Krassimir Markov, Vladimir Ryazanov,
Vitalii Velychko, Levon Aslanyan
(editors)

New Trends
in
Classification and Data Mining

I T H E A
SOFIA
2010

Krassimir Markov, Vladimir Ryazanov, Vitalii Velychko, Levon Aslanyan (ed.)

New Trends in Classification and Data Mining

ITHEA®

Sofia, Bulgaria, 2010

First edition

Recommended for publication by The Scientific Concil of the Institute of Information Theories and Applications FOI ITHEA

This book maintains articles on actual problems of classification, data mining and forecasting as well as natural language processing:

- new approaches, models, algorithms and methods for classification, forecasting and clusterisation. Classification of non complete and noise data;
- discrete optimization in logic recognition algorithms construction, complexity, asymptotically optimal algorithms, mixed-integer problem of minimization of empirical risk, multi-objective linear integer programming problems;
- questions of complexity of some discrete optimization tasks and corresponding tasks of data analysis and pattern recognition;
- the algebraic approach for pattern recognition - problems of correct classification algorithms construction, logical correctors and resolvability of challenges of classification, construction of optimum algebraic correctors over sets of algorithms of computation of estimations, conditions of correct algorithms existence;
- regressions, restoring of dependences according to training sampling, parametrical approach for piecewise linear dependences restoration, and nonparametric regressions based on collective solution on set of tasks of recognition;
- multi-agent systems in knowledge discovery, collective evolutionary systems, advantages and disadvantages of synthetic data mining methods, intelligent search agent model realizing information extraction on ontological model of data mining methods;
- methods of search of logic regularities sets of classes and extraction of optimal subsets, construction of convex combination of associated predictors that minimizes mean error;
- algorithmic constructions in a model of recognizing the nearest neighbors in binary data sets, discrete isoperimetry problem solutions, logic-combinatorial scheme in high-throughput gene expression data;
- researches in area of neural network classifiers, and applications in finance field;
- text mining, automatic classification of scientific papers, information extraction from natural language texts, semantic text analysis, natural language processing.

It is represented that book articles will be interesting as experts in the field of classifying, data mining and forecasting, and to practical users from medicine, sociology, economy, chemistry, biology, and other areas.

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

Printed in Bulgaria

ISBN 978-954-16-0042-9

C\o Jusautor, Sofia, 2010

SYNTHESIS OF CORRECTOR FAMILY WITH HIGH RECOGNITION ABILITY*

Elena Djukova, Yurii Zhuravlev, Roman Sotnezov

***Abstract**: The model of recognizing procedures based on construction of family of logic correctors is proposed. For these purposes the genetic approach is used. This method allows, firstly, to reduce calculation costs and, secondly, to construct correctors with high recognition ability. The proposed model is tested on real problems.*

***Key words**: logical recognition procedures, covering of the Boolean matrix, logical corrector, algebra-logic data analysis, genetic algorithm.*

***ACM Classification Keywords**: I.5.1 Computing Methodologies - Pattern Recognition - Models*

Introduction

In the paper questions of logic and algebra-logic data analysis are considered. The most important problem in this direction is concerned with generation of informative fragments from feature description of objects. These fragments play the role of elementary classifiers and allow to differ objects from different classes. As a rule, the correctness of recognition algorithm (the ability to classify the objects from a training sample correctly) is provided by the correctness of each used elementary classifier [Djukova, Zhuravlev, 1997], [Djukova, Zhuravlev, 2000].

For generalization of this approach the correct recognition procedures can be constructed from the arbitrary sets of valid feature values. As a corrective function a monotone Boolean function can be used. In this case the constructing a corrector of minimal complexity can be reduced to the search of minimal cover of a Boolean matrix, which can be constructed from the training sample and has large size even in the simplest case. The idea of constructing the logic corrector was proposed in the work [Djukova, Zhuravlev, Rudakov, 1996]. Unfortunately, the problem has not been given sufficient attention, in particular, because of the large computational complexity.

In the paper a new model of recognition procedures based on constructing the logic corrector family is proposed. For these purposes the genetic approach is used. This method allows, firstly, to reduce calculation costs and, secondly, to construct correctors with high recognition ability. Proposed model is tested on real problems.

Description of voting model on logic correctors

The problem of recognition by precedents is considered in standard formulation. [Zhuravlev, 1978]. The set of objects M that can be represented as a union of disjoint subsets (classes) $K_1,\dots,K_l$ is investigated. Objects of set M are described by the set of integer attributes $x_1,\dots,x_n$. Each attribute has a finite number of valid values. As the initial information the set of object descriptions T from M is given (training sample). For these objects it is known which classes they belong. It is required on training sample and the description in system of attributes $x_1,\dots,x_n$ of some object S to determine to what class it belongs.

Let $H=\{x_{j_1},\dots,x_{j_r}\}$ be a set of r attributes and $\sigma=(\sigma_1,\dots,\sigma_r)$, where σ_i - one of valid values of feature x_{j_i} under $i=1,2,\dots,r$. The pair (H,σ) is called an elementary classifier (e.c.). The e.c. (H,σ) generates the predicate $P_{(H,\sigma)}(S)$ which is defined on objects $S\in M$, $S=(a_1,\dots,a_n)$, and such that

$$P_{(H,\sigma)}(S)=\begin{cases}1, & if\ \ a_{j_i}=\sigma_i,\\ 0, & otherwise.\end{cases}$$

The set of e.c. $U=\{(H_1,\sigma_1),\dots,(H_q,\sigma_q)\}$ *is called a (monotone) correct set for the class* K*,* $K\in\{K_1,\ldots,K_l\}$*, if there exists the (monotone) function of the logic algebra* F_K*, dependent on* q *variables, such that*

$$F_K\left(P_{(H_1,\sigma_1)}(S),P_{(H_2,\sigma_2)}(S),\dots,P_{(H_q,\sigma_q)}(S)\right)=\begin{cases}1, & if\ \ S\in K\cap T,\\ 0, & if\ \ S\in \overline{K}\cap T.\end{cases}$$

(here and in the following text $\overline{K}=\{K_1,\ldots,K_l\}\setminus\{K\}$).

The function F_K is called a (monotone) corrector for the class K, denote as $\omega_U(S)$ the binary set $\left(P_1(S),P_2(S),\dots,P_q(S)\right)$.

Let $a^{'}=\left(a_1^{'},\ldots,a_n^{'}\right)$, $a^{''}=\left(a_1^{''},\ldots,a_n^{''}\right)$ be binary sets. The note $a^{'}\succ a^{''}$ means that $a_i^{'}\geq a_i^{''}$ for all $i=1,2,\ldots,n$.

Let S', $S''\in M$ and U be the correct set of e.c. Let's put

$$\delta(S,S',U)=\begin{cases}1, & if\ \ \omega_U(S)\succ\omega_U(S'),\\ 0, & otherwise\end{cases}$$

in the case when the U is the monotone correct and

$$\delta(S,S',U)=\begin{cases}1, & if\ \ \omega_U(S)=\omega_U(S'),\\ 0, & otherwise\end{cases}$$

in the case when the U is the correct set that is not monotone.

Let $W_K=\{U_1,U_2,\dots,U_t\}$ be the set of (monotone) correct sets of e.c. for the class K, then the score for the class K for recognizing object S has a form

$$\Gamma(S,K)=\frac{1}{|T\cap K|}\sum_{U\in W_k}\sum_{S'\in T\cap K}\delta(S,S',U)$$

We obviously have the next

Statement 1. The set of e.c. $U=\{P_{(H_1,\sigma_1)}(S),P_{(H_{21},\sigma_2)}(S),\dots,P_{(H_q,\sigma_q)}(S)\}$ is monotone correct set for the class K if and only if for any two objects S' and S'' from training sample such that $S'\in K$, $S''\notin K$ there exists $i\in\{1,\ldots,q\}$ that

$$P_{(H_i,\sigma_i)}(S')=1 \text{ and } P_{(H_i,\sigma_i)}(S'')=0 \qquad (1)$$

The condition of monotony in the last statement can be removed if replace (1) on

$$P_{(H_i,\sigma_i)}(S')\neq P_{(H_i,\sigma_i)}(S'') \qquad (2)$$

Let U be the correct set of e.c. for the class K. The set U is called irreducible if from the condition $U'\subset U$ it follows that the set of e.c. U' is not correct set for the class K. The set U is called minimal if there is not exist smaller capacity correct set of e. c. for the class K.

Let L be an arbitrary Boolean matrix. The set of columns H of matrix L is called a covering if each row of the matrix L in crossing even with one of the columns in H gives 1. The covering is called irreducible if any its own subset is not covering. Let $c = (c_1,...,c_n)$ be the vector of weights of columns of the matrix L. The sum of columns weights of the covering is called a weight of covering. The covering, minimal on weight, is called a minimal covering. Note that in the case of unit column weights vector the minimal covering is the covering with minimum columns number. In the further it is considered, that the Boolean matrix has a unit vector of weights if it is not told opposite.

The training object $S^{'} = (a_1^{'},...,a_n^{'})$ generates the e.c. (H,σ), $H = \{x_{j_1},...,x_{j_r}\}$, $\sigma = (\sigma_1,...,\sigma_r)$, if $\sigma_i = a_{j_i}$ under $i = 1, 2, ..., r$. The set $U_K = \{(H_1,\sigma_1),\ldots,(H_{N_K},\sigma_{N_K})\}$ of all e.c. of the class K is considered. The difference between the number of training objects from the class K that generate the e.c. and the number of training objects from the $\overline{K}$ that also generate the same e.c. is called a weight of this e.c..

For pair of objects S' and S'' we shall construct the binary vector $B(S',S'') = (b_1,\ldots,b_{N_k})$ where

$$b_j = \begin{cases} 1, & if\ P_{(H_j,\sigma_j)}(S') = 1\ and\ P_{(H_j,\sigma_j)}(S'') = 0 \\ 0, & otherwise, \end{cases}$$

$j = 1, 2, ..., N_K$ .For the class K we shall construct the Boolean matrix L_K from all rows $B(S',S'')$ such that $S' \in K$ and $S'' \notin K$.

By constructing each column in matrix L_K corresponds to some elementary classifier from the set U_K . Let R be the of e.c. that corresponds to the column set H of matrix L_K . Following two statements are true.

Statement 2. The set of e.c. R is the monotone correct set for the class K if and only if H is the covering of the matrix L_K .

Statement 3. The set of e.c. R is the monotone irreducible (minimal) correct set for the class K if and only if H is the irreducible (minimal) covering of the matrix L_K .

In case of correct sets that is not monotone the set of all e.c. $U_K^{'} = \{(H_1^{'},\sigma_1^{'}),\ldots,(H_N^{'},\sigma_N^{'})\}$ is formed from parts of object descriptions from all classes. In this case the Boolean matrix $L_K^{'}$ is constructed from rows $D(S',S'') = (d_1,\ldots,d_N)$ where

$$d_j = \begin{cases} 1, & if\ P_{(H_j^{'},\sigma_j^{'})}(S') \neq P_{(H_j^{'},\sigma_j^{'})}(S'') \\ 0, & otherwise, \end{cases}$$

$j = 1, 2, ..., N$ for all pairs of objects $S' \in K$ and $S'' \notin K$.

Each column of the matrixes L_K and $L_K^{'}$ is associated with a weight of the according elementary classifier.

In the present work two models of recognizing procedures are constructed and investigated. The first model is founded on constructing the one correct set of e.c. which is close to minimal on complexity. The second model is founded on constructing the family of the most informative corrects sets of e.c.

As a rule, even for problems of small dimension the number of elementary classifiers is great and procedure of construction of the minimal correct set with use matrixes L_K and $L_K^{'}$ demands significant computing resources, therefore a question on development of effective methods of the decision of problems of algebra-logic correction of elementary classifiers. Because of NP-completeness of the set-covering problem the exact algorithms for the search of solution are practically inapplicable. For problems with big dimensions the approximate algorithms are used. The algorithms using the genetic approach concern to such algorithms.

In the present it is shown that the voting procedure on the one minimal correct set of e.c. cannot provide comprehensible quality of recognition. Therefore in the present work the model which builds family of logic proof-readers is offered. For constructing this model the genetic algorithm from [Sotnezov, 2008] is used. Thus in a population correct set of e.c. with good recognizing ability are selected.

The construction of family of logic correctors on the basis of the genetic approach

The training sample T is divided on two subsamples: base (T_0) and tuning (T_1) according to a technique described in [Djukova, Peskov, 2005]. The sample T_0 is used for construction matrixes L_K and L'_K , the sample T_1 is used for an estimation of quality of recognition of correctors found by genetic algorithm.

The quality of recognition $\tau(U)$ for the correct set of e.c. U of the class K is estimated under the formula

$$\tau(U) = \frac{1}{|T_1 \cap K|} \sum_{S \in T_0 \cap K} \sum_{S' \in T_1 \cap K} \delta(S, S', U) - \frac{1}{|T_1 \cap \overline{K}|} \sum_{S \in T_0 \cap K} \sum_{S' \in T_1 \cap \overline{K}} \delta(S, S', U)$$

Thus, quality of recognition of the corrector U is equal to difference between the number of objects from the class K which are correctly recognized by the corrector U , and numbers of objects from other classes which also are recognized by the corrector U to the class K .

Work of genetic algorithm reminds development of a biological population in which each object, is characterized by a set of genes. Updating of such population eventually happens according to the law "survives the most adapted". Thus there is an opportunity of reception of new objects by means of operators of crossing and a mutation who in special way combine genes of parents.

Let $L = (a_{ij})_{m \times n}$ be the Boolean matrix constructed on sample T_0 and $c = (c_1, ..., c_n)$ is a vector of its column weights. The covering of the matrix L we shall represent as the integer vector $Q = (q_1, ..., q_m)$ where q_i is the number of one of columns which cover the i -th row.

By the greedy heuristic from [Sotnezov, 2009] the initial family of decisions $P = (Q_1, ..., Q_N)$ that is called a population is formed. Elements of the set P are called individuals.

For the individual Q_j the function of the fitness $f(Q_j)$ describing quality of the found decision is defined. The individual Q_j describes the covering of the matrix which corresponds to some correct set of e.c. U_j . As a function of fitness $f(Q_j)$ we shall use value

$$\tau(U_j) - \min_{i \in \{1,2,...,N\}} \tau(U_i) + 1$$

For each individual Q_j from the population P the probability p_j is calculated by the formula

$$p_j = \frac{1/f_j}{\sum_{i=1}^{N} 1/f_i} \qquad (5)$$

where f_j - the fitness of the individual Q_j .

On the next step of genetic algorithm with the set of probabilities $\{p_i\}$ $i = 1, 2, ..., N$, two parental individuals $Q^{(1)} = (q^{(1)}_1, ..., q^{(1)}_m)$ и $Q^{(2)} = (q^{(2)}_1, ..., q^{(2)}_m)$ are selected. These individuals are used for generating child individual $Q = (q_1, ..., q_m)$ with the crossing operator by following rules.

Let f_1 and f_2 be fitnesses of individuals $Q^{(1)}$ and $Q^{(2)}$ accordingly, then the i -th component of the child Q is equal to

$$q_i = \begin{cases} q_i^{(1)}, \text{ with probabilty } \dfrac{c_{q_i^{(2)}} \cdot f_2}{c_{q_i^{(1)}} \cdot f_2 + c_{q_i^{(2)}} \cdot f_2} \\ q_i^{(2)}, \text{ with probability } \dfrac{c_{q_i^{(1)}} \cdot f_1}{c_{q_i^{(1)}} \cdot f_1 + c_{q_i^{(2)}} \cdot f_2}. \end{cases}, \ i = 1, 2, ..., m$$

In difference from most often used one-point and two-point crossovers proposed operator of crossing considers the structure of parental individuals and their relative fitness. The more relative fitness of the parental individual has the bigger probability, that its gene will be copied in the descendant.

The using only the operator of crossing for updating the population can lead to formation of individuals with approximately identical set of columns. It means, that the algorithm converges in some local minimum in which neighbourhood there will be all new descendants. For overcoming local minima the operator of a mutation is used. This operator in random way changes (mutates) the given number of genes in the description of the child. As especially strong influence of the operator of a mutation on the individual should occur at convergence of the search process of the optimum decision it is offered to increase the number of mutating genes $k(t)$ with growth of the number of algorithm steps according to the formula

$$k(t) = K\left(1 - \frac{1}{C \cdot t + 1}\right),$$

where t - the number of the algorithm step, K and C are variable parameters which characterize the number of mutating genes on the last step of genetic algorithm and the speed of change of the mutating genes number accordingly.

After using of the operator of crossing and a mutation we receive the integer vector Q that corresponds to some covering H of the matrix L. If H is not the irreducible covering the procedure of feasibility restoration of the decision is applied. The procedure of feasibility restoration of the decision can be described as follows. Let M_j be the set of rows of the matrix L, covered by the column j, then

1. for each $i \in \{1, 2, ..., m\}$ the value w_i as number of columns from H covering i -th row is defined;
2. for each $j \in H$ in decreasing order of weights the set w_i, $i \in M_j$ is analyzed. If all such w_i is more than one the column j leaves the set H' and for each $i \in M_j$ the value w_i is reduced by one;

As a result of the feasibility restoration procedure application the individual Q that corresponding to irreducible covering of the matrix L is received. The individual Q replaces one of individuals of the population P if 1) in a population there is not the individual identical to Q 2) the set $R_Q = \{Q' \in P \mid f(Q') \geq f(Q)\}$ is not empty.

For replacement in random way the individual from R_Q is chosen. The first condition is necessary for prevention of occurrence of identical individuals and, as consequence, degeneration of a population. The second condition means, that in a population the most adapted individuals (that is individuals with the least weights of coverings) is occurred.

The genetic algorithm stops the work if the population has been updated $N_{\max}$ time. That means it has been received $N_{\max}$ unique, more adapted individuals. The number $N_{\max}$

The population received as a result contains the description of family of correctors with high recognizing ability. Thus the individual with the greatest fitness corresponds to the logic corrector with the maximal recognizing ability. The changing fitness function of of the population individuals, it is possible to construct logic correctors of various type. For example, if as fitness function use weight of the corresponding covering of the expanded matrix of comparison, the genetic algorithm will give out the logic correctors close to minimal.

The model testing on real problems

The model was tested for two cases. In the first case one minimal corrector, and in the second family of correctors was used. Problems for testing have been taken from repository UCI [Asuncion, Newman, 2007]. Characteristics of problems are resulted in Table 1.

Problem	The number of attributes	The number of objects in the first class	The number of objects in the second class
A	24	51	237
B	19	51	218
C	35	38	107
D	9	626	332
E	16	168	265

Table 1. Characteristics of test problems

The algorithm of voting on the corrector close to minimal (algorithm A2) and algorithm of voting on family of correctors (algorithm A1) were compared to the algorithms realized in system Recognition [Zhuravlev, Rjazanov, Senko, 2006] on the general percent of recognition R_1 - the percent of number of correctly recognized control objects to the number of all control objects and weighed percent of recognition R_2 , which are defined as follows. Let r_i and n_i be accordingly the number of correctly recognized objects and the total number of objects in the class $K_i, i = 1,2,\ldots,l$, then

$$R_1 = \frac{\sum_{i=1}^{l} r_i}{\sum_{i=1}^{l} n_i},$$

$$R_2 = \sum_{i=1}^{l} \frac{r_i}{n_i}.$$

In Table 2 for each compared algorithm on each problem there are specified: the general percent of recognition R_1 (the first line of each cell), general percent of recognition for each of classes (the second line of each cell) and the weighed percentof recognition R_2 (the third construction of each cell).

Recognizing algorithm	Problem A	Problem B	Problem C	Problem D	Problem E
A1	**84.91%** (71.43, 88.23) **79.83%**	**85.56%** (83.33, 86.08) **84.71%**	85.45% (75.0, 91.43) **83.21%**	99.2% (100, 97.87) 98.94%	**96.04%** (94.6, 97.0) **95.84%**
A2	67.20% (16.0, 80.0) 48.0%	63.64% (30.0,82.86) 56.43%	54.64% (22.22, 62.03) 42.12	98.96% (100,97.16) 98.58	93.87% (92.7, 95.4) 94.05%
Algorithm of calculation of estimations	80.2% (0.0, 100.0) 50.0%	72.7% (50.0, 85.7) 67.85%	81.4% (0.0,100.0) 50.0%	72.7% (94.7, 34.8) 64.75	59.7% (0, 100) 50%
Binary decision tree	80.2% (0.0, 100) 50.0%	80.0% (45.0, 100.0) 72.5%	81.4% (0,100) 50.0	63.4% (100,0) 50.0	59.7% (0, 100) 50%
Logical patterns of classes	80.2% (42.9, 89.4) 62.65%	54.5% (20.0, 74.3) 47.15%	75.3% (50,81) 65.5%	**99.5%** (100, 98.6) **99.3%**	50.3% (36.7%,59.6%) 48.15%
SVM	81.1% (52.2,89.2) 70.7%	80.0% (50.0,97.1) 73.55%	**89.7%** (50.0, 98.7) 74.35%	63.4% (100, 0) 50%	59.7% (0 , 100) 50%
Voting on irreducible tests	80.2% (47.6, 88.2) 67.9%	76.4% (40, 97.1) 68.55%	87.6% (72.2, 91.1) 81.65%	73.5% (77.5, 66.7) 72.1%	56.4% (71.7 , 46.1) 58.9%

Table 2. Results of testing

Conclusion

The algorithm of voting on family of correctors A1 considerably overcomes other considered algorithms on problems with small number of values in object attributes (problems D and E) on parameters R_1 and R_2. On other problems the algorithm A1 due to consideration of each of classes separately also overcomes other algorithms on parameter R_2.

Bibliography

[Djukova, Zhuravlev, 1997] E.V. Djukova, Yu.I. Zhuravlev. Discrete Methods of Information Analysis in Recognition and Algorithm Synthesis // Pattern Recognition and Image Analysis. 1997. Vol.7. No.2. P. 192-207.

[Djukova, Zhuravlev, 2000] Djukova E.V., Zhuravlev Yu. I. Discrete Analysis of Feature Descriptions in Recognition Problems of High Dimensionality. // Zh. Vychisl. Mat. Mat. Fiz. 2000, 40(8), 1264-1278 (2000) [Comput. Math. Math. Phys. 40(8), 1214-1227 (2000)]

[Djukova, Zhuravlev, Rudakov, 1996] E.V. Djukova, Yu.I. Zhuravlev, K.V. Rudakov. On an algebra-logic synthesis of correct recognition procedures on base of elementary classifiers // Zh. Vychisl. Mat. Mat. Fiz. 1996 36(8), 215-223

[Zhuravlev, 1978] Zhuravlev Yu.I. On an Algebraic Approach to Recognition or Classification Problems. //

Problemy Kibernetiki, Moscow: Nauka, 1978, no.33, pp. 5-68 [in Russian]

[Sotnezov, 2008] Sotnezov R.M. Genetic Algorithms in problems of discrete optimization and recognition,

International Conference on "Pattern Recognition and Image Analisys: new Information Technologies", Nizhni Novgorod, Russian Federation, 2008. V.2, P 173-175

[Sotnezov, 2009] Sotnezov R.M. Genetic Algorithms for Problems of Logical Data Analysis in Discrete Optimization and Image Recognition // Pattern Recognition and Image Analysis, 2009, Vol. 19, No 3, pp. 469-477

[Djukova, Peskov, 2005] Djukova E.V., Peskov N.V. Constructing recognition procedures on base of elementary classifiers// Problemy Kibernetiki, Moscow: Fizmatlit, 2005. no 14 p. 57-92 [in Russian]

[Asuncion, Newman, 2007]Asuncion A., Newman D.J. UCI Machine Learning Repository, University of California, Irvine. - 2007. www.ics.uci.edu/~mlearn/MLRepository.html

[Zhuravlev, Rjazanov, Senko, 2006] Zhuravlev Yu.I., Rjazanov V.V., Senko O.V. «Recognition». Mathematic Methods. Programm System. Practical Applications // Moscow FAZIS, 2006. 176 p.

Author's Information

***Djukova Elena Vsevolodovna** – postdoctoral (Doktor Nauk) degree in the physical and mathematical sciences, Major scientist, Dorodnicyn Computing Centre of the Russian Academy of Sciences, Vavilova st., 40, Moscow, 119333, Russia; e-mail: edjukova@mail.ru*

Major Fields of Scientific Research: logical data analysis, pattern recognition, discrete mathmeatics, logical recognition procedures, computational complexity of discrete problems, synthesis of asymptotically optimal algorithms for solving discrete problems.

***Zhuravlev Yurii Ivanovich** – academician of the Russian Academy of Sciences, postdoctoral (Doktor Nauk) degree in the physical and mathematical sciences, vice-president of Dorodnicyn Computing Centre of the Russian Academy of Sciences in scientific research, Vavilova st., 40, Moscow, 119333, Russia, e-mail: zhur@ccas.ru*

Major Fields of Scientific Research: mathematical cybernetic and theoretical informatics; discrete analysis; theory of local algorithms of information processing; forecasting and recognition methods; development of mathematical methods of decision-making based on incomplete, contradictory, heterogeneous information

***Sotnezov Roman Mihailovich** – post graduate student in Moscow State University, faculty of Computational Mathematic and Cybernetic, Leninskie gory, 1, Moscow, 119234, Russia; e-mail: rom.sot@gmail.com*

Major Fields of Scientific Research: pattern recognition, discrete mathematics, optimization problems