulfillment, terms and advancement are reported.

7. Management Method on the basis of development and education. The set of Disciplines is organized in the Management System, each of them corresponds to some rating scale. Each task in the Management System is put in correspondence with the set of minimal rating values for each of the Disciplines. The employee, whose rating values for each of the Disciplines are not less than is required for the task, can be admitted to solving this task.

8. Management Method on the basis of correspondence of the tasks and employees. In the method the history of all employees’ activities is provided according to their performed tasks. When the new tasks appear, the employees who successfully solved the similar tasks in past, are admitted to solve those new tasks.

9. Management Method on the basis of past situations analysis. This method implies keeping all the information on each solved task – the state of the system at the moment of starting the task solving, the method of solving, solving result (successful/non-successful) and the state of the Management System at the moment of completing the task solving. This method helps to avoid the past mistakes and use the successful examples of tasks solving.

Within the framework of designing the Subsystem of Management Procedures usage, the procedures that would be used in the abovementioned methods, are defined.

Procedures set for the method on the basis of job descriptions:

- Roles distribution procedure:

  Description: This procedure provides the relations installation between the set of human resources and the set of roles, defined in accordance with the job descriptions and the type of the defined task. So, each of the Subjects and Objects of the management is assigned with the set of roles.

  Definition: (resources, tasks types, job descriptions) → ({} <objects, {} roles>)

Use Environment: Starting of the new task solving, failure in solving the current task (schedule exit, resources problems, re-defining of time terms)

- Positions distribution procedure:

Description: This procedure provides the relations installation between the set of human resources and the set of positions, defined in accordance with the job descriptions and the type of the defined task. So, each of the Subjects and Objects of the management is assigned with the set of positions.

Definition: (resources, tasks types, job descriptions) → ({} <objects, positions>)

Use Environment: Starting of the new task solving, failure in solving the current task (schedule exit, resources problems, re-defining of time terms)

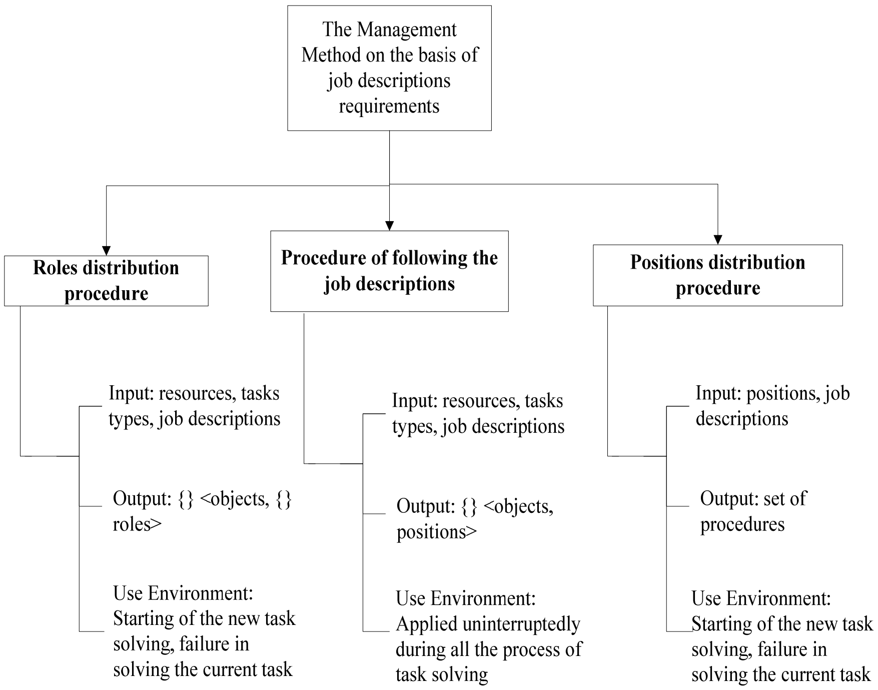
- Procedure of following the job descriptions:

Description: This procedure represents the set of management procedures, characterized by the definite set of job descriptions and office position.

Definition: (positions, job descriptions) → (set of procedures)

Use Environment: Applied uninterruptedly during all the process of task solving.

Thus, the Management Method on the basis of job descriptions requirements is defined by the procedures set shown on the Scheme 3.



Scheme 3. The Management Method on the basis of job descriptions requirements

Procedures set for the method on the basis of incentive and punitive measures.

- Procedure of sanctions definition:

Description: This procedure represents defining of relations between the effectiveness of solving the tasks of the given type and executed by the definite volume of resources, and the type of sanction, applied to the task executor.

Definition: (tasks types, resources) → ({} <results, sanctions>)

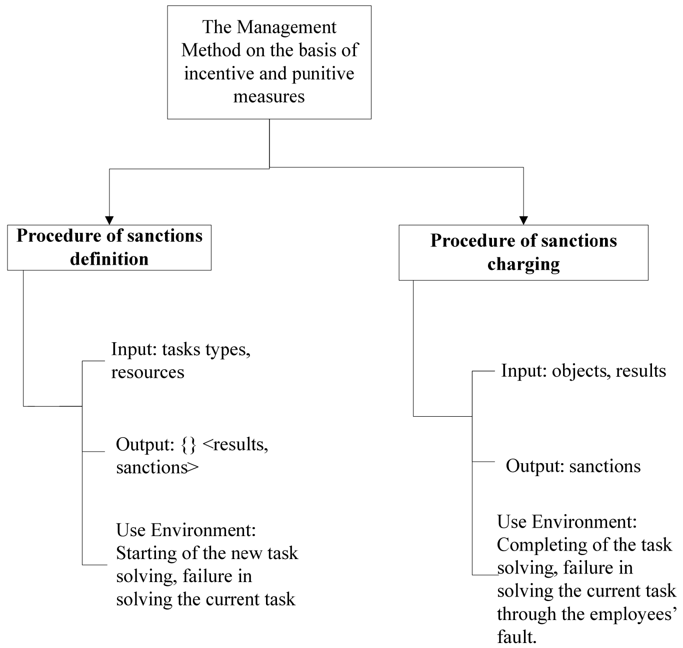
Use Environment: Starting of the new task solving, failure in solving the current task (schedule exit, resources problems, re-defining of time terms)

- Procedure of sanctions charging:

Description: This procedure provides sanctions charging to the Objects and Subjects of Management.

Definition: (objects, results) → (sanctions)

Use Environment: Completing of the task solving, failure in solving the current task through the employees' fault.



Scheme 4. The Management Method on the basis of incentive and punitive measures

Procedures set for the method, in which the interests and psychological peculiarities of the person are considered:

- Procedure of defining the psychological peculiarities for the tasks:

  Description: This procedure sets the relation between the type of the task and position during the task solving, and the set of necessary psychological peculiarities.

  Definition: (tasks types, positions) → ({}(psychological peculiarities))

  Use Environment: New type of the task or position appearance in the Management System.

- Procedure of appointing to the positions:

  Description: This procedure provides appointment of the definite employee with definite set of psychological peculiarities at the corresponding position within the task of defined type.

  Definition: (objects, {}psychological peculiarities, tasks types) → (positions)

  Use Environment: Starting of the new task solving, change in the psychological peculiarities of employee.

- Procedure of tracking the psychological peculiarities:

  Description: This procedure defines the psychological peculiarities of employee according to his behavior history.

  Definition: (objects, behavior history) → (<objects, {}psychological peculiarities>)

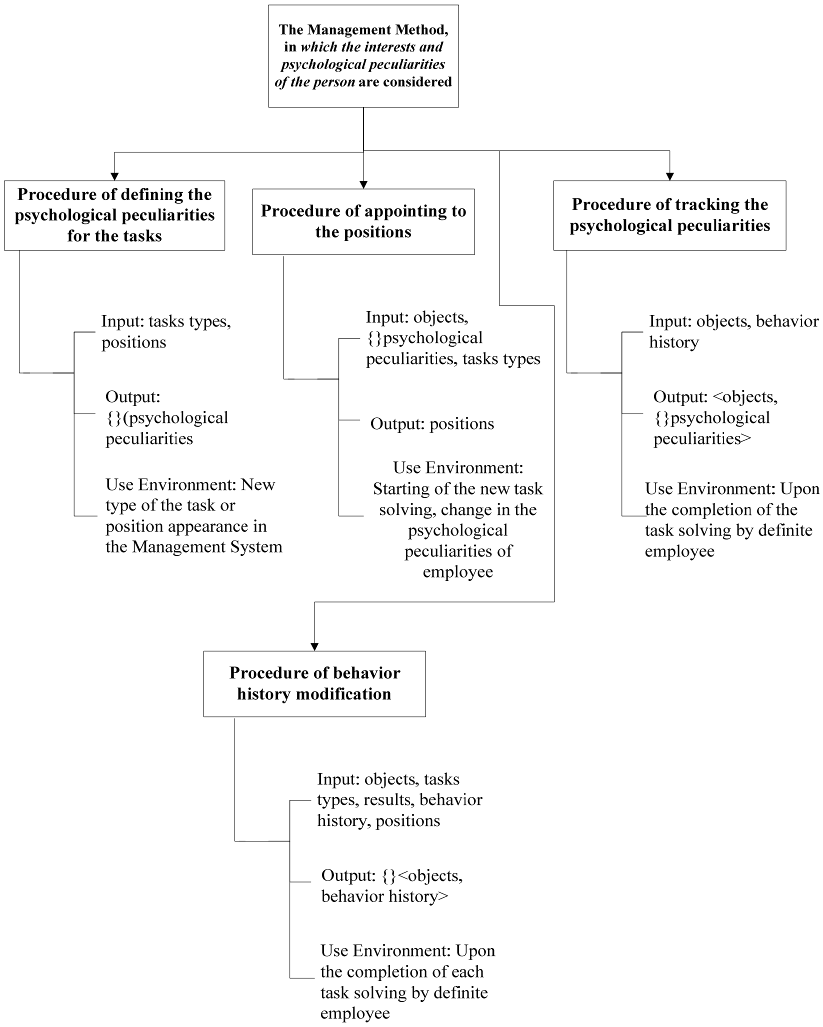
  Use Environment: Upon the completion of the task solving by definite employee.

- Procedure of behavior history modification:

  Description: This procedure modifies the behavior history of the employee each time upon the completion of his task solving according to the result.

  Definition: (objects, tasks types, results, behavior history, positions) → ({}<objects, behavior history>)

  Use Environment: Upon the completion of each task solving by definite employee.



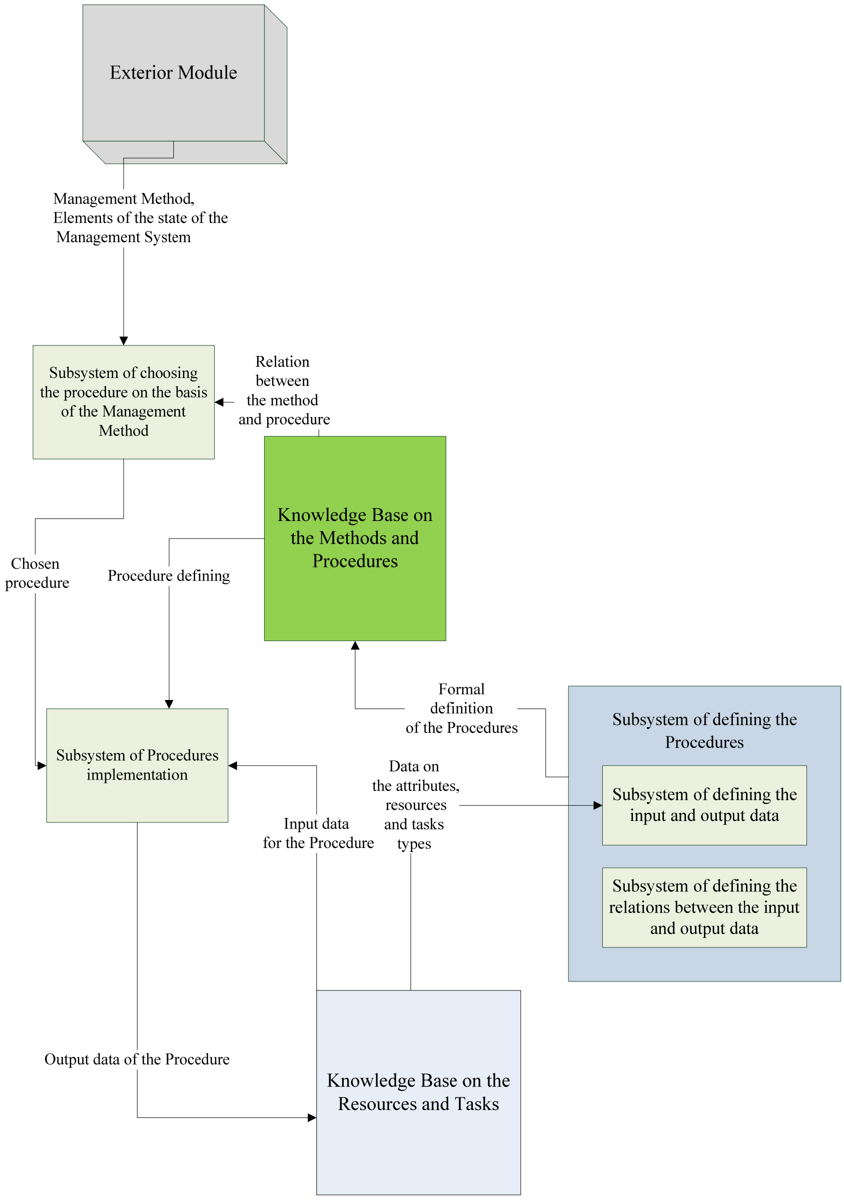
Scheme 5. The Management Method, in which the interests and psychological peculiarities of the person are considered

After detailed examination of the Management Methods and the Procedures representing these Methods, within the framework of solving the task of intellectual management of the proposed Implementing Management Methods Subsystem, we can get the generalized model of this Subsystem, providing the widening of the used methods, procedures and their interrelations variety. This model is represented on the Scheme 6.

The development of the Implementing Management Methods Subsystem is carried out within the framework of developing the Unreliable Elements (Human Resources) Intellectual Management System (during their professional tasks solving). Let us examine the Architecture of this System, as is shown on Scheme 7.

The Elements of Architecture (Scheme 7)

The knowledge on the Unreliable Elements Management, which is put into the System, is represented by the enlargeable knowledge base. Using this knowledge, the System will make intellectual management decisions. Since the resources in the examined Intellectual System are human resources, it is necessary to take into account the human “unreliability” in the System: to formalize the human peculiarities that hamper or block the activity upon the prescribed decisions plan, to define the possible and justified motivation methods to prevent these “blocking’ impacts. At the same time it is necessary to consider the peculiarities of each particular “unreliable element”. Creation of the Knowledge Bases and data base, shown on Scheme 3, will provide for fulfillment of these management purposes.

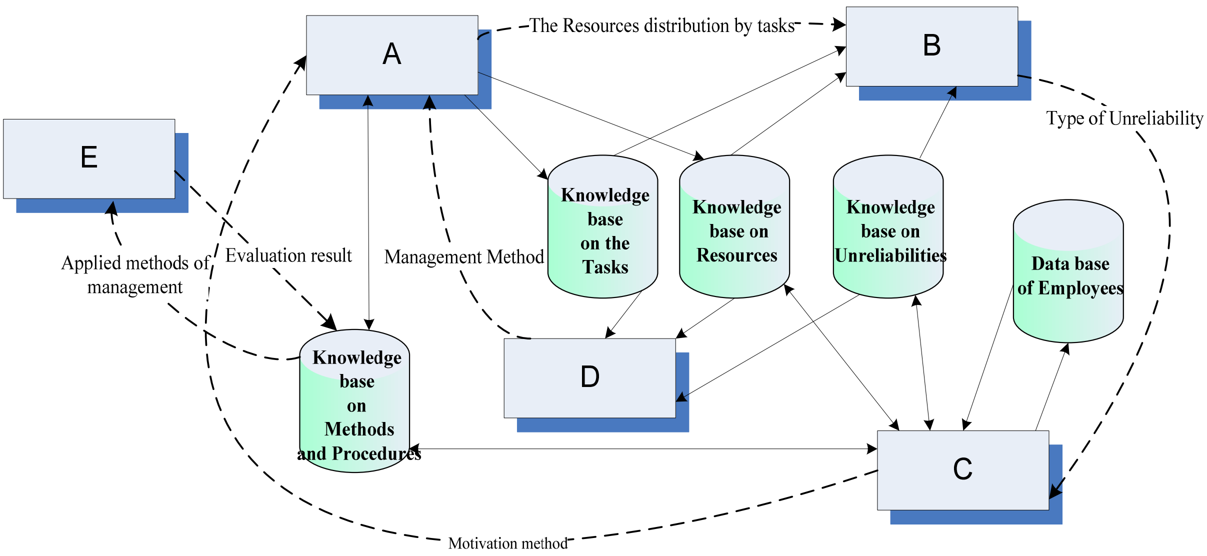


Scheme 6. General Structure of the Methods and Management Procedures Choosing System

It is also necessary to define the algorithm, which allows to track the process of the professional activity tasks fulfillment, interfere in the process of the activity and implement the managing procedures necessary for particular case and particular “unreliable element”, in order to provide the fulfillment of the decisions plan for professional activity (Subsystem B).

The suggested approach to the Management System designing allows to track the effectiveness of the managing procedures implementation and corresponding human resources reaction to the implemented impacts. Knowledge on effectiveness and human reactions will allow the system to provide self-training in future, thereby improving the possibilities of the System implementation during its further operation (Subsystem E).

Thus, the conclusion could be made, that the functionality, put into the Intellectual Management System, is characterized by the influence (by implementation of particular management event – the number of management procedures) on the “unreliable element” functioning. Such influences may cause the need of the decisions plan revising.



A – Subsystem of procedures implementation on the basis of the chosen methods
B – Subsystem of plan execution and unreliabilities definition monitoring
C – Subsystem of motivation method defining
D – Subsystem of managing method defining
E – Subsystem of implemented managing methods effectiveness evaluation

Scheme 7. The Unreliable Elements Management System (during their professional tasks solving)

Conclusion

The Article suggests the author's view on the approach of the automatized Intellectual Management of people's activity, considering the mechanisms of influencing the human resources as the "Unreliable Element" within the Management System. The Article suggests the approach to development of the Management Procedures Implementation Model on the basis of the chosen method within the Unreliable Elements Management System.

The main components of the Model are defined. Also touched upon are the questions of System Architecture, which can provide the control over the tasks fulfillment and increase the effectiveness of professional tasks execution by the people, using the Management Procedures.

The suggested approach allows to create the intellectual programmed instrument which would, unlike the already existing programmed management systems, cover all the characteristics and abilities of the human resources and will increase human resources reliability during their professional tasks solving.

Bibliography

[Benger, 2006] T.V. Ryabtsev, E.I. Antonova, R.V. Benger, "The Model of Intelligent system of activity people control in solving professional problems". In: "Artificial Intelligence 2006" magazine, Institute of Artificial Intelligence, Ukraine, Donetsk, 2006.

Authors' Information

Roman Benger, Elena Antonova – Far East State University, Institute of Mathematics and Computer Sciences, P.O. Box: 690002, Okeanskiy avenue, 99 - 136, Russia, Vladivostok; e-mail: stainberg@mail.ru

THE ROLE OF DBMS IN ANALYTICAL PROCESSES OF THE LOGISTIC OF STOCK RESERVES

Julian Vasilev

Abstract: One of main problems of corporate information systems is the precise evaluation of speed of transactions and the speed of making reports. The core of the problem is based on the DBMS that is used. Most DBMS which are oriented for high performance and reliability of transactions do not give fast access to analytical and summarized data and vice versa. It is quite difficult to estimate which class of database to use. The author of the article gives a concise observation of the problem and a possible way to be solved.

Keywords: Database management systems (DBMS), Information technology, Cache, Interbase, post-relational DBMS.

ACM Classification Keywords: H.2.8 Database Applications, H.4 information systems applications.

Introduction

Stock reserves affect all activities in the enterprise. Their management is directly connected to the appliance of different methods. According to Gatorna [1,333] "…logistics includes two main spheres of activity: management of materials entering the production and the management of distribution of final products."

According to the information needs of the operational management we meet requirement of different groups of users who give priorities for fastness of defined actions. For instance, people who get stock from suppliers and store it in the factory warehouse want high speed of their transactions. By analogy with supplies, the sales department needs high speed in making, saving and printing invoices. Having in mind these requirements, the key role of the DBMS (Database management system) is to provide high speed of transactions. As we mentioned these transactions concern registering documents in the information system of the enterprise. In practice certain raw materials are stored and "wait" their participation in production. At the end of the enterprise final products form also stock reserves. This type of production activities is usual for Bulgarian enterprises. Without involving logistics, these activities are organized in accounting software.

Layout

According to managers, production operations are subject to monitoring using certain indicators, such as effectiveness, profit, costs, revenues. A full control of material flows is a precondition for the science logistics to give a solution to several problems in the sphere of creating an order for supply. Supply department needs to know the amount of order, the frequency of sending, stocks included, supplier. Analytical processes refer to a higher level of data aggregation and extracting synthesized data. Some authors [2,100] give account of necessity of expert systems in logistics for solving complex problems. According to other experts [3,182] in building "computers are used for solving complex problems of planning and for techno-economical problems when building plans for material and technical providence". According to other authors [5] "multi-dimensional data structures are the base of the conception of direct analytical process". Top managers are interested in the dynamics of several indicators in order to monitor the state of the organization. The warehouse of the enterprise generates big amounts of data. Data increases throughout the time. Observations in practice show that new technical data storage devices are with bigger volume, than their precedents, so data storage is not a problem for IT specialists. Problems occur when we need to show dynamic indicators. The use of DBMS, accenting to high speed of transactional performance, for example Cache (a product of Intersystems), guarantees quick and reliable registration of data. DBMS of that class usually make reports slower than other classes of DBMS such as Oracle or MS SQL Server. The application of last two databases is combined with a spread application development instruments for OLAP (on-line analytical data processing) data analysis, which make them adjusted to solve managers' problems of high level. Information systems based on Oracle require significant hardware resources. Another disadvantage is the speed of transactions. Keeping indexed data in relational structures is a requirement for fast speed of OLAP instruments. But DBMS need much time to keep indexed data actual. With increase of data more time is needed for a transaction to be saved. We made an observation. Results are summarized in Table 1.

Table 1: Comparison between Cache, MS SQL Server and MS Access

DBMS/Indictor	Cache	MS SQL Server	MS Access
SQL access	yes	yes	yes
Support of huge data structures	yes	yes	yes
Speed of grouping and summarizing data	low	**high**	middle
Support of OLAP	no	yes	no
Speed of saving transactions	**high**	middle	low

We consider that the key problem of the DBMS role is the seeking and finding of an objective compromise between high speed of transactions and high speed of analytical processes in management. Most IT experts prefer popular DBMS such as Oracle, MS Access, Interbase, DB2, MS SQL Server. Advertisements in IT magazines and newspapers, application development environments make them attractive. Managers need to monitor a set of economic indicators concerning activities in the enterprise. The compromise is oriented to high speed of registration accounting documents from one side and to the logistics on the other side. A huge document flow requires high speed of transactional performance. If it is not provided, customer services are slowed down. Moreover the company needs personnel to process these documents. Customer services have to be fast and with high quality. These are main priorities of marketing. Whereas future sales are subject to research of the science “Forecasting”, future supplies are a matter of organization of the procurement department and are subject of discussion by scientists in the sphere of logistics. Forecasting sales and organization of future supplies is deeply connected to analysis of big arrays of data, meaningless methods which can be applied. Backing our opinion other authors [4, 57] think that “Forecasting of material requirements is based on values from historical data. Proving of future needs is helped by statistical methods”.

Each DBMS is optimized either by the processing of transactions (OLTP – on-line transaction processing) or – by analytical processing of data - OLAP (on-line analytical data processing). Their performance is compared by the execution of test done by independent companies, such as TPC – Transactional Processing Council. SQL Server support tools for optimizing queries, such as SQL Query Analyzer, Query Governor – for the determination of priorities of execution of queries. The creators of DBMS Cache have built several tools to fasten processing of queries. One of them is called “Write Demon”. All queries which require operations including adding or modifying data are non-synchronic – the server accepts the query and gives back to the client a flag for successful execution. In the meanwhile, Write Demon adds or modifies data. Another process is called “Garbage Collector”. Its purpose is processing of queries for deleting data. It functions likewise “Write Demon”. More information about comparing DBMS-s can be read in an article, published by Bloor Research [7, 64].

As we said the fastness of a DBMS can be in the sphere of transaction processing or analytical processing. If we study the problem carefully we will find out that the approach for making indexes in the DBMS is the nostrum. If we compare Intersystem’s Cache and MS SQL Server the approaches for making indexes are different. SQL Server uses B-tree technology for indexing. It supports cluster and non-cluster indexes. The common rule is: the more indexes we have in a table, the more time we wait on adding data, the less time we wait on extracting summarized data. Cache uses block structure to save physically records of data. The organization of blocks resembles the file allocation structure in operating systems: UNIX, DOS, Windows. Cache is optimized for multi-user access and processing of high intensity flow of transactions. Grouping and searching of data is executed slower than relational DBMS. For the purpose of retrieving data faster, designers of information systems have to create artificial structures analogical to index data. This approach leads to: firstly, increasing the size of the database because doubled data is stored and secondly, business logic complicates. In contrast to relational DBMS where adding of data takes more time, the same operation in a post-relational DBMS takes less time. For the purpose of our research we made a comparison between Cache and Interbase. Results are summarized in Table 2.

Table 2: Comparison between Cache and Interbase

DBMS/Indicator	Cache	Interbase
Time to insert 100 000 records in the database, each is 255 bytes long	8 seconds	350 seconds
Filtering 100 000 records and retrieving 5 000 records by a given criteria	4 seconds	8 seconds

It is indisputable that there is a wide choice of DBMS. In accordance with the opinion of some authors [6] “For building databases commercial DBMS can be used such as ORACLE, MS-SQL, DB2 and others, as well as free versions.” We consider that for building corporate systems IT experts cannot rely on DBMS which does not have support. Experiments were done on Win XP, with 128 MB RAM, Intel Celeron processor 466 MHz. To visualize resulting data we used following technologies: HTML (hyper text markup language), ASP (active server pages) and GUI (graphic user interface) application. The GUI application visualizes data faster than ASP technology which needs an IIS (Internet information server).

In a post-relational DBMS (such as Cache) records are stored as blocks. Each block has: a unique address, Boolean flag (free or occupied), data range, address of the next block. Adding data means rerouting the address of previous block to the next block and changing the flag from “free” to “reserved”.

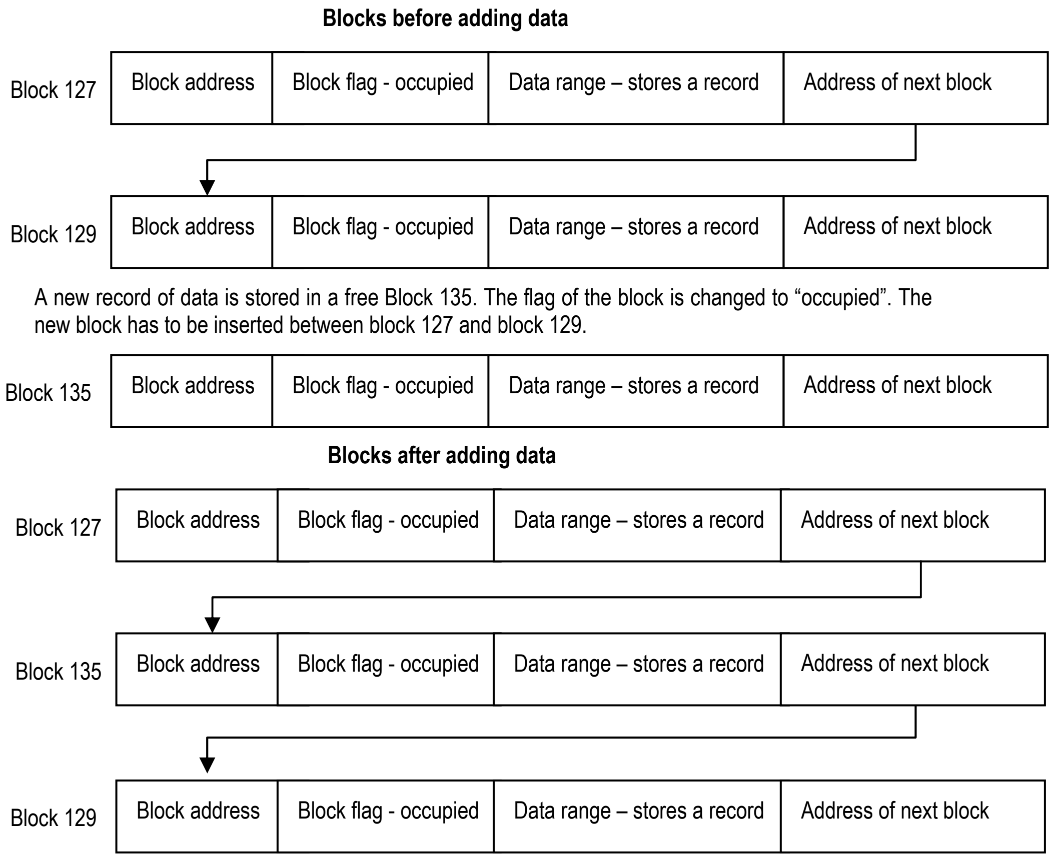


Fig. 1. Adding data in block structures

This organization of adding data resembles the management of RAM (random access memory) and the management of queues and stacks. The difference is that blocks are stored on hard disk drives. Deleting of data is executed in the following order – the flag of the block is changed to “free”, bits in data range are set to “0” and the pointer “address to next block” is set to “null”. In figure 2 we can see sample deletion of block 135 (which we already added).

Computer programmers don’t manipulate directly blocks. They write “Insert” or “Delete” SQL clauses. These clauses are interpreted by Cache and transformed to block operations. In this way programmers use the language MUMPS (Massachusetts University Multiprogramming System), COSL (Cache Object Script Language) or SQL queries.

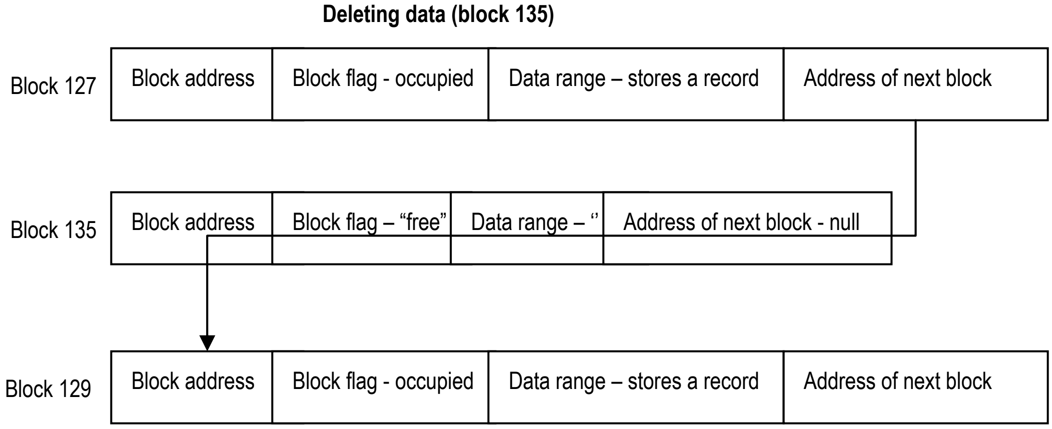


Fig. 2 Deleting data in block structures

For now, there is not a DBMS which is optimized in both directions – transactional processing and analytical processing. That is why, when a company wants to choose a DBMS, it has to estimate which category of DBMS needs. In the sphere of logistics – for the registering a high intensity flow of transactions we need an OLTP system. From another side, for the decision support processes we need OLAP instruments and an analytical DBMS. These facts open another problem – defining the interface between both DBMSs.

Conclusions

Estimating DBMS we need to bare in mind not only its popularity but its orientation to fast transactions or high speed of analytical processes. The right choice is based on logistic processes, hardware resources, personnel and an objective forecast for the increase of data arrays. The last factor is usually ignored and sometimes it is decisive. Its correct evaluation is done after several years of functioning of a store or accounting system. One of the ways to solve this problem is to be built an application server, which acts as a Windows process which transports data from one DMBS to another. For instance we can have an information system based on Cache, and a second one based on MS SQL Server. The first one is used for registering documents such as invoices and stock receipts, the second one for OLAP analysis. The mediator is the upper mentioned application server. In our opinion it is a temporary solution until further development of DBMSs.

Bibliography

[1] М Гаторна, Д. Основи на логистиката и дистрибуцията, Бургас, 1996.
[2] Гаджинский, А. Логистика, Москва, 2000.
[3] Генов, Г. Логистика в строителството, ИУ Варна, 1998.
[4] Коралиев, Я. Логистика на снабдяването, София, 1998.
[5] Несторов, К. Организация на данни в програмни продукти за счетоводно отчитане, сп. Известия, ИУ Варна, бр. 2, 2002г.
[6] Петров, П. Проблеми на изграждането на автоматизирани информационни системи в големи производствени предприятия сп. Известия, ИУ Варна, бр. 1, 2002г.
[7] Howard, P., On-Line Database Technology, Bloor Research Volume 1, 2004

Author's Information

Julian Vasilev – Chief assistant professor, Department of Informatics, Varna University of Economics; 77, Kniaz Boris I str.; Varna; Bulgaria; e-mail: vasilev@ue-varna.bg

ON THE MODELS OF DEVELOPMENT AND DISTRIBUTION OF SOFTWARE

Neli Maneva, Krassimir Manev

Abstract: The notion model of development and distribution of software (MDDS) is introduced and its role for the efficiency of the software products is stressed. Two classical MDDS are presented and some attempts to adapt them to the contemporary trends in web-based software design are described. Advantages and shortcomings of the obtained models are outlined. In conclusion the desired features of a better MDDS for web-based solutions are given.

Keywords: models of development and distribution of software, efficiency, web-based solutions.

ACM Classification Keywords: D.2.7 Distribution, Maintenance, and Enhancement; D.2.9 Management.

Introduction

It is a fact that the software industry is one of the most successful and the volume of sales of software is growing each year. According to a study of *Forester Research* [CIO-3, 2007] the volume of sales of corporative software only for 2006 is up to 50 billion USD. The efficiency of the software solutions and especially of the business software solutions is one of the most important challenges for the software industry. Some statistics show (see for example [CIO-11,2006]) that about 50% of the installed Enterprise Resource Planning systems (the leading class of corporative software products) in companies from USA and Western Europe are inefficient. Such negative trends are crucial and have to be analyzed in order to identify the reasons for them. In this paper we will try to introduce a notion that will permit us to analyze the problem.

Let us denote with the notion *model of development and distribution of software* (MDDS) the system of principles, rules, and especially ***relations*** established between a developer and the potential users during the whole lifecycle of a software product. It will happen that, beside the other usual characteristics of a specific software product – its functionality, usability, easy maintenance, etc. – the used MDDS could be crucial for the efficiency of the product. This is especially valid for software developed for small and medium companies with limited resources.

In second section two classical MDDS – *On Purpose* and *On Demand* – are discussed. An attempt is made to list some important characteristics of these MDDS and to stress their role for the efficiency of the software product. Third section outlines the requirements to the MDDS that are generated by the development of the contemporary web-based software solutions and how these requirements could be satisfied on model level. Some conclusions and recommendations for features of an MDDS adequate to the development of web-based software solutions are given in forth section.

Model of development and distribution of software

Let us consider a simplified *development and distribution of software process* (DDSP) with only two kinds of participants – *developers* and *users*. Since the beginning of the software production two classical MDDS were applied – *On Purpose* and *On Demand*.

On Purpose model. When a developer identifies a task/need that is concerning many users and that could be solved/satisfied by a software product then he creates such product and put it on the market. The potential users find on the market one or more such products, chose one of them and use it to solve their tasks or satisfy their needs. We will call such model of development and distribution of software *On Purpose* (OP) and the corresponding software products – *OP-products*. For example, as OP-products essentially operating systems, software development environments, DBMS and other *system software* are developed and distributed.

Being dedicated to serve many potential users OP-products have relatively low price. Due to the same reason OP- products are trying to cover as many as possible aspects of the solved problem or satisfied need and sometime they include too much functionalities. Any way, it is possible that functionality, specific for some user, is missing in an OP-product. Terms of delivery of these products are very short because they are complete and ready for distribution.

On Demand model. When a user identifies a task/need that is concerning him and on the market there is no software product able to solve/satisfy task/need then the user finds developers that are able to create such product, then chose one of them and assign the development of the product to him. We will call such model of development and distribution of software *On Demand* (OD) and the corresponding software products – *OD-products*. For example, as OD-products are developed and distributed different kind of *applications* for end users – information and resources management systems, etc.

Terms of delivery of such products are usually long and sometime unpredictable. Being dedicated to serve a single user the OD-product have relatively high price. Due to the same reason OD-products cover almost precisely the needs of the user. Anyway, it is still possible that, during the process of usage, some specific for the user functionality is happening to miss in the OD-product.

In both cases it is coming a moment when the user of OP-product has to buy a new version and, respectively, the user of OD-product has to demand redesign of the product. Having in mind the trend of shortening the life cycle of the software products, this is leading to additional expenses for users. The developers could try to escape the stressed negatives by some build-in components for *tuning* of the product. Tuning could be very easy – giving values to set of parameters of the product, as well as very difficult – writing a code in some (inner for the product or standard) programming language. Successful tuning of a software product will depend too much on the *qualification* of the user.

These two classical models are extreme cases of MDDS and it is possible to define different middle cases – more close to one or the other of the extreme cases. But they are demonstrating how important is to establish mutually beneficial relationship between developers and users. For the users it is very important to obtain the product that meets its requirement – especially the functionality and performance, under objectively given constrains – price, term of delivery, level of complexity of usage and tuning etc. Developers, from the other side, have to be able to organize efficient development process in order to make profit by supplying products that meet requirements of users. Contemporary MDDS has to be able to help for establishing such mutually beneficial relationship.

MDDS for web-based software solutions

The age of web-based software solutions raises some new objectives to the MDDS making the both classical models inappropriate. Instead of a single large task/need to be solved/satisfied by the product, there is a diversity of smaller tasks/needs dynamically changed over the time. For each user there is a subset, also dynamically changed, of such smaller tasks/needs. So it is **no more necessary** that a single software product to solve all such tasks or satisfy all such needs. It is quite more efficient if the product is able to supply the corresponding solution as *content* or to satisfy corresponding need as a *service* through the (local or global) net. Something more, web application users ask for services and contents *highly tailored* to their particular contexts of usage. It is **no more possible** for a single software product to meet all requirements of the user.

The contemporary software *solution* is rather a set of interacting *applications* (some of them web-based, some of them working locally). Each application could be purchased as OP-product or as OD-product. It is not so difficult to identify some typical applications that will be necessary for any web-based software solution (i.e. web-server and web-client, DB-server etc.). Such applications are developed and distributed as OP-products and are available on the market. The other applications in the solution have to reflect the particularities (business logic) of the user and it is more natural to be developed as OD-products. The challenge of building integrated web-solutions is to ensure the interoperability of the applications.

The answer of big developers to this challenge is to organize development of as many as possible of the necessary applications around some general concept or approach, called *technology* (i.e. Java-technology of SUN, or .NET-technology of Microsoft). They also have to provide some tools (programming languages, IDE, API, etc.) to support integration of the solution. There is no standard definition what a “technology” really is and what has to comprise, so the products of different developers could be quite different by implemented functionality. Small developers who accept the technology (*technology-partners*) could develop some applications that are missing and even to develop better versions of some existing applications. We will call such model *Technology-based MDDS*.

Technology-based model. The user is choosing one of the possible technologies, buy the necessary applications that fit the technology and that are available (OP-products) on the market. Then the user searches a developer which is able to create the missing parts of the solution as OD-products and to integrate them in the solution. In most of the cases such developers are the corresponding technology-partners.

Technology-based MDDS has some positive aspects – the technology is usually supported by a stable company and large set of partners who are ameliorating the technology, applications and instrumental tools. But some negative aspects could be outlined, too. The most negative aspect of such model is the *dependence* of the user on the technology. For example, any attempt of the user to extend the solution with new functionality has to be implemented "within" the technology. Issuing of new generation of elements of the technology could lead to necessity of total upgrading of all bought to the moment elements. And more, when an element of the technology is of low quality, comparing with the concurrent products with same purpose, it is very difficult to eliminate this element from the solution and to replace it with a better one.

Minimalistic model. The Technology-based model is appropriate for big users that have resources to keep their solutions up to date with the technology. Small users usually prefer more cheap solutions. For example, the famous combination Appache web-server + PHP + MySQL DB-server is applied in many web-solutions of small users. Such solutions are very cheap because the three products are free and all expenses of the user are for integrating the solution. We could accept such form of development of web-solutions as a model and to call it *Minimalistic*. It is clear that the Minimalistic model has much more negatives then positive aspects – for example, minimal support from the developers of the free applications, missing of enough instrumental tools, very low level of performance and security etc. Instead of the dependence of the technology small users are caught by the dependence of the limited resources.

The Minimalistic model has its place on the market of web-based software solutions. It is acceptable for some non-commercial solutions – private sites, sites of non-profit societies etc. where the goal is to inform interested people. But it is not appropriate for business companies, where on the software solutions lay much more responsibilities than just to provide necessary information.

Since 2000/2001 an idea is circulating, know as Software as a Service (SaaS) associated with the paper of Tim O'Reilly [O'Reilly, 2004]. It originates from the popular conception "open source" and in principle is an apology, on an abstract level, of a dream for total standardization and interoperability of all software products. The goal is to give to the users freedom to choose any component of their solutions on the base of needs and available resources, i.e. to reduce as much as possible the dependence of the user on technologies and products.

SaaS model. Recently there are some attempts to turn the idea into realistic MDDS – *Software as a service model*. Unfortunately, as it is shown in the description of SaaS, given in Wikipedia, the attractive idea was transformed in non-attractive model. The essence of the model is that the developer (of software products) in simplified bipolar DDSP is replaced by a *provider of services*. If the user need some service (for example, to keep data and search them) than the provider is organizing (on his own servers) the corresponding DB and give to the user the necessary web-based interface in order to input, update and search the necessary data. If the user would like to publish a site in Internet then the provider is developing the site and is hosting it. Maintenance of the content of the site is made also by provider on the base of data supplied by the user.

Positive element of the SaaS model is the possibility the user to solve a task or to satisfy a need without buying the corresponding software product but use it as a service. For tasks or needs that are arising rarely such form could be the most efficient. The negatives of SaaS model, formulated as it is mentioned above, are obvious. It is clear what will happen if the user is loosing for some period of time the Internet connection with provider or provider announce that he lost some of the user's data. The dependency of the technology or of the limited resources is replaced by the dependence of provider.

Conclusion

As a result of this overview we have to conclude that especially for the medium and small users of business web-oriented software solution there is **no established and appropriate model** for development and distribution of software. This leads to big inefficiency of the implemented solutions – for relatively large expenses made by the users they do not obtain the products that are able to solve their tasks or to satisfy their needs in an adequate way.

It is not easy to establish and reason a new MDDS that is adequate of the needs of the medium and small clients but it is worth to invest some efforts in it. Taking into account the discussion above, we could define the following features of such model:

- It has to guarantee the **independence of the user** from the technologies, i.e. the user have to be free to chose for each component of the solution the existing technology that is the best for this component;
- It has to guarantee that the user will obtain a service with a **quality that is relevant to the paid cost**;
- It will be very good if the model has the **tolerance for the qualification** of the user and to allow, etc.

In order to prove feasibility of such model a lot of programming and experimental work should be done. Each feature of the model has to be examined on sample applications. For example, implementing of efficient and fast interoperability of technologically non-homogenous components is a great challenge.

Bibliography

[CIO-11,2006] Editorial, CIO, No. 11, 2006, Internet version. http://www.cio.bg/?call=USE~home;&page=paper&n=1148.
[CIO-3,2007] Editorial, CIO, No. 3, 2007, Internet version. http://www.cio.bg/?call=USE~home;&page=paper&n=1391&pn=2
[O'Reilly,2004] http://tim.oreilly.com/articles/paradigmshift_0504.html

Authors' Information

Neli Maneva – Senior Research Fellow; Institute of Mathematics and Informatics, BAS, Acad. G.Bonthev St., bl.8, Sofia-1113, Bulgaria; e-mail: neman@gbg.bg
Krassimir Manev – Associated Professor, Faulty of Mathematics and Informatics, Sofia University, 5 J. Bourchier str, Sofia-1164, Bulgaria; e-mail: manev@fmi.uni-sofia.bg

THE LOCALIZATION PROCEDURES OF THE VECTOR OF WEIGHTING COEFFICIENTS FOR FUZZY MODELS OF CHOICE

Elena Prisyazhnyuk

Abstract: The author analyzes the localization procedures of the vector of weighting coefficients which are based on presenting the function of value by additive reduction adapted to fuzzy models of choice

Keywords: fuzzy sets, coefficients of value.

Introduction

Different kinds of uncertainty are to some extent characteristics of practically any situation of making decision in which the expert information is used. The result of the research [Ларичев, 2002; Борисов, 1989] show that the main difficulty is caused by the necessity to appraise numerical values of the objects (variants, criteria) or to give numerical evaluation on the ratio scale between them. It is known that verbal definitions usage allow to define the preferences of a person who makes decisions (PMD) more steadily. This approach seemed to be even more justified because in prevailing number of the cases it is enough to have the approximated characteristic of the data set and the expert info usage does not demand high accuracy.

The experts' evaluation in fuzzy models of choice is described by the functions of belonging to the fuzzy set. The functions of belonging can be interpreted differently: as "subjective probability", expert's confidence degree in object's belonging to the concept described by fuzzy set, the opportunity of its interpretation by this concept and so on [Борисов, 1989]. The choice is characterized by preference relation R the meaning of which in fuzzy models consists of the fact that it can point out for every two objects:

- the fact of preference of object α[1] over α[2]. The preference function μ_R(α[1], α[2]) in this case is substantially interpreted as the expert's degree of confidence in the fact that α[1] is not less preferable than α[2]. The confidence degree can be described both numerically and verbally, for example, by the linguistic variables

"confidence degree"={very low, low, average, high, very high}. The linguistic preference relation of this type corresponds to the decision making situations when PDM doubts in preferences existence in relations of the certain objects and because of that it's difficult for him to express them only in terms "yes" (definitely dominates) or "no" (is definitely dominated);

- preference power of object α[1] over α[2]. Fuzzy relation R in this case shows the idea of preference power; besides PMD is sure in the very fact of preference (and in this sense his preferences are clear). Here μ_R(α[1], α[2]) can be interpreted as degree with which α[1] is definitely better (preferable) than α[2].

The problem of the objects' evaluation in terms of value theory settles down to the problem of axiomatic argument and its value function creation. The classical methods used for defining value which represents binary preference relation R($U(a^1) \geq U(a^2) \Leftrightarrow a^1 R a^2$ для $\forall a^1, a^2 \in A$), are generally pretty hard. The basis for their usage, in particular, is a sufficient condition of its existence which is set, for example, by Debre theorem [Пономаренко, 1994]: the preference relation must be complete, reflexive, transitive and continuous, the set of decisions – connected. If Debre theorem conditions are not completed and the function of value which introduces relation R does not exist, it's difficult to use the classical methods.

The procedure of the problem formalization with the help of exchange of fuzzy "vector value evaluation" by additive reduction is suggested.

The task set

Let's A as the universal (distinctly described) set of objects α[j], $j \in J$, where *J* – set of indexes of objects. Each of objects $a^j \in A$, $j \in J$ is characterized by the set of parameters $a^j = (a^j{}_1,...,a^j{}_i,...,a^j{}_n)$. Let's mark the set of indexes of objects' parameters *I*, $I = \{1,...,n\}$. Each object, $a^j \in A$, $j \in J$ corresponds to its vector evaluation in the dimension of objects' parameters Ω^n.

Further on we'll analyze not the set of the values of objects' parameters $α^j{}_i$∈A, $j \in J$ itself but the corresponding set $\omega(a^j{}_i)$, $i \in I$, $j \in J$, where ω – some monotonous reorganization which defines the degree of quantitative characteristics declination from the optimum meanings for each parameter $a^j{}_i$, $i \in I$, $j \in J$ and reorganizes all the meanings of objects' parameters towards the normal type in interval [0,1].

Let the expert consequently define his preferences on the set A as fuzzy binary relations of preference R.

The following approach to the task solution is suggested: we suppose that evaluating the object the expert (consciously or subconsciously) means its vector value. If we consider "vector" function of value as fuzzy additive reduction the task is considered as defining weighting coefficients of reduction (1) – (2):

$$\sum_{i \in I} \rho_i \omega(a_i^1) < \sum_{i \in I} \rho_i \omega(a_i^2) , \tag{1}$$

$$\rho = (\rho_1,...,\rho_n) , \ i \in I , \ \rho_i > 0 , \ \sum_{i \in I} \rho_i = 1 , \tag{2}$$

where (2) – normal vector of relative importance of objects' parameters for the expert's statement about the fuzzy preference relation between the objects.

So, the task is in "localization" (defining intervals of changes) of weighting coefficients of additive reduction (1)-(2). Similar task was analyzed in [Волошин, 2003], in this article we suggest the generalization of the method for fuzzy models of choice.

The localization procedures of the vector of weighting coefficients

Let the person who makes decisions (PMD) considers the object α[1] more preferable than object α[2], and μ – preference degree, $\mu \in [0,1]$. We'll define it by $\mu_{\succ}\left(a^1, a^2\right)$. Then $1 - \mu_{\succ}\left(a^1, a^2\right)$ – preference degree of α[2] over object α[1]. Let's use heuristics: we think that righteousness of inequality (1) comes from PMD subjective idea about object preference α[1] over α[2]. Then for the case of fuzzy preference relation $\mu_{\succ}\left(a^1,a^2\right)$ between objects α[1] and α[2] we'll consider right:

$$\mu_{\succ}\left(a^1,a^2\right) \Leftrightarrow \sum_{i\in I}\rho_i \frac{\omega(a^1{}_i)}{\mu} \le \sum_{i\in I}\rho_i \frac{\omega(a^2{}_i)}{1-\mu} . \tag{3}$$

It is necessary to create on the basis of preference relations on the set of effective objects A which are consequently defined by PMD, the intervals of allowed values of objects' weighting coefficients (hyper parallelepiped of weighting coefficients - HWC) as:

$$\rho \in K = \prod_{i\in I}[\rho_i^H,\rho_i^B], \quad \rho = (\rho_i,\ i\in I), \quad 0<\rho_i^H \le \rho_i^B <1, \tag{4}$$

$$\sum_{i\in I}\rho_i = 1, \rho_i > 0, i\in I . \tag{5}$$

It is supposed that defining preferences PMD consequently, particularly, in his ideas sets the preference relations between the objects. They satisfy the qualities of transitiveness. Binary comparison is given with the help of the line reduction (1) and HWC as the result of the suggested procedure functioning is reduced step by step ($K^{S+1} \subseteq K^S$, $s = 1,2,...$). So transitiveness of considered binary relation is preserved.

For reorganization of all the meanings of objects' parameters $a^j{}_i$, $i\in I$, $j\in J$ to the unlimited kind in the interval [0,1] such formula is suggested:

$$\omega(a^j{}_i) = \frac{a^{opt}{}_i - a^j{}_i}{a^{opt}{}_i - a^0{}_i}, \tag{6}$$

where $\alpha^j{}_i \in A$, $i\in I$, $j\in J$; $\alpha^{opt}{}_i \in A$, $i\in I$ - the best meaning of i-parameter on the set of effective objects; $\alpha^0{}_i \in A$, $i\in I$ - the least meaning of i-parameter on the set of effective objects. Let's consider that α^{opt} and α^0 can be set directly by PMD or defined as maximum (minimum) parameters values which are achieved on the set of admitted.

Taking into consideration (6), the generalized criteria which reflects the total declination of j-object, $j\in J$, from optimum meanings will be presented as

$$D(a^j,a^{opt}) = \sum_{i\in I}\rho_i\omega(a^j{}_i) = \sum_{i\in I}\rho_i \frac{a^{opt}{}_i - a^j{}_i}{a^{opt}{}_i - a^0{}_i}, \quad j\in J .$$

The last formula is the proximity metrics of values vector closeness of object's parameters $\alpha^j \in A$, $j\in J$, to some ideal (optimum) values vector $\alpha^{opt} = (\alpha^{opt}{}_1, \alpha^{opt}{}_2, \ldots \alpha^{opt}{}_n)$, weighted in the dimension of parameters. Formula (3) for fuzzy preference relations between objects $\mu_{\succ}\left(a^1,a^2\right)$ will be presented as:

$$\frac{D(a^1,a^{opt})}{D(a^2\ ,a^{opt})} = \frac{\sum_{i\in I}\rho_i\omega(a^1{}_i)}{\sum_{i\in I}\rho_i\omega(a^2{}_i)} \le \frac{\mu}{1-\mu} .$$

Last inequality can be interpreted in the following way: the statement "object α[1] is preferable than object α[2] with preference degree μ" means that in the dimension of objects' parameters the point which corresponds to object α[1] is located closer to the ideal point than the point which corresponds to object α[2] with degree $\frac{\mu}{1-\mu}$.

Statement 1. Objects $a^1 \in A$ and $a^2 \in A$ are called equal if in "equable" dimension of parameters Ω^n the corresponding points are located within equal distance from the point which corresponds to the ideal object.

Statement 2. Object $a^2 \in A$ is called μ-equal to object $a^1 \in A$ if in "equable" dimension of parameters Ω^n points $\omega\left(a^2\right)\frac{\mu}{1-\mu} = \left\{\omega(a^2{}_1)\frac{\mu}{1-\mu}, \omega(a^2{}_2)\frac{\mu}{1-\mu}, ..., \omega(a^2{}_n)\frac{\mu}{1-\mu}\right\}$ and $\omega(a^1)$ are equal.

Statement 3. Weighting coefficients of parameters $\rho = (\rho_i,\ i \in I)$ which correspond to μ-equal objects in the dimension of preferences R^n define the intervals' limits of weighting coefficients of objects' parameters.

Argumentation. Let's define sets of indexes of objects' parameters $a^1 \in A$ and $a^2 \in A$ through $I_1 = (i : \omega(a^1{}_i) > \omega(a^2{}_i)) \neq 0$, $I_2 = (i : \omega(a^1{}_i) \leq \omega(a^2{}_i)) \neq 0$, $i \in I = I_1 \cup I_2$. We can renew inequality (2) taking into account defining sets of indexes in the following way:

$$\sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^1) + \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^1) \leq \frac{\mu}{1-\mu} \sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^2) + \frac{\mu}{1-\mu} \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^2) . \quad (7)$$

Then the condition of μ-equality of objects α 1 and α 2 will be given as

$$\sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^1) + \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^1) = \frac{\mu}{1-\mu} \sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^2) + \frac{\mu}{1-\mu} \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^s \omega(a_i^2) . \quad (8)$$

We can pass over from (7) to (8) if we increase weighting coefficients of parameters which belong to the set of I_1. At the same time we reduce weighting coefficients of parameters which belong to the set of indexes I_2. So weighting coefficients of parameters ρ_i, $i \in I_1$ achieve their upper borders and weighting coefficients of parameters ρ_i, $i \in I_2$ achieve correspondingly their lower borders. As $\omega(\alpha^1{}_i)$, $\omega(\alpha^2{}_i)$, $i \in I$ are fixed quantities and $\rho^s{}_i \in K^s$, the equality we get can be defined as

$$\sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^{(s)B} \omega(a_i^1) + \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^{(s)H} \omega(a_i^1) = \sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^{(s)B} \omega(a_i^2) + \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^{(s)H} \omega(a_i^2) . \quad (9)$$

Spreading (9) on the case of μ- equality of objects α 1 and α 2 we'll get finally

$$\sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^{(s)B} \omega(a_i^1) + \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^{(s)H} \omega(a_i^1) = \frac{\mu}{1-\mu} \sum_{\substack{i\in I_1\\ \rho_i^s\in K^s}} \rho_i^{(s)B} \omega(a_i^2) + \frac{\mu}{1-\mu} \sum_{\substack{i\in I_2\\ \rho_i^s\in K^s}} \rho_i^{(s)H} \omega(a_i^2) , \quad (10)$$

where $\rho_i^{(s)B}$, $\rho_i^{(s)H}$, $i \in I$ - correspondingly the upper and the lower borders of I-interval of weighting coefficients on s-step of algorithm. Equality (10) is an equivalent to equality (8).

So HWC on s+1 step will be equal

$$K^{s+1} = \prod_{i\in I_1} [\rho_i^{(s)H}, \rho_i^{(s+1)B}] \times \prod_{i\in I_2} \left[\rho^{(s+1)H}, \rho^{(s)B}\right] , \quad (11)$$

Equality (9) probes righteousness of statement 1: weighting coefficients defined with the help of the described above method limit HWC frames.

As only the fact of PMD preference is known, it is given in the form (3), so for defining vector component $\rho = (\rho_i,\ i \in I)$ we'll suggest the hypothesis about righteousness towards inequality of such a type:

$$\rho_{\substack{i\\ i\in I_1}} \omega(a^1{}_i) + \rho_{\substack{j\\ j\in I_2}} \omega(a^1{}_j) \leq \frac{\mu}{1-\mu} (\rho_{\substack{i\\ i\in I_1}} \omega(a^2{}_i) + \rho_{\substack{j\\ j\in I_2}} \omega(a^2{}_j)) , \quad (12)$$

where $k = k_1 \cdot k_2$; k_1 – quantity of parameters with indexes i, $i \in I_1$; k_2 – quantity of parameters with indexes j, $j \in I_2$.

It is obvious that performing inequality (12) is a sufficient condition for performing inequality (7).

Let's pass over in equality system (12) to equalities and exclude k-n-1 equality according to the rule: in each step we exclude equality which add maximum to expression:

$$\max((\omega(a^1{}_i) - \omega(a^2{}_i),\ i \in I_1,\ \omega(a^2{}_j) - \omega(a^1{}_j),\ j \in I_2) .$$

The last condition means that such equalities are excluded which create unwarranted weights enlargement of one group of objects' parameters at the expense of the other group.

Let's add to the system n-1 of inequalities of type (12) as n-equality the condition of weights coefficients setting (5). Let's pass over to equalities and reorganize. We'll get finally the system of n-equality of the type:

$$\rho_{\substack{i \\ i\in I_1}}(\omega(a^1{}_i) - \frac{\mu}{1-\mu}\omega(a^2{}_i)) - \rho_j(\frac{\mu}{\substack{1-\mu \\ j\in I_2}}\omega(a^2{}_j) - (\omega(a^1{}_i)) = 0\,, \quad (13)$$

$$\sum_{i\in I}\rho_i = 1, \rho_i > 0, i \in I\,.$$

From the equality system (13) we precisely define the components of weighting coefficients vector $\rho = (\rho_i,\ i \in I)$ which according to statement 1 limits in vector dimension R^n the intervals of weighting coefficients of objects' parameters.

Taking into consideration the information given above the correctness of the following statement is obvious.

Statement 2. The condition of objects selection ω^j, $j \in J$, from the set A^s is unbelongingness of HWC vector which passes through the coordinates beginning and point $\omega(\alpha^j)$, $\alpha^j \in A^s$, $j \in J$, namely $\rho(\omega(\alpha^j)) \notin K^{(s+1)}$. Vector of weighting coefficients is defined according to the formula given in [Волкович, 1993]:

$$\rho = \rho(\omega(a^j)) = \{\rho_i : \rho_i = \prod_{\substack{t\in I \\ t\neq i}}\omega(a^j{}_t) / \sum_{\substack{q\in I \\ l\neq q}}\prod_{l\in I}\omega(a^j{}_l)\}\,.$$

The person-computer procedure of localization of hyperparallelepiped of weighting coefficients is described in the following sequence.

Step 1. Pointing out the set of effective objects A^0 on the universal set A by one of the methods which are described in the article [Волкович, 1993]. The very first HWC is set as equal to single hypercube.

Step 2. PMD's choice of two objects α^1 and α^2 from the set of effective objects A^s in HWC K^s, s =1,2,... (step of HWC limiting) stating preference and equivalence.

Step 3. Constructing equation system of type (13). Finding solution of the equation system.

Step 4. Specifying HWC limits according to formula (11). If hypercube $K^{(s+1)}$ satisfies PMD's requirements it means the end of the procedure. Otherwise we pass over to the next step.

Step 5. Pointing out the set of effective objects $A^{(s+1)}$ ($A^{(s+1)} \subseteq A^{(s)}$) in HWC $K^{(s+1)}$ and presenting them to PMD for the choice of next two objects with stating for them the preference attitude.

Step 6. Uniting the objects, chosen by the expert on the previous step to the set of discussed objects and analysis on the given set of transitiveness. If the transitiveness is not destroyed then increase of iteration number s=s+1 and passing by to step 2. If the transitiveness is destroyed then the exclusion of these objects from the set of considered objects and passing by to step 6.

Conclusions

The suggested procedures do not demand the complete metrics of binary comparisons of objects and allow to restore the function of expert's value on the fuzzy binary relations' set. The reflection of weighted coefficients vector in the form of the intervals allows to present adequately the level of uncertainty in fuzzy models of decision making.

Bibliography

[Ларичев, 2002] Ларичев О.И. Свойства методов принятия решений в многокритериальных задачах индивидуального выбора // Автоматика и телемеханика, № 2, 2002. – С. 146-158.

[Борисов, 1989] Борисов А.Н., Алексеев А.В. и др. Обработка нечеткой информации в системах принятия решений. – М: Радио и связь, 1989. – 304 с.

[Пономаренко, 1994] Пономаренко О.І. Системні методи в економіці, менеджменті та бізнесі. – К.: Наукова думка, 1994. – 242 с.

[Волошин, 2003] Волошин А.Ф., Гнатиенко Г.Н., Дробот Е.В. Метод косвенного определения интервалов весовых коэффициентов параметров для метризированных отношений между объектами // Проблемы управления и информатики, 2003, № 2. – С. 83-89

[Волкович, 1993] Волкович В.Л., Волошин А.Ф., др. Модели и методы оптимизации надежности слождных систем. – К.: Наукова думка, 1993. – 312 с.

Author's Information

Elena Prisyazhnyuk – Kirovohrad Pedagogical University, Kirovohrad, Ukraine, e-mail: elena_drobot@ukr.net

THE NEW SOFTWARE PACKAGE FOR DYNAMIC HIERARCHICAL CLUSTERING FOR CIRCLES TYPES OF SHAPES

Tetyana Shatovska, Tetiana Safonova, Iurii Tarasov

Abstract: In data mining, efforts have focused on finding methods for efficient and effective cluster analysis in large databases. Active themes of research focus on the scalability of clustering methods, the effectiveness of methods for clustering complex shapes and types of data, high-dimensional clustering techniques, and methods for clustering mixed numerical and categorical data in large databases. One of the most accuracy approach based on dynamic modeling of cluster similarity is called Chameleon. In this paper we present a modified hierarchical clustering algorithm that used the main idea of Chameleon and the effectiveness of suggested approach will be demonstrated by the experimental results.

Keywords: Chameleon, clustering, hypergraph partitioning, coarsening hypergraph.

ACM Classification Keywords F.2.1 Numerical Algorithms and Problems

Introduction

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group in many applications. Data clustering is under vigorous development. Contributing areas of research include data mining, statistics, machine learning, spatial database technology, biology, and marketing. Owing to the huge amounts of data collected in databases, cluster analysis has recently become a highly active topic in data mining research. As a branch of statistics, cluster analysis has been studied extensively for many years, focusing mainly on distance-based cluster analysis. Active themes of research focus on the scalability of clustering methods, the effectiveness of methods for clustering complex shapes and types of data. Chameleon is a clustering algorithm that explores dynamic modeling in hierarchical clustering. In its clustering process, two clusters are merged if the interconnectivity and closeness between two clusters are highly related to the internal interconnectivity and closeness of objects within the clusters. The merge process based on the dynamic model facilitates the discovery of natural and homogeneous clusters and applies to all types of data as long as a similarity function is specified. Chameleon is derived based on the observation of the weakness of two hierarchical clustering algorithms: CURE and ROCK. CURE and related schemes ignore information about the aggregate interconnectivity of objects in two different clusters, whereas ROCK and related schemes ignore information about the closeness of two clusters while emphasizing their interconnectivity. In this paper, we present our experiments with hierarchical clustering algorithm CHAMELEON for circles cluster shapes with different densities using hMETIS program that used multilevel k-way partitioning for hypergraphs and a Clustering Toolkit package that merges clusters based on a dynamic model. In CHAMELEON two clusters are merged only if the inter-connectivity and closeness between two clusters are comparable to the internal inter-connectivity of the clusters and closeness of items within the clusters. The methodology of dynamic modeling of clusters is applicable to all types of data as long as a similarity matrix can be constructed. We present a modified hierarchical clustering algorithm that measures the similarity of two clusters based on a new dynamic model with different shapes and densities. The merging process using the dynamic model presented in this paper facilitates discovery of natural and homogeneous not only circles cluster shapes.

1 Related work

In this section, we give a brief description of existing clustering algorithms.

A hierarchical method creates a hierarchical decomposition of the given set of data objects. A hierarchical method can be classified as being either agglomerative or divisive, based on how the hierarchical decomposition is formed. The agglomerative approach, also called the bottom-up approach, starts with each object forming a separate group. It successively merges the objects or groups close to one another, until all of the groups are merged into one, or until a termination condition holds. The divisive approach, also called the top-down approach, starts with all the objects in the same cluster. In each successive iteration, a cluster is spitted up into smaller clusters, until eventually each object is in one cluster, or until a termination condition holds.

Hierarchical methods suffer from the fact that once a step is done, it can never be undone. This rigidity is useful in that it leads to smaller computation costs by not worrying about a combinatorial number of different choices. However, a major problem of such techniques is that they cannot correct erroneous decisions. There are two approaches to improving the quality of hierarchical clustering: (1) perform careful analysis of object "linkages" at each hierarchical partitioning, such as in CURE and Chameleon, or (2) integrate hierarchical agglomeration and iterative relocation by first using a hierarchical agglomerative algorithm and then refining the result using iterative relocation, as in BIRCH [Zhang, 1996].

Most clustering algorithms either favor clusters with spherical shape and similar sizes, or are fragile in the presence of outliers. CURE overcomes the problem of favoring clusters with spherical shape and similar sizes and is more robust with respect to outliers. CURE employs a novel hierarchical clustering algorithm that adopts a middle ground between centroid-based and representative-object-based approaches. Instead of using a single centroid or object to represent a cluster, a fixed number of representative points in space are chosen. The representative points of a cluster are generated by first selecting well-scattered objects for the cluster and then "shrinking" or moving them toward the cluster center by a specified fraction, or shrinking factor. At each step of the algorithm, the two clusters with the closest pair of representative points (where each point in the pair is from a different cluster) are merged. ROCK is an alternative agglomerative hierarchical clustering algorithm that is suited for clustering categorical attributes. It measures the similarity of two clusters by comparing the aggregate interconnectivity of two clusters against a user-specified static interconnectivity model, where the interconnectivity of two clusters is defined by the number of cross links between the two clusters, and link is the number of common neighbors between two points. In other words, cluster similarity is based on the number of points from different clusters who have neighbors in common [Guha, 1999].

ROCK first constructs a sparse graph from a given data similarity matrix using a similarity threshold and the concept of shared neighbors. It then performs a hierarchical clustering algorithm on the sparse graph.

There are two major limitations of the agglomerative mechanisms used in existing schemes. First, these schemes do not make use of information about the nature of individual clusters being merged. Second, one set of schemes (CURE and related schemes) ignore the information about the aggregate interconnectivity of items in two clusters, whereas the other set of schemes ignore information about the closeness of two clusters as defined by the similarity of the closest items across two clusters.

2 Overview of CHAMELEON: Clustering Using Dynamic Modeling

Chameleon is a clustering algorithm that explores dynamic modeling in hierarchical clustering [Karypis, 1999a]. Chameleon represents its objects based on the commonly used k-nearest neighbor graph approach. This graph representation of the data set allows CHAMELEON to scale to large data sets. Each vertex of the k-nearest neighbor graph represents a data object, and there exists an edge between two objects if one object is among the k-most similar objects of the other. The k-nearest neighbor graph captures the concept that neighborhood radius of an object is determined by the density of the region in which this object resides [Mitchell, 1997].

During the next step a sequence of successively smaller hypergraphs are constructed – Coarsening Phase. Two primary schemes have been developed for selecting what groups of vertices will be merged together to form single vertices in the next level coarse hypergraphs. The first scheme called edge- coarsening (EC) [Alpert, 1997] selects the groups by funding a maximal set of pairs of vertices (i.e., matching) that belong in many hyperedges. The second scheme that is called hyperedge-coarsening (HEC) [Karypis, 1997] finds a maximal independent set of hyperedges, and the sets of vertices that belong to each hyperedge becomes a group of vertices to be merged together. At each coarsening level, the coarsening scheme stop as soon as the size of the resulting coarse graph has been reduced by a factor of 1.7[Karypis, 1999b]. The third phase of the algorithm is to compute a k-way partitioning of the coarsest hypergraph such that the balancing constraint is satisfied and the partitioning function as mincut is optimized. During the fours phase - uncoarsening phase, a partitioning of the coarser hypergraph is projected to the next level finer hypergraph, and a partitioning refinement algorithm is used to optimize the objective function without violating the partitioning balancing constraints. At the final iteration of algorithm CHAMELEON determines the similarity between each pair of clusters by taking into account both at their relative inter-connectivity and their relative closeness. It selects to merge clusters that are well inter-connected as well as close together with respect to the internal inter-connectivity and closeness of the clusters. By selecting clusters based on both of these criteria, CHAMELEON overcomes the limitations of existing algorithms that look either at the absolute inter-connectivity or absolute closeness.

3 Performance Analysis

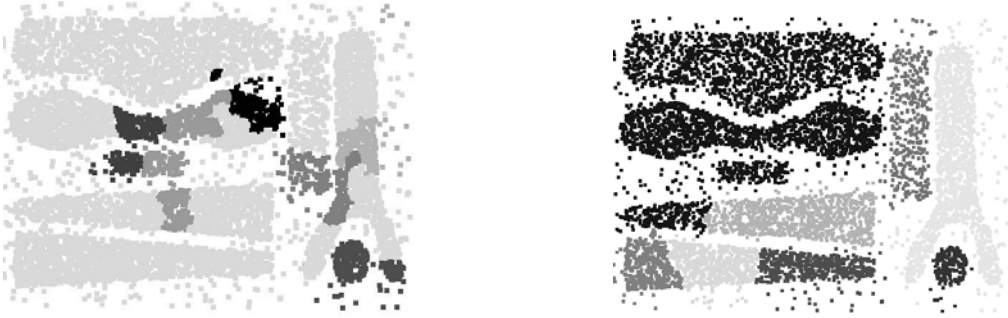
The overall computational complexity of CHAMELEON depends on the amount of time it requires to construct the K – nearest neighbors graph and the amount of time it requires to perform the two phases of the clustering algorithm. In [Karypis, 1999a] was shown that CHAMELEON is not very sensitive of values k for computing the k-nearest neighbor graph, of the value of MINSIZE for the phase I of the algorithm, and of scheme for combining relative inter-connectivity and relative closeness and associated parameters, and it was able to discover the correct clusters for all of these combinations of values for k and MINSIZE. In this section, we present experimental evaluation of clustering using hMETIS hypergraph partitioning package for k-way partitioning of hypergraph and for recursive bisection [Karypis, 1998] and CLUTO 2.1.1– A Clustering Toolkit [Karypis, 2003].

a) k-way partitioning by hMETIS b) final clusters by CLUTO

Fig.1 Data set "disk in disk" with k=5 nearest neighbors and asymmetric k-NN

We experimented with five different data sets containing points in two dimensions: "disk in disk", t4.8k, t5.8k, t8.8k, t7.10k [Karypis lab.].The first data set, has a particularly challenging feature that two clusters are very close to each other and they have different densities and circles shapes. We choose the number of neighbors k=5, 15, 40, MINSIZE = 5%. Looking at Figure 1, a) we can see the results of the k-way partitioning of hypergraph by hMETIS package [Karypis lab.] and b) merging process by CLUTO package [Karypis lab.] with k=5 nearest neighbors. Looking at Fig.1 we can see that in both cases we have not correctly identified the genuine clusters.

The data set t8.8k has eight clusters of different shapes, size and orientation, some of which are inside the space enclosed by other clusters. Moreover, it also contains random noise such as a collection of points forming vertical streaks. Looking at Fig. 2 with k=5 nearest neighbors we can see that hMETIS also compute k-way partitioning of hypergraph with mistakes closer to the border of two classes and CLUTO can not effectively merge clusters for such type of dataset using asymmetric k-NN, with k=5. It means that algorithm of the partitioning phase is very sensitive to the value of k for spherical shapes of clusters and to the types of k-NN graph (symmetric and asymmetric). It is very important to choose an optimal value of k, because with k=16 and more, and only for symmetric k-NN with weights of edges equal to the number of common neighbors we obtain final clustering with minimum percentages of errors.



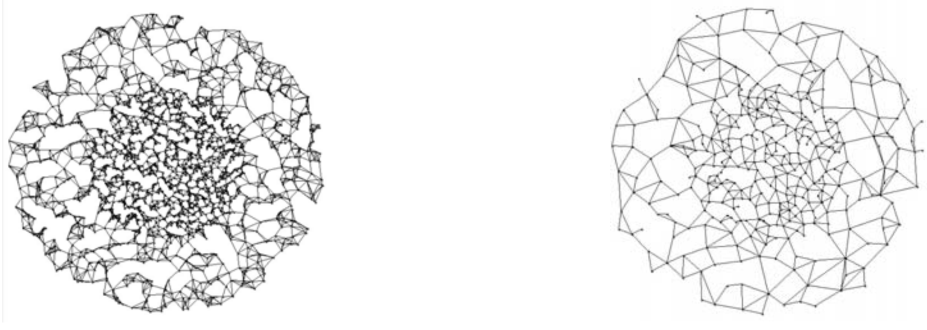
a) k-way partitioning by hMETIS b) final clusters by CLUTO

Fig.2 Data set “t8.8k” with k=5 nearest neighbors and asymmetric k-NN

4 Modeling the cluster similarity

As we remark above the CHAMELEON operates on a sparse graph in which nodes represent data items and weighted edges represent similarities among the data item (symmetric graph) [Karypis, 1999a]. In our algorithm during first phase we construct an asymmetric k-NN graph and there exists an edge between two points if for one of it there exist closest neighbor among all existing neighbors according to the value of k. Note that the weight of an edge connecting two objects in the k-NN graph is a similarity measure between them, as usual a simple distance measure (or inversely related to their distance).

In our algorithm the weight of an edge we compute as weighted distance between objects. Fig. 3 represents the k-NN graph for data set “disk in disk” with k=5. During coarsening phase the set of smaller hypergraphs is constructed. In the first stage of coarsening process we choose the set of vertices with maximum degrees and matched it with a random neighbor. On the other stages we visit each vertex in a random order and matched it with adjacent vertex via heaviest edge. Note that usually the weight of an edge connecting two nodes in a coarsened version of the graph is the number of edges in the original graph that connect the two sets of original nodes collapsed into the two coarse nodes. In our case we compute the weight of the hyperedge as the sum of the weights of all edges that collapse on each other during coarsening step. We stop the coarsening process at each level as soon as the number of multivertices of the resulting coarse hypergraph has been reduced by a constant less then 2 (Fig. 3).



a) k-NN graph with k=5 b) hypergraph after third level of coarsening

Fig 3 Data set “disk in disk”

On the next level of algorithm we produce a set of small hypergraphs using k-way multilevel paradigm [Karypis, 1999b]. We start the process of partitioning by choosing k most heavier multivertices, where k = 8, 16, 32. After that we gathering one by one all neighbors from each chosen vertex and obtain the initial partitioning w.r.t to the balancing constant. The problem of computing an optimal bisection of a hypergraph is NP-hard. One of the most commonly used objective function is to minimize the hyperedge-cut of the partitioning; i.e., the total number of hyperedges that span multiple partitions [Karypis, 1999b]. One of the most accuracy algorithm of partitioning the hypergraph is Kernighan-Lin / Fiduccia – Mattheyses algorithm, in which during each pass, the algorithm repeatedly finds a pair of vertices, one from each of the subdomains, and swaps their subdomains. The pairs are selected so as to give the maximum improvement in the quality of the partitioning even if this improvement is negative). Once a pair of vertices has been moved, neither are considered for movement in the rest of the pass. When all of the vertices have been moved, the pass ends. At this point, the state of the bisection at which the minimum edge-cut was achieved is restored. In our experiments we use a greedy refinement algorithm developed by George Karypis [Karypis, 1999b], but as the gain function for each vertex we compute the differences between the sum of the weights of edges incident on vertex that go to the other partition and the sum of edges weights that stay within the partition. We choose the vertex with maximum positive gain and move it if it result in a positive gain, so we works only with boundary vertices.

a) Data set "disk in disk" b) Data set "t8.8k.txt"

Fig 4. Clustering results using a new approach to the sub-clusters merging, k=5

After the partitioning of hypergraph into the large number of small parts we start to merge the pair of clusters for which both relative inter-connectivity and their relative closeness are high [Karypis, 1999a]. In our research we use George Karypis formula to compute the similarity between sub-clusters. Looking at the Fig.1 b we can see that for data set "disk in disk" was obtained not correct clustering results. Thus we suggests to modified the above mentioned expression by change the relative inter-connectivity to a new expression that estimate the average weights of edges in each sub-graph and the number of edges that connect two partitions to the number of edges that stay within the smallest partition.

Looking at the Fig.4 we can see the correct clustering results for the same data set "disk in disk" using our suggested expression. For another above mentioned data sets we obtain as well accuracy results. In all experiment we use k=5 and in our approach the correctness of classification really doesn't depend of the value of k and of the k-NN type.

Conclusion

In this paper, we present our experiments with hierarchical clustering algorithm CHAMELEON for circles cluster shapes with different densities using hMETIS program that used multilevel k-way partitioning for hypergraphs and a Clustering Toolkit package that merges clusters based on a dynamic model. In CHAMELEON two clusters are merged only if the inter-connectivity and closeness between two clusters are comparable to the internal inter-connectivity of the clusters and closeness of items within the clusters. The methodology of dynamic modeling of clusters is applicable to all types of data as long as a similarity matrix can be constructed.

Experimental results showed that hMETIS compute k-way partitioning of hypergraph with mistakes closer to the border of two classes and CLUTO cannot effectively merge clusters using asymmetric k-NN, with k=5.

We present a modified hierarchical clustering algorithm that measures the similarity of two clusters based on a new dynamic model with different shapes and densities. The merging process using the dynamic model presented in this paper facilitates discovery of natural and homogeneous not only circles cluster shapes.

Experimental results showed that this method is not sensitive to the value of k and doesn't need a specific k-nearest neighbor graph creating.

Bibliography

[Alpert, 1997] C. J. Alpert, J. H. Huang and A. B. Kahng, Multilevel circuit partitioning. In: Proc. of the 34th ACM/IEEE Design Automation Conference. 1997.

[Guha, 1999] S. Guha, R. Rastogi, K. Shim ROCK: Robust Clustering using linKs, (ICDE'99)

[Karypis, 1997] G. Karypis, R. Aggarwal, V. Kumar and Sh. Shekhar, Multilevel hypergraph partitioning: Application in VLSI domain. In: Proceedings of the Design and Automation Conference. 1997.

[Karypis, 1998] G. Karypis, and V. Kumar, hMETIS 1.5.3: A hypergraph partitioning package. Technical report. Department of Computer Science, University of Minnesota, 1998

[Karypis, 1999a] G. Karypis, E.-H. Han, and V. Kumar. CHAMELEON: A Hierarchical Clustering Algorithms Using Dynamic Modeling. IEEE Computer, 32(8):68–75, 1999.

[Karypis, 1999b] G. Karypis and V. Kumar. Multilevel k-way hypergraph partitioning. In Proceedings of the Design and Automation Conference, 1999.

[Karypis, 2003] G. Karypis, CLUTO 2.1.1. A Clustering Toolkit. Technical report. Department of Computer Science, University of Minnesota, 2003

[Karypis lab.] http://www.cs.umn.edu/˜karypis.

[Mitchell, 1997] T. M. Mitchell. Machine Learning. McGraw Hill, 1997

[Zhang, 1996] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH : an efficient data clustering method for very large databases. SIGMOD'96.

Authors' Information

Shatovska Tetyana – Kharkiv National University of Radio Electronics, Computer Science department, P.O.Box: Kharkiv-116, Lenin av. 14, Ukraine; e-mail: tanita_uk@mail.ru

Safonova Tetiana - Kharkiv National University of Radio Electronics, Computer Science department, P.O.Box: Kharkiv-116, Lenin av. 14, Ukraine; e-mail: safonovatm@mail.ru

Tarasov Iurii - Kharkiv National University of Radio Electronics, Computer Science department, P.O.Box: Kharkiv-116, Lenin av. 14, Ukraine; e-mail: ytarasovmail@rambler.ru

TECHNOLOGY OF STORAGE AND PROCESSING OF ELECTRONIC DOCUMENTS WITH INTELLECTUAL SEARCH PROPERTIES

Yuri Kalafati, Konstantin Moiseyev, Sergey Starkov, Svetlana Shushkova

Abstract: The technology of record, storage and processing of the texts, based on creation of integer index cycles is discussed. Algorithms of exact-match search and search similar on the basis of inquiry in a natural language are considered. The software realizing offered approaches is described, and examples of the electronic archives possessing properties of intellectual search are resulted.

Keywords: dynamic systems, associative search, integer index cycles, text indexation, archiving and retrieval.

Introduction

The use of dynamic systems for record and data processing was first offered in [Dmitriev etc., 1991]. The basic idea consists in the fact that a correspondence is put between a given set of information blocks and a set of limiting cycles of a discrete nonlinear dynamic system: one-dimensional [Dmitriev, 1991], [Andreyev etc., 1992], [Andreyev etc., 1997] or multidimensional [Andreyev etc., 1994] maps. The final sequence of symbols of a certain alphabet is meant by the information block and each symbol is put in conformity with the dynamic system variable. This approach has been realized to record the text and graphic information by means of piecewise-linear maps and the scope for associative searching of information by its fragments is shown. The research has been developed further [Andreyev etc., 1999] and to optimize the search an integer index was suggested to be used instead of dynamic variables. In spite of the fact that the principles of information representation with the integer index cycles are well-known and successfully applied for data storage in various DB, the creation of electronic archives possessing the opportunities of intellectual search, including associative search, exact-match search, search similar to inquiries in the natural language is of special interest and is to be developed. Consider some key moments of a technology being developed.

Indexation algorithm

Assume, the first page presented by the next text line is transmitted for indexation:

a b c a c a d a b a c a b c (1)

As a first step, the symbol – an attribute of the beginning of a page is added to the initial alphabet. This symbol will have the number 256. Transform the initial text to a cyclic sequence of symbols and add a symbol 256 to the end of a sequence of symbols. To avoid excessive information, the concept of an expanded alphabet is entered in the initial sequence. Each element of an expanded alphabet is based on two already existing symbols. For example, if it is necessary to escape the repeated sequence of symbols a b c, a new symbol 257 = ' a ' + ' b ' and 258 = 257 + ' c ' is entered and so the sequence a b c in the initial text is replaced with a symbol 258. The text (1) is indexed using the above concept. Let us search the repeated pairs of symbols in the initial text. Take the pair of symbols from a given example– the pair a b. As a given pair of symbols is found more than once in the text, we add this pair to the expanded alphabet:

Symbol 257 = ' a ' + ' b '

Transform the initial sequence of symbols, having replaced the pair symbols a b by a new symbol with the index 257. We have:

257 c a c a d 257 a c 257 c 256 (2)

Then find the recurrence of other pairs and create new symbols. Based on this sequence, create a page description array, each element of which consists of three symbols: a_n, a_{n+1}, a_{n+2} , and order the array obtained in a pair (a_n, a_{n+1}). A necessary condition for system operation is the uniqueness of a pair (a_n, a_{n+1}) in the array formed. The following actions are to be done in adding the next page:

The text transmitted for indexation is processed using the expanded alphabet formed earlier. First there is a search in the added text of a pair of symbols corresponding to the element 257 in the expanded alphabet. If such pairs are found, they are replaced by the symbol 257. Then there is a search of a pair of symbols corresponding to the element 258 and so on.

The repeated pairs of symbols are searched in the sequence of symbols obtained. If these are found, a new element of the expanded alphabet is created and the corresponding pairs are replaced with a new element.

Then each pair of symbols from the sequence created is being searched in the page description array available. If such a pair is found, a new symbol of the expanded alphabet is created and the corresponding pair of symbols in the added text is replaced by a new symbol. In the program realization available there also proceed the change in already existing array and the replacement of a pair of symbols with the newly created symbol. The description of this algorithm updating is beyond the scope of this paper.

Transform the obtained cyclic sequence to a set of elements for a page description array. Add these elements to an array and reorder it.

The expanded alphabet is presented as a two-dimensional integer array. An array index is the number of a symbol in the expanded alphabet. The first element of an array has the index 257. Each couple of integers stored in the array element are components of the corresponding alphabet symbol. The ordered page description array is the array consisting of elements (a_n, a_{n+1}, a_{n+2}). This array is ordered in a pair (a_n, a_{n+1}) and this pair is unique within the whole indexed text *sets.* Input points *on indexed pages.* Any pair of symbols of each page representation is stored as a two-dimensional array of integers. Array index is the number of the page indexed, and the array element content is the information necessary and sufficient to begin the procedure of extracting corresponding text page.

Text extraction from the page

To extract the text page from the constructed index it is sufficient to have information about any pair of symbols in the expanded alphabet (a_{n0}, a_{n0+1}) contained in the page. It should be remembered that the page description array is ordered in the first pair (a_n, a_{n+1}) and the condition required. i.e. the uniqueness of this pair in the whole file index, is fulfilled. Using the procedure of binary search we are searching the pair (a_{n0}, a_{n0+1}) in a page description array. From the element found in a file we take an element a_{n0+2}. Then in the page description array we search the pair (a_{n0+1}, a_{n0+2}). From the element found in a file we take the element a_{n0+3} and so on. We repeat this operation until after search of the pair (a_i, a_{i+1}) and extraction of the element a_{i+2} the equality: $a_{i+1} = a_{n0}$ and $a_{i+2} = a_{n0+1}$ is true.

Thus, we have obtained the cyclic sequence describing the page text in symbols of the expanded alphabet. Then based on the procedure of decoding the symbol of then expanded alphabet (see below) we transform the cyclic sequence presented in symbols of the expanded alphabet in a sequence of ASCII symbols.

Description of base algorithm search

To demonstrate the principal opportunity of information search with a given way of information indexation the following mechanism is suggested. A sentence or even a text paragraph is taken as an input for searching. The search inquiry is coded by means of the expanded alphabet available. Assume that the search inquiry was transformed to the sequence $a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8$ of elements of the expanded alphabet. Take the first pair of the search inquiry (a_1, a_2) and search for it in the ordered page description array. If such a pair of symbols is not found, pass to the pair of symbols (a_2, a_3) and so on. In case of a successful search, for example, of the pair (a_2, a_3), we have (a_2, a_3, a_x) an element of the page description array. Compare the element a_4 and the element a_x. If these elements coincide, find (a_3, a_4, a_y) the element of a page description array. If elements a_5 and a_y coincide, the search is considered to be successful.

Two versions of search are realized based on the suggested indexation algorithm of text information:

Exact-match search

A given version of search is intended for solving a problem of search by a word or by a short search inquiry. It allows to find all inputs of search inquiry in the indexed text array. According to the offered algorithm of indexation there are some possible variants for line representation corresponding to the search inquiry inside the indexed file. First, the whole line of a search inquiry can be entered inside a symbol of the expanded alphabet. All such symbols as well as symbols created on their basis are to be searched in a page description array. Each found element of a page description array which is the input point to one of the pages of the file indexed is considered to be the result of search.

Second, the line, corresponding to the search inquiry in the indexed page, can be divided into two parts, one of which is the end and the second is the beginning of one of symbols of the expanded alphabet. To find all such results of search we should construct two files of symbols. The first one is an array of all symbols of the expanded alphabet terminating with a given line of ASCII symbols. The second one is an array of all symbols of the expanded alphabet beginning with a given line of ASCII symbols. The corresponding pair of symbols should be searched for each element of the first array and each element of the second array in the page description array. In successful search the found element of a page description array is a result of search. Then the required line can consist of three parts. The first part is the end and the third one is the beginning of one of the symbols of the expanded alphabet. The second part is one of the elements of the expanded alphabet. As in the previous case, compile two files of symbols for the first and third parts of the search inquiry. For each symbol of the first part we create a pair of symbols formed by an element of this array and a symbol of the expanded alphabet corresponding to the second part of the search inquiry and search this pair in a page description array. If search is successful, search the third symbol of a file element in the file of symbols constructed for the third part of search inquiry. In case of success the found element of a page description array is considered to be the result.

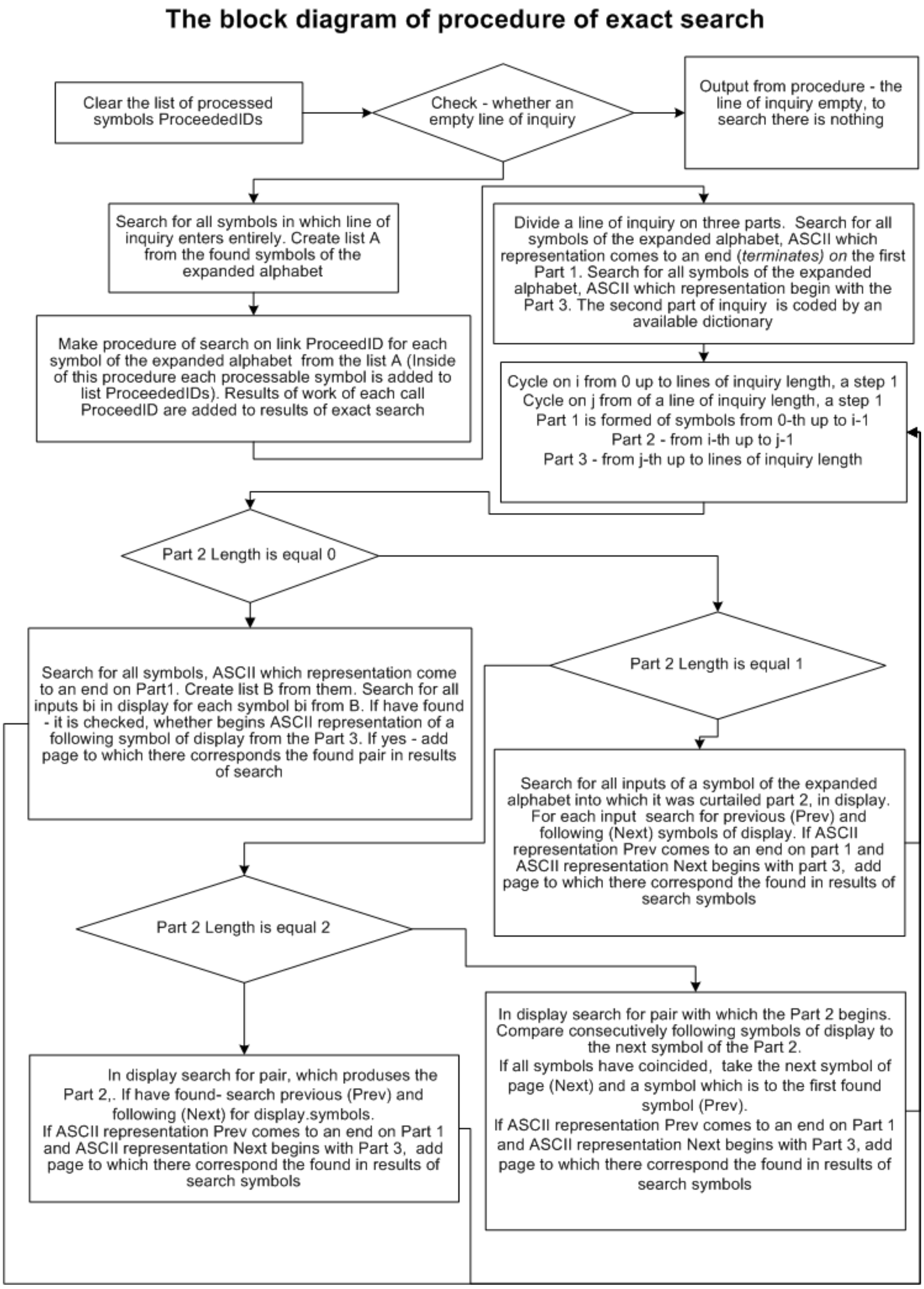


Fig. 1. Procedure of exact-match search.

Finally, if the number of parts by which a desired line in the indexed page is presented exceeds three, we can search the symbols displaced in the centre of the line. Assume that the line is presented by five parts. The first and fifth parts are, accordingly, the end and the beginning of symbols of the expanded alphabet. Each of the remaining parts is presented by one of the symbols of the expanded alphabet. First, we search for the pair of the symbols formed by the second and third parts. If this part is found in a page description array, the third symbol of this array element can be compared with the fourth part. If they coincide, a pair of symbols formed by the third and the forth parts is being searched in a page description array. Take the third element from the found element of a file and we search it in the fifth part, i.e. in a file of symbols beginning from with the end of search inquiry. If search is successful, it is necessary to compare only a part of the found page which is before the second part of a line with the first part. In case of success we have the result of search. Fig.1 presents a block-diagram of the exact search procedure.

Search of similarity

This variant of search is meant for finding information closest to the search inquiry. Offer in the natural language, a paragraph or even the whole page of the text can be transmitted as the search inquiry. The search inquiry transferred to the input of search of similar is coded by means of the expanded alphabet available.

On the basis of a list of symbols for each indexed page the following sum is calculated:

$$P_i = \sum_{k=1}^{N} (length_k)^{\alpha} * (count_k)^{\beta} ,$$

where $length_k$ is the length in ASCII symbols of the k-th element in a list of symbols, $count_k$ is the quantity of the k-th element in a list in page i, and are the external parameters. Then, the obtained Pi values are ordered and pages with the highest values are given to the user as results of search.

Conclusion

Now a described algorithm of text processing and algorithms of text-through search are realized and used in CCT Archive and CCT Publisher Companies Controlling Chaos Technologies software products. Software products are intended for the creation of electronic archives of not structured documents with an opportunity of text-through information search, and for creation and preparation for CD and DVD electronic books, encyclopedias, archives of magazines. Examples of successful application of software products are the electronic archives of magazines « Chemistry and the Life », "Quantum", "Knowledge-force".

Fig. 2. gives results of search system operation with electronic archive of magazine " Quantum " as an example. At the upper left is inquiry in the natural language on which the search was carried out, below is the ranged list of the documents found. To the left is the document page with the allocated inputs.

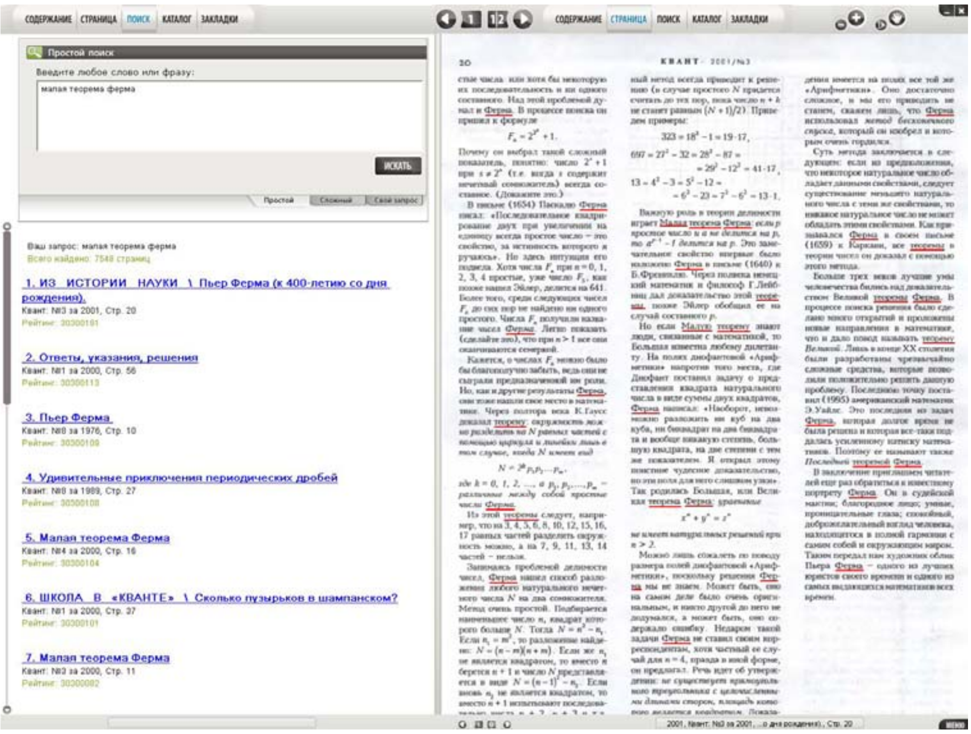


Fig. 2. Search system operation with electronic archive as an example

Below are the basic time characteristics managed to be reached with the present program realization of the algorithms described. All values are obtained using an ordinary personal computer, by the text size we mean the number of ASCII symbols in a text but not the size of files containing this text.

The maximal size of the indexed text is about 100 Mb

Text indexation rate is about 1 Mb per min (the average indexation rate 100 Mb)

Time of index opening is not more than 1 min.

Search time is about 1 sec.

It should be noted that the technology being developed is not language dependent and can be adjusted to any language systems. Development of ideas put in searching the similar allows one to solve such problems as search of plagiarism, *rubrication* and *text clusterization*, Internet content filtration and anti-spam system creation.

Bibliography

[Dmitriev etc., 1991] Dmitriev A.S., Panas A.I., Starkov S.O. Storing and recognition information based on stable cycles of 1-D map // Phys. Lett. A.–1991–. V.155, –№8–pp.494-499.

[Dmitriev, 1991] Dmitriev A.S. Record and recognition of information in one-dimensional dynamic systems // РЭ. –1991–V.36– №1–p. 101-108.

[Andreyev etc., 1992] Andreyev Y.V., Dmitriev A.S., Chua L.O., Wu C.W. Associative and random access memory using one-dimensional maps // IJBC–1992–V.3–№3–pp.483-504.

[Andreyev etc., 1997] Andreyev Yu. V., Dmitriev A.S., and Starkov S.O. Information Processing in 1-D Systems with Chaos, *IEEE Transactions on Circuits and Systems*, 1997, vol. 44, No. 1, pp. 21-28.

[Andreyev etc., 1994] Andreyev J.V., Belskij J.L., Dmitriev A.S. Record and restoration of the information with the application of steady cycles of two-dimentional multivariate displays // РЭ–1994–V.39–№1–p.114-123.

[Andreyev etc., 1999] Andreyev Yu. V., Dmitriev A.S., and Ovsyannikov A.V. Information searching system "Forget-Me-Not" based on complex dynamics of nonlinear systems, *Proc. 7th Int. Workshop NDES-99*, 1999. Ronne, Denmark. pp.273-276.

Authors' Information

Kalafati Yuri Dmitrievich – Senior Scientist, PhD, Institute or Radio Engineering & Electronics, RAS, Moscow; e-mail: kalafati@controlchaostech.com

Moiseyev Konstantin Vladimirovich – Director, Controlling Chaos Technologies, Moscow; e-mail: moiseyev@controlchaostech.com

Starkov Sergey Olegovich – Professor, Obninsk State Technical University of Nuclear Power Engineering, Kaluga region, Obninsk; email: Starkov@iate.obninsk.ru

Shushkova Svetlana Alexandrovna – Undergraduate, Obninsk State Technical University of Nuclear Power Engineering, Kaluga region, Obninsk; e-mail: shushkovasvetlana@ramler.ru

COMMON SCIENTIFIC LEXICON FOR AUTOMATIC DISCOURSE ANALYSIS OF SCIENTIFIC AND TECHNICAL TEXTS *

Elena Bolshakova

Abstract: The paper reports on preliminary results of an ongoing research aiming at development of an automatic procedure for recognition of discourse-compositional structure of scientific and technical texts, which is required in many NLP applications. The procedure exploits as discourse markers various domain-independent words and expressions that are specific for scientific and technical texts and organize scientific discourse. The paper discusses features of scientific discourse and common scientific lexicon comprising such words and expressions. Methodological issues of development of a computer dictionary for common scientific lexicon are concerned; basic principles of its organization are described as well. Main steps of the discourse-analyzing procedure based on the dictionary and surface syntactical analysis are pointed out.

* The work is supported by the grant № 06-01-00571 of Russian Fond of Fundamental Researches (RFFI).

Keywords: scientific and technical prose, common scientific words and expressions, discourse markers, scientific discourse operations, discourse-compositional analysis.

ACM Classification Keywords: I.2.7 [Artificial Intelligence]: Natural language processing – Text analysis

Introduction

Intensive use of scientific phraseology is admittedly the most distinctive feature of functional style of scientific and technical prose [10]. The prose comprises documents of various genres and types: scientific papers, annotations, reviews, technical reports, etc. The phraseology includes both scientific and technical terms and various word expressions of common nature, such as English expressions *to test hypothesis, key concept, to be more precise, mentioned above, for this reason* etc. and Russian: *принятая гипотеза, по указанной причине, обосновать вывод, описанный ниже* and so on. Such lexical items are usually called *common scientific words and expressions* [5].

Both scientific terms and common scientific expressions are equally necessary to create a coherent and consistent scientific text. However, they differ in their functions within scientific discourse (scientific speech). Specific terms denote concepts, objects, and processes of particular scientific domains, whereas common scientific expressions are domain independent, they are used to organize scientific text narrative by expressing the logic of scientific reasoning, by structuring the text under development, by introducing cross-text references.

Lexicon of common scientific words and expressions is a syntactically quite heterogeneous set. It comprises, besides content (autosemantic) words, functional (auxiliary) words. Noun and verb-noun combinations, adverb and participle expressions, compound prepositions and conjunctions are included as well. Certain common scientific words and expressions are known as *discourse markers* [14, 15]: for instance, *Eng. in other words, Rus. другими словами*. Some word combinations are stable expressions exploited as ready-for-use colloquial formulas (clichés) [1], such as *Eng. as it was stated above*, *to outline directions of further research; Rus. из вышесказанного следует, как показало проведенное исследование*. It is worth noting that some clichés are common for scientific and technical prose, the others are specific for particular genres.

The paper describes the results of our research on elaboration of a computer dictionary of common scientific words and expressions, as well as on development of a procedure for discourse analysis of scientific and technical texts. The work is done within the overall framework of creating computational models of scientific and technical prose [2], supposing that really effective and deep automatic analysis of scientific and technical texts requires taking into account functional features of scientific and technical prose, in the first place, peculiarities of its phraseology and discourse.

The work started with an empirical study of scientific texts in several fields of exact sciences (mainly in computer science), so that scientific papers as the “core” of the functional style were analyzed. The study was initially performed for Russian scientific texts, and then expanded to English. In both languages the principal features of scientific discourse and lexicon proved to be the same.

As the work progressed, the importance of the common scientific lexicon became increasingly obvious, despite of its relatively small size. So we began to develop a computer dictionary comprising a wide range of common scientific words and expressions and providing a classification of their syntactic and semantic features. For Russian, the dictionary is now partially implemented; for English, only the classification work was done so far.

Our approach considers any lexical device signaling scientific discourse as discourse marker, and we include into our dictionary expressions specific for scientific and technical texts, such as English *by definition* or Russian *по определению*. As a result, the dictionary covers a wider set of lexical units than style-independent DiMLex lexicon of discourse markers [14] developed for German and English and mainly consists of conjunctions and conjunctive adverbs.

Instead of concept of discourse relationship proposed in well-known discourse theory RST [8] for explaining relations between adjacent phrases in text, we rely on the concept of *scientific discourse operation* fruitful for recognition of discourse-compositional structures specific for scientific and technical texts.

As units of common scientific lexicon may be served as surface cues, we assume the hypothesis that shallow text analysis based on the lexicon is adequate for detecting discourse structure of scientific text, without a deeper syntactical-semantic analysis of all its sentences. Our discourse-analyzing procedure is now under development; it differs from the procedures developed for Japanese texts [7, 11] and based on deep syntactical analysis of sentences and a relatively narrow set of style-independent discourse markers.

The paper starts with an overview of specific features of scientific discourse derived from our empirical study, which also determined the set of common scientific words and expressions we consider as discourse markers. Illustrative examples from both languages – English and Russian are given. Then organizing principles for the computer dictionary of common scientific lexicon are described; and basic steps of the procedure for recognition of discourse-compositional structure of a given scientific text are presented. Potential applications of the dictionary and the procedure are pointed out in conclusion.

Scientific Discourse and Common Scientific Lexicon

The global purpose of scientific communication is to convey new ideas and results of scientific research, as well as to explain and rationalize them. Therefore, scientific discourse (speech) involves reasoning that is organized as a sequence of mental operations of informing and arguing. Among these typical operations one can notice assuming hypotheses, defining new terms and concepts, exemplification, resuming and so on. We call such intellectual operations *scientific discourse operations*.

As a rule, these mental-discourse operations are introduced into texts and more or less explicitly marked by authors of the texts with the aid of lexical devices – *common scientific words and expressions*. Therefore the words and expressions have metatext function and are often called *discourse markers* [15]. Accordingly to the applied discourse operations, scientific text is composed of *discourse segments*; in general, each segment comprises several adjacent sentences and includes discourse markers.

The most evident markers of scientific discourse are ***mental performative expressions*** like *Eng. we conclude, we would assume* or *Rus. мы докажем, мы предположим*. For Russian, they are described in detail in [12]. Performative expressions are based on “mental” verbs, e.g., *Eng. to conclude, to consider, to admit, to propose; Rus. заметим, рассмотрим, выразим* and so on. As a rule, these verbs explicate particular steps of scientific reasoning and have valences (complementing arguments): *Eng. we consider N*, *we conclude that S*; *Rus. рассмотрим N*, *подчеркивается, что S*. Besides pure “mental” verbs (*to conclude, to assume,* etc.), verbs of physical action (*to see, to show,* etc.) are used as mental.

Various forms of mental performative expressions are identified in scientific texts: *canonical forms, verbal variants, impersonal forms, and descriptive variants*. Canonical forms are based on mental verbs in the second person plural, often with the corresponding pronoun: e.g., *Eng. we resume, let us proceed to, we will proceed; Rus. мы покажем, мы рассмотрим*.

Verbal variants – e.g., *Eng. summing up, strictly speaking; Rus. подводя итоги, строго говоря* – are often used together will canonical forms: e.g., *Eng. refining the definition, wee see that…; Rus. суммируя вышесказанное, укажем…* . Impersonal forms – such as *Eng. it should be added, it was found, it is reasonable to assume; Rus. необходимо/нетрудно заметить, представляется, что…* – often include words of author’s estimation (*should, reasonable, necessary*). Verbal and impersonal forms are used in texts to paraphrase canonical forms (e.g., *it was found* instead of *we found*) or to give some cross-text references (e.g., *as it was stated above*). Though they are less explicit forms than canonical, they are functionally equivalent.

One can also find in scientific and technical texts ‘hidden’ performatives, which we call descriptive variants: e.g., *These data are given in Table 3* stands for *We gave these data in Table 3*. Additional examples of descriptive variants are: *Eng. N is briefly described, N are given in; Rus. N кратко описано*.

Mental-discourse operations might be expressed by various parenthetic words and expressions: indicators of order (e.g., *Eng. first* or *lastly; Rus. во-первых, наконец*), markers of equivalency (e.g., *Eng. in other words; Rus. иными словами*), various connectives between textual parts (e.g., *Eng. nevertheless* or *so far; Rus. тем не менее, благодаря тому, что*) and so on. The metatext nature of these discourse markers is more obvious; they are used in texts of all styles.

As typical just for scientific and technical texts, we should point out abstract nouns, such as *problem*, *analysis*, *model, concept, conclusion* and so on. They are intended to name mental constructs by which scientific information is semantically structured. We call such nouns *common scientific variables*, since they have the obligatory attributive valence (*problem of N, model of N*). Common scientific variables are mainly used with mental performative verbs, thereby forming stable noun-verb combinations, such as *Eng. to test hypothesis* or *to draw conclusions; Rus. подвергнуть анализу, проводить аналогию, опровергнуть гипотезу* [5]. Meanings of such verbs are close to Mel'čuk's Lexical Functions [9] with corresponding nouns as arguments.

Common Scientific Lexicon and Scientific Discourse Operations

Based on our empirical study, we propose for discourse analysis a particular set of scientific discourse operations, the most significant are presented in the Table 1.

We also propose to use the set of scientific discourse operations for semantical classification of heterogeneous collection of common words and word combinations, i.e., to classify them according to their discourse-organizing functions in scientific texts, irrespectively of their grammatical form and syntactic features.

Table 1. Scientific discourse operations

Operation	Russian Examples	English Examples
Description or statement	*укажем, что; характеризуя*	*let us to describe; we point out that*
Elaboration or adding information	*в частности; в дополнение к*	*to be more precise; in addition*
Expressing relations of causal, conditional, and concession type	*по этой причине; следовательно*	*hence; provided that; however*
Actualization of the topic	*перейдем к; рассмотрим*	*as for; let us consider; regarding*
Emphasizing	*особо подчеркнем; необходимо отметить*	*first of all; it is necessary to emphasize*
Presupposition	*предположим/допустим, что*	*we would assume; it may be admitted*
Definition	*будем называть; по определению*	*by definition; we call it/them,*
Comparison	*по сравнению с*	*as compared with*
Contraposition	*с одной стороны*	*on the one hand; as opposite to*
Illustration or exemplification	*к примеру; например*	*as illustrated below; for example*
Generation or resuming	*суммируя вышесказанное; в общем*	*in general; summing up*
Enumeration or ordering	*во-первых; наконец*	*next; finally*
Labeling with a scientific variable	*идея; модель; результат*	*result; idea; model*
Expressing of author's attitude	*целесообразно считать; по всей видимости*	*in our opinion; it seems reasonable*

Two following text fragments taken from English and Russian texts illustrate the usage of various scientific discourse operations and corresponding common scientific words and expressions (they are underlined):

> *Использование VBA–функции в данном случае является весьма целесообразным. Поясним это на примере с изменением процентной ставки. Допустим, вместо собственной функции мы использовали бы Len… Однако мы сталкиваемся с серьезной проблемой – как формировать такую структуру? … Именно по этой причине здесь не используется эта функция, хотя, на первый взгляд, она была бы вполне уместна.* (1)

> *For dealing with discourse markers, we do not regard this distinction as particular helpful, though. As we have illustrated above and will elaborate below, these words can carry a wide variety of semantic and pragmatic overtones, which render the choice of a marker meaning-driven, as opposite to a mere consequence of structural decisions.* (2)

It should be pointed out that besides units of common scientific lexicon, non-lexical devices are used to organize scientific discourse. In particular, such devices as sections, paragraphs, items, rubrics, and numeration are intended to structure scientific texts and to form their composition. All structuring and discourse-organizing devices present an interconnected system: devices can complement or substitute one another. For example, section headings are really substitutes for performative expression *we proceed to*, whereas numeration often complements performative formulas: e.g. *Let us enumerate main statements: 1)…2)…* . This interconnected system is rather excessive, since for most discourse operations there exist collections of similar lexical markers allowing for flexible paraphrasing.

In general, some discourse operation with its lexical and non-lexical devices can be used to implement another operation. For example, for categorization, a definition of new term is often required. As a result, certain discourse segments are embedded into some others, and in this way hierarchical structure of scientific text is formed.

Compiling the Dictionary of Common Scintific Words and Expressions

To develop the computer dictionary, collections of Russian and English common scientific words and word combinations were compiled from few available text dictionaries of scientific phraseology [4, 5] and from scientific texts in several fields of science (mainly in computer science and artificial intelligence), through their manual scanning. While selecting a word or expression for our collection, we used the following non-formal criteria:

- discourse-organizing function of the word or expression should be evident;
- it should be rather frequently used in texts of several scientific fields.

In addition, inter-language correspondences were used: for Russian expressions English equivalents were looked for, and vice versa.

Each compiled collection of common scientific words and expressions (for Russian and English) was divided into functional classes in accordance with the proposed list of discourse operations. Within each class, all words and word combinations that are semantically close and interchangeable in the texts as discourse markers were gathered into a group, thereby giving a subclass of functionally equivalent markers. Each group of functional equivalence often includes words of different parts of speech and contains from 2 to 9 units, the number depending on the language. For example, the resulted group of the consequence relationship includes for English: *hence, therefore, as a result, consequently, it follows that, we conclude that* etc., and for Russian: *значит*, *итак*, *таким образом*, *тем самым, как видим* etc. For both English and Russian, we obtain 53 groups corresponding to particular discourse organizing functions.

We should note that resulted groups of expressions differ in nature of objects being marked: while expressions of some groups indicate particular relations between text segments, the other mark particular text sentences or text segments.

Semantic and Syntactic Information in the Dictionary

To develop lexical entries of our computer dictionary, we considered requirements for its use by automatic text processing system, first of all, by discourse-analyzing procedure. The dictionary contains both units that correspond to words of common scientific lexicon and units representing word combinations. The former comprises all words of the lexicon, including those encountered in texts only within scientific expressions.

For a particular word, each unit stores adequate morphosyntactic information, including the part of speech and the flexional class (if any), as well as pointers to dictionary units describing available combinations with this word. In turn, unit for a particular word combination represents information about syntactical properties of the combination: stable *vs*. free, continuous *vs*. discontinuous.

For each dictionary unit considered as discourse marker, our computer dictionary provides semantic information that facilitates recognition of underlying discourse operation, namely:

- Functional class and group of the unit within the proposed semantic classification;
- Contextual conditions necessary for unit to be discourse marker within texts;
- Information about size and boundaries of implied discourse segment (one or several sentences; the beginning or the end of discourse segment is to be marked by this unit).

As most word combinations of common scientific lexicon have syntactic valences, we propose to represent information about valences with the aid of special ***lexicosyntactic patterns*** [3]. Each lexicosyntactic pattern fixes lexemes (constituent words of the particular combination) and their grammatical forms, as well as specifies syntactic conditions necessary for filing its empty slots (valences of the fixed lexemes). An example of such a pattern is *“let us consider” NP* with *NP* denoting a noun phrase; this pattern describes English expression that corresponds to discourse operation of topic actualization. Another example is *NP “we will call” T*, where *T* denotes an author’s term and *NP* is a noun phrase explaining its meaning; the pattern represents the typical English expression for definition of new terms.

Lexicosyntactic patterns proved to be a convenient device for describing stable colloquial expressions comprising both phrasal formulas (like Eng. *the paper describes main features of*, *argument can be made against*) and predicative constructs (such as *to take as starting point for*). So a formal language for specifying lexicosyntactic patterns was elaborated, as well as a methodology for acquiring new patterns for the particular discourse operation from scientific and technical texts. Based on the acquiring methodology, a collection of patterns was created, which describes typical Russian single-sentence definitions of new terms. An example of lexicosyntactic pattern for discourse operation of defining a new term is given below:

«под» NP1 <case=ins> V<пониматься; tense=pres, person=3> NP2 <case=nom> < NP1.numb=V.numb>

where particular lexemes of the pattern are quoted, letter *V* denotes the verb, NP1 and NP2 denote noun phrases, and grammatical conditions are written within angle brackets – they specify values of grammatical parameters (tense, person, case, number) or establish their equality. The pattern describes typical Russian expression often encounted in scientific and technical texts, e.g. *Под графемной конструкцией понимается графическая форма, построенная из базисных, проблемно-ориентированных и/или графических конструкций* (fixed lexemes of the pattern are underlined) or *Под данными при такой формализации понимаются последовательности символов в некоторых алфавитах*.

Automatic Recognition of Discourse Structure

We consider discourse-compositional structure of scientific text as hierarchical structure of sequenced and embedded discourse segments, which corresponds to applied discourse operations and applied structuring devices. The structure may be represented as a tree, with tree nodes corresponding to discourse segments, and tree links fixing semantic (e.g., causal) and structural (e.g., embedding) relations between segments.

Our study of scientific discourse and common scientific lexicon showed that the lexicon has its own functional semantics, which makes it possible to superficially read scientific texts, i.e. to identify underlying discourse operations, to derive discourse-compositional structure of the texts, and thereby to comprehend logic of scientific reasoning, without deep understanding of these texts. So we are developing our procedure for recognition of discourse-compositional structure of scientific texts on the basis of shallow text analysis and the described computer dictionary. In order to reconstruct discourse-compositional tree for a given text, the recognition procedure takes the following steps:

1. Grapheme analysis of words, delimiting of sentences, and detecting of text composition elements: section headings, paragraphs, items, rubrics, and numeration.
2. Morphologic analysis of words and identification of all occurrences of common scientific words and word combinations.
3. Recognition of dictionary discourse markers in the given text through matching text fragments with those dictionary lexicosyntactic patterns that contain identified common scientific words.
4. Delimiting of discourse segments, based on recognized discourse markers and semantic information presented in the dictionary for functional groups and classes. In general case, the result of the segmentation is ambiguous: several plausible discourse trees fit the sequence of recognized markers.
5. Estimation of plausible discourse-compositional trees resulted at the previous step and selection of the most plausible one. A number of heuristic rules are used for this purpose (in particular, an exemplifying segment is rather embedded into another segment than includes it).

To implement steps 3 and 4, surface syntactical analysis of sentences is required, which takes into account: *i)* agreement and coordination of words; and *ii)* overall grammatical structure of sentences.

It should be noted that reliability of recognition of discourse-compositional structure depends on various factors, among them are the number and types of discourse markers encountered in the text. In order to increase the reliability, the other linguistic devices, such as anaphoric links and repetitions of lexical units in adjacent sentences are to be considered.

Conclusion

In the paper we have overviewed the features of scientific discourse and the wide spectrum of common scientific words and expressions, with their role in scientific discourse. We concerned the methodological issues of development of the computer dictionary comprising common scientific lexicon and providing semantic and syntactic information valuable for automatic discourse analysis of scientific and technical texts. We have also outlined heuristic multi-step procedure intended to recognize discourse-compositional structure of a given scientific text, with the aid of the dictionary and surface syntactical analysis of sentences in the text.

Potential applications for the proposed discourse-analyzing procedure include:

- Eliciting of knowledge represented in scientific and technical texts, in particular, extraction of new terms and their definitions introduced into texts by authors [13];
- Text abstracting, which may be based on processing of detected markers, e.g. the expression *we illustrate our approach with N* transforms into *the approach is illustrated with N* [6];
- Document browsing and intra-document information retrieval, which are especially topical for large-size technical documents;
- Computer-aided writing and editing of scientific and technical texts [2].

Some applications will supposedly be investigated after implementation, testing, and refinement of the dictionary of common scientific lexicon and the recognition procedure.

Bibliography

1. Bolshakova, E.I. Phraseological Database Extended by Educational Material for Learning Scientific Style. In: ACH/ALLC 2001: The 2001 Joint Int. Conference. Conf. Abstracts, Posters and Demonstrations, New York, 2001, p. 147-149.
2. Bolshakova, E.I. Towards Computer-Aided Editing of Scientific and Technical Texts // International Journal on Information Theories & Applications. V. 10, N 2, 2003, p. 204-210.
3. Bolshakova, E.I., Vasilieva N.E., Morozov S.S. Lexicosyntactic Patterns for Automatic Processing of Scientific and Technical Texts. In: Proc. of 10th National Conference on Artificial Intelligence with International Participation 2006. Moscow, Fizmatlit, Vol 2, 2006, p. 506-514 (in Russian).
4. Dictionary of Word Combinations Frequently Used in English Scientific Literature. Nauka Publ., Moscow, 1968.
5. Dictionary of Verb-Noun Combinations of the Common Scientific Speech. Nauka Publ., Moscow, 1973 (in Russian).
6. Ellouze, M., Hamadou A.B. Relevant Information Extraction Driven with Rhetorical Schemas to Summarize Scientific Papers. In: Advances in Natural Language Processing. Third International Conference, PorTAL 2002. E. Ranchhod and N.J. Mamede (Eds.). Lecture Notes in Computer Science, N 2389, Springer-Verlag, 2002, p. 111-114.
7. Kurohashi, S., Nagao M. Automatic Detection of Discourse Structure by Checking Surface Information in Sentences. In: COLING 94 Proceedings of the 15th Int. Conf. On Computational Linguistics. Vol. II, Kyoto, Japan, 1994, p. 1123-1127.
8. Mann, W.C., Thompson S.A. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. Text, 8 (3), 1988, p. 243-281.
9. Mel'čuk, I. Dependency Syntax: Theory and Practice. SONY Press, NY, 1988.
10. Mitrofanova, O. Language of Scientific and Technical Literature. Moscow University Press, 1973 (in Russian)
11. Ono, K., Sumita K, Miike S. Abstract Generation Based on Rhetorical Structure Extraction. In: COLING 94 Proceedings of the 15th Int. Conf. On Computational Linguistics. Vol. II, Kyoto, Japan, 1994, p. 344-348.
12. Ryabtseva, N.K. Mental Performatives in Scientific Discourse. Voprosy yazykoznaniya, V 4, 1992, p. 12-28 (in Russian).
13. Paice, C., Jones P. The Identification of Important Concepts in Highly Structured Technical Papers. Proc. of 16th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburg, 1993, p.69-78.
14. Stede, M., Umbach C. DiMLex: A Lexicon of Discourse Markers for Text Generation and Understanding. Proceedings of Int. Conf. On Computational Linguistics COLING-ACM'98, Vol. 2, 1998, p. 1238-1242.
15. Wierzbicka A. Metatext w tekscie. In: O spojnosci tekstu. Wroclaw-Warszawa-Krakow-Gdansk, 1971, p.105-121.

Authors' Information

Elena I. Bolshakova – Moscow State Lomonossov University, Faculty of Computational Mathematics and Cybernetic, Algorithmic Language Department; Leninskie Gory, Moscow State University, VMK, Moscow 119899, Russia; e-mail: bolsh@cs.msu.su

CONDITIONS OF EFFECTIVENESS OF PATTERN RECOGNITION PROBLEM SOLUTION USING LOGICAL LEVEL DESCRIPTIONS OF CLASSES

Adil Timofeev, Tatiana Kosovskaya

Abstract: Earlier the authors have suggested a logical level description of classes which allows to reduce a solution of various pattern recognition problems to solution of a sequence of one-type problems with the less dimension. Here conditions of the effectiveness of the use of such a level descriptions are proposed.

Keywords: compound images, logical description of classes, effectiveness.

Introduction

Various pattern recognition problems which may be described in the terms of predicates (which characterize the whole object or its parts) were reduced in [1] to the proof of deducibility of propositional and predicate calculus formulas from a set of atomic formulas.

Upper bounds of the number of steps of an algorithm solving pattern recognition problems with logical description were proved in [3]. For example, such an upper bound for an algorithm solving the problem of analysis of a compound object is polynomial but the degree of such a polynomial depends of the number of objective variables included to the class description. As a rule such a number is rather large.

A level description of classes offered in [2] allows to reduce the solution of various pattern recognition problems to the solution of a sequence of one-type problems with the less dimension. Here conditions of decreasing of the number of steps of algorithm solving the described pattern recognition problems with the use of many-level description are proposed.

Setting of a problem of compound objects logical recognition

Let Ω be a set of finite sets $\omega = \{\omega_1, ..., \omega_t\}$. The set ω will be called a recognizable object. Let $p_1, ..., p_n$ be a collection of predicates which characterize an object (global indication) or describe properties or relations between elements of ω (local indication). The set Ω is a union of K (may be intersected) classes Ω_k.

Logical description $S(\omega)$ of an object ω is a set of all true formulas in the form $p_i(x)$ or its negation written for all parts x of the object ω.

Logical description of a class (DC) Ω_k is such a formula $A_k(x)$ that $A_k(x)$ contains as an atomic only formulas of the form $p_i(y)$ where y is a subset of x; $A_k(x)$ has no quantifiers; if for some ordering ω' of the object ω the formula $A_k(\omega)$ is true then $\omega \in \Omega_k$

These descriptions may be used for solving the following problems.

Identification problem. To check weather object ω or its part belongs to the class Ω_k.

This problem was reduced in[1] to the proof of deducibility of the formula $\exists y\ (y \subset \omega\ \&\ A_k(y))$ from the description $S(\omega)$.

Classification problem. To find all such numbers k that $\omega \in \Omega_k$..

This problem was reduced in[1] to the proof of deducibility of disjunction of formulas $A_k(\omega')$ (for some ordering ω' of the object ω) from the description $S(\omega)$ and pointing out all such numbers k for which the corresponding disjunct is true for ω.

Problem of analysis of a compound object. To find and classify all parts x of the object ω.

This problem was reduced in[1] to the proof of deducibility of disjunction of formulas $\exists y\ (y \subset \omega\ \&\ A_k(y))$ from the description $S(\omega)$ and pointing out all parts of ω which may be classified.

2. Level logical description of classes

Objects the structure of which allows to extract more simple fragments and to describe these objects in the terms of properties and relations between such fragments are regarded. In particular it may be done be means of selecting «frequently» appeared subformulas of formulas $A_k(x)$ with «small complexity». A system of equivalences in the form $p_j^1(x_j^1) \Leftrightarrow P_j^1(y_j^1)$ (where x_j^1 – new first-level variables, p_j^1 – new first-level predicates, $P_j^1(y_j^1)$ – subformulas of formulas $A_k(x)$) is written. The result of substitution of $p_j^1(x_j^1)$ instead of $P_j^1(y_j^1)$ into $A_k(x)$ is denoted by $A_k^1(x^1)$.

Such a procedure may be repeated with $A_k^1(x^1)$ but not later than $A_k^l(x^1)$ contains at least two subformulas in the same form.

3. Conditions of effectiveness of level description with the use of global indications

Let $p_1, ..., p_n$ be global indications (i.e. they are boolean variables). Then class descriptions are disjunctive normal forms (DNF) and any subformula of formulas $A, ..., A_K$ which appears at least two times is a simple conjunction.

Definition. Atom is variable or its negation.

Definition. Simple conjunctions $B_1, ..., B_m$ are called disjoint if there not exists such an atom that is included simultaneously into two different conjunctions.

Notifications.

a – a number of occurrences of boolean variables in formulas in DNF $A_1, ..., A_K$,

$P_1^1, ..., P_{n1}^1$ – subformulas of $A_1, ..., A_K$,

N_j^1 – a number of occurrences of subformula P_j^1,in $A_1, ..., A_K$,

v_j^1 – a number of occurrences of boolean variables in P_j^1,

$A_1^1, ..., A_K^1$ – the result of substitutions of atomic formulas p_j^1 instead of P_j^1 into $A_1, ..., A_K$,

Theorem 1. *If formulas $P_1^1, ..., P_{n1}^1$ are disjoint then for the equality $a^1= d\, a$ (for some $0< d < 1$) it is necessary and sufficient*

$$\Sigma_{j=1}^{n1} (v_j^1 -1) N_j^1 = (1-d)\ a . \quad (1)$$

Corollary 1.1. *If formulas $P_1^1, ..., P_{n1}^1$ are disjoint then for decreasing the number of occurrences of boolean variables in formulas $A_1^1, ..., A_K^1$ in comparison with the number of occurrences of boolean variables in formulas $A_1, ..., A_K$ it is necessary and sufficient*

$$\Sigma_{j=1}^{n1} (v_j^1 -1) N_j^1 > a . \quad (2)$$

Corollary 1.2. *If formulas $P_1^1, ..., P_{n1}^1$ are disjoint and $N_j > N$ for some N then for decreasing the number of occurrences of boolean variables in formulas $A_1^1, ..., A_K^1$ in comparison with the number of occurrences of boolean variables in formulas $A_1, ..., A_K$ it is necessary and sufficient*

$$\Sigma_{j=1}^{n1} (v_j^1 -1) \geq a / N . \quad (3)$$

The next theorem gives a necessary condition for not disjoint formulas $P_1^1, ..., P_{n1}^1$.

Theorem 2. *For the equality $a^1= d\, a$ (for some $0< d < 1$) it is necessary and sufficient*

$$\Sigma_{j=1}^{n1} (v_j^1 -1) N_j^1 \geq (1-d)\ a . \quad (4)$$

If $p_1, ..., p_n$ are boolean variables then both the identification problem and the classification problem may be solved with the use of resolution method or sequent propositional calculus the number of rule applications of which is not more then the number of occurrences of boolean variables in formula A_k (in formulas $A_1, ..., A_K$ for classification problem) [3]. Note that the upper bound of number of steps needed for calculation of p_j^1 equals to such a bound for classification problem if instead of A_k we take P_j^1.

Theorem 3. *If $a^1 = d\, a$ then for decreasing of number of rule application steps while using the 2-level description it is sufficient*

$$\Sigma_{j=1}^{n1} v_j^1 \leq (1-d)\ a. \quad (5)$$

4. Examples of two-level descriptions

Illustrate an application of the received conditions with a model example of 2-level description.

Let the set of recognizable objects is divided into 3 classes and objects may be described by means of 5 boolean variables *x, y, z, u, v*. Classes descriptions which allow to identify and classify an object have the form

A_1 = ~x & ~y V x & ~y & z V x & y & z & ~v
A_2 = ~x & y & ~z & ~u V ~x & y & u & ~v V x & ~z & ~ u V x & z & u & ~v
A_3 = ~x & y & z & ~u V ~x & y & u & v V x & ~z & u & v V x & y & z & v

The number of occurrences of boolean variables in formulas A_1, A_2, A_3 is a=40.

Example 1. Let the following subformulas are extracted.

P_1^1 = x & y & z
P_2^1 = ~x & y & u
P_3^1 = ~x & y & ~u
P_4^1 = x & z
P_5^1 = y & z

The number of occurrences of boolean variables in these subformulas is ν_1^1=3, ν_2^1=3, ν_3^1=3, ν_4^1=2, ν_5^1=2. The number of occurrences of each of these subformulas is N_j^1=2.

The formulas are not disjoint and we can use the condition (4) $\Sigma_{j=1}^{n1}(v_j^1 -1)N_j^1 \geq (1-d)\ a.$ In this example $\Sigma_{j=1}^{n1}(v_j^1 -1)\ N_j^1 = 16.$ Hence for decreasing the length of description it is necessary $16 \leq 40$ (1-d), i.e. $d \geq 0.4$.

In fact

A_1^1 = ~x & ~ y V ~y & p_4^1 V ~v & p_1^1 & p_4^1 & p_5^1
A_2^1 = ~z & p_3^1 V ~v & p_2^1 V x & ~z & ~u V u & ~v & p_4^1
A_3^1 = p_3^1 & p_5^1 V v & p_2^1 V x & ~z & u & v V v & p_1^1 & p_4^1

The number of occurrences of boolean variables in A_1^1, A_2^1, A_3^1 is a^1=29. As $a^1=d\ a$ we have d= 29/40 = 0.725.

Verify, weather we may guarantee that such a 2-level description provides a decreasing of an upper bound of number of steps of a solution of classification problem, i.e. weather the condition (5) $\Sigma_{j=1}^{n1} v_j^1 \leq (1-d)\ a$ is fulfilled. In this example $\Sigma_{j=1}^{5} v_j^1$ =13, $(1-d)\ a$ = 11. Hence the condition (5) is not fulfilled and we cannot guarantee a decreasing of an upper bound of number of steps of a solution of classification problem.

Example 1. Let the following subformulas are extracted.

P_1^1 = ~x & y
P_2^1 = x & z
P_3^1 = y & z

The number of occurrences of boolean variables in these subformulas is ν_1^1=2, ν_2^1=2, ν_3^1=2. The number of occurrences of each of these subformulas is N_1^1=4, N_2^1=4, N_3^1=3.

The formulas are not disjoint and we can use the condition (4) $\Sigma_{j=1}^{n1}(v_j^1 -1)N_j^1 \geq (1-d)\ a.$

In this example $\Sigma_{j=1}^{n1}(v_j^1 -1)\ N_j^1 = 11.$ Hence for decreasing the length of description it is necessary $11 \leq 40$ (1-d), i.e. $d \geq 0.725$.

In fact

A_1^1 = ~x & ~y V ~y & p_2 V ~v & p_2^1 & p_3^1
A_2^1 = ~z & ~u & p_1^1 V u & ~v & p_1^1 V x & ~z & ~u V u & ~v & p_2^1
A_3^1 = ~u & p_1^1 & p_3^1 V u & v & p_1^1 V x & ~z & u & v V v & p_2^1 & p_3^1

The number of occurrences of boolean variables in A_1^1, A_2^1, A_3^1 is a^1=32. As $a^1=d\ a$ we have d= 32/40 = 0.8.

Verify, weather we may guarantee that such a 2-level description provides a decreasing of an upper bound of number of steps of a solution of classification problem, i.e. weather the condition (5) $\Sigma_{j=1}^{n1} v_j^1 \leq (1-d)\ a$ is fulfilled. In this example $\Sigma_{j=1}^{3} v_j^1$ = 6, $(1-d)a$= 8. Hence the condition (5) is fulfilled and we can guarantee a decreasing of an upper bound of number of steps of a solution of classification problem.

Conditions of effectiveness of level description with the use of local indications

Let $p_1, ..., p_n$ characterize properties and relations of a recognizable object elements. In such a case it was proved in [3] that the number of steps of an algorithm solving identification problem is bounded by the number of arrangement of m_k from t : A_t^{mk}. For classification problem and problem of analysis of a compound object such a bound is $\Sigma_{k=1}^K A_t^{mk}$. (Here m_k – the number of objective variables in the description of the k-th class.

Consequently the number of steps of an algorithm solving these problems is an exponent of the number of objective variables in the description of classes (and a polynomial of a high degree for any particular description). Moreover, if it is possible to construct an algorithm which in a polynomial (over the length of classes descriptions) number of steps solves such problems then one of the most difficult problems of XXI century **P=NP** will be solved. However with the use of level logical descriptions it is possible to decrease an exponent in the upper bound of the number of steps of such an algorithm.

Notifications.

$m_1, ..., m_K$ – number of objective variables in formulas $A_1, ..., A_K$,

r – a number which is greater than number of objective variables in every formula $P_1^1, ..., P_{n1}^1$,

x_k – the string of variables of the formula A_k,

x_j^1 – new variables of the 1st level defined be equivalences $p_j^1(x_j^1) \Leftrightarrow P_j^1(y_j^1)$,

$s1$ – the number of variables occurred in $A_1, ..., A_K$ but not occurred in $P_1^1, ..., P_{n1}^1$.

Theorem 4. *Checking weather formulas $A_1, ..., A_K$ are true on the set $\omega = \{\omega_1, ..., \omega_t\}$ is equivalent to checking equivalences $p_j^1(x_j^1) \Leftrightarrow P_j^1(y_j^1)$ and weather formulas $A_1^1, ..., A_K^1$ are true on the same set.*

For decreasing the number of steps of an algorithm solving the problem of analysis of compound object it is sufficient

$$n_1 t^r + t^{s1+n1} < t^m. \quad (6)$$

Conclusion

Hence level logical description of classes of objects is described. In the frameworks of such an approach the conditions of decreasing of upper bounds of number of steps of an algorithm solving various pattern recognition problems including recognition of compound objects (compound images and scenes, complex signals and so on) are done.

The work has been done at partial support of the project 1.6 of the Program № 15 (“GRID”) of Presidium of RAS and RFBR grant № 06-08-01612.

Bibliography

[1]. Timofeev A.V., Kosovskaya T.M. Logic-Axiomatical Method of Knowledge Representation and Recognition for Robots. In: Computers in Industry 15 (1990). Proc. Intelligent Manufactoring Systems – IMS'89. Elsevier Science Publishers B.V. 1990.

[2]. Kosovskaya T.M. Level descriptions of classes for providing a solution in pattern recognition problems. In: Proc. III International Conf. “Discrete models in the theory of control systems”. Moscow, Dialog-MSU, 1998. (in Russian)

[3]. Kosovskaya T.M. Proofs of the number of step bounds for solving of some pattern recognition problems with logical description. In: Herald of St.Petersburg University, Series 1 (Mathematics), 2007 (?). (be published, in Russian)

Authors' Information

Timofeev Adil V. – Dr.Sc., Professor, Head of the Laboratory of Information Technologies in Control and Robotics of Saint-Petersburg Institute for Informatics and Automation of Russian Academy of Sciences, tav@iias.spb.su

Kosovskaya Tatiana M. – Doctor's Degree Student, Laboratory of Information Technologies in Control and Robotics of Saint-Petersburg Institute for Informatics and Automation of Russian Academy of Sciences, Russia;

Ph.D., Assistant Professor, Department of Mathematics and Mechanics, Saint-Petersburg State University, Russia, e-mail: kosov@nk1022.spb.edu

TABLE OF CONTENTS OF VOLUME 15, NUMBER 2