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

ON THE COMPLEXITY OF SEARCH FOR CONJUNCTIVE RULES IN RECOGNITION PROBLEMS

Elena Djukova, Vladimir Nefedov

Abstract: *New results are obtained in logical data analysis and in the development of recognition procedures based on constructing conjunctive rules. Asymptotically optimal methods of constructing normal forms for characteristic functions of classes are developed. The technique of effective search is improved for the conjunctive rules in the case of incomplete data. Theoretical results are confirmed by experimental estimations.*

Keywords: *logical recognition procedures, enumeration problem complexity, dualization problem, algorithm with a polynomial delay, asymptotically optimal algorithm, normal form of a logical function, maximal conjunction, transformation of normal forms.*

ACM Classification Keywords: *F.2.1 Theory of Computation – Analysis of Algorithms and Problem Complexity – Numerical Algorithms and Problems – Computations on matrices, G.2.1. Mathematics of Computing – Discrete Mathematics – Combinatorics – Counting problems.*

Introduction

Logical analysis of Integer data in recognition is based on constructing normal forms of double-valued logical functions which are characteristic functions of classes [Zhuravlev, 1978], [Djukova, 1977], [Djukova, 1982], [Djukova, 1987], [Djukova, 1989], [Djukova, Zhuravlev, 1997], [Djukova, Zhuravlev, 2000], [Djukova, 2004]. The characteristic function of a class is completely defined by the training set and the concerned recognition algorithm model. If the descriptions of objects are complete (for each precedent all feature values are defined) then the problem appears concerned with the transformation of the perfect conjunctive normal form (CNF) of the function F_K into the disjunctive normal form (DNF), i.e. the problem of multiplication of logical terms with exactly n variables in each term, where n is the number of features. In the case of incomplete data some variables may be absent in some terms and thus the initial CNF would not be perfect. Constructing the required DNF brings us to constructing the set of the conjunctive rules of the class K .

The search for conjunctive rules is a difficult computational problem. Constructing the reduced DNF of the function F_K is of a particular complexity (its length as a rule grows exponentially with the growth of the initial DNF length). This problem is one of difficult generation problems. The effectiveness of algorithms for enumeration problems is usually estimated by the complexity of a step (the time delay of a step). As the major enumeration problem the dualization is regarded. The dualization is the problem of constructing the reduced DNF of a monotone Boolean function $f(x_1,\ldots,x_n)$ define by the CNF of the following type

$$D_1 \,\&\, \ldots \,\&\, D_u \,. \tag{1}$$

The search for effective algorithm for this problem is carried since the middle of the last century, starting from the classical works by Yablonskiy [Chegis, Yablonskii, 1958]. In [Gurvich, Khachiyan, 1999] an “incremental quasi-polynomial algorithm” is suggested. The number of elementary operations performed by that algorithm on each step is bounded by a quasi-polynomial over u , n and the number of the maximal conjunctions found on the previous steps. The existence of an algorithm working wth a (quasi)-polinomial delay depending only on u and n is not ascertained by the time.

In [Djukova, 1982], [Djukova, 1987] an approach is proposed, which allows to solve the considered problem of the transformation of the normal forms of f efficiently not always but almost always (for almost all CNFs of type (1) as $n \to \infty$). In this case an algorithm is allowed to perform “unnecessary” (empty) steps, but the number of such steps should have "almost always" lower growth degree than the number of all algorithm steps does. For this purpose the conception of an asymptotically optimal algorithm was introduced. This conception has been defined more exactly in a series of followed works [Djukova, 2003], [Djukova, 2004], [Djukova, 2007]. Let us adduce this definition.

Let $B(f)$ be a set of all maximal conjunctions of the function f, let $Q(f)$ be a finite sequence of elementary conjunctions which contains $B(f)$. It is supposed that some conjunctions may be found in $Q(f)$ more than once.

We will say that an algorithm A constructs $Q(f)$ with a *polynomial delay* if exactly one conjunction from $Q(f)$ is built on each step with no more than $d(u,n)$ elementary operations performed, $d(u,n)$ is bounded above by a polynomial over u and n. The elementary operation is an examination of one symbol of a variable in the CNF.

Let an algorithm A construct the sequence $Q(f)$ with a polynomial delay checking each built conjunction from $Q(f)$ to be in $B(f)$ (such check can be performed at a polynomial time according to statements 2 and 3 formulated below). The number of steps of the algorithm A equal to the length of the sequence $Q(f)$ is denoted by $N_A(f)$.

The algorithm A is called *asymptotically optimal*, if for almost all CNFs of type (1) as $n \to \infty$ it is true that

$$N_A(f) \approx |B(f)|$$

(the number of steps of the algorithm A is asymptotically equal to the power of $B(f)$).

Later the concept of an asymptotically optimal algorithm was spread for the case of a double-valued logical function F defined on k-ary n-dimensional arrays and determined by a CNF of u elementary disjunctions [Djukova, 1987], [Djukova, 1989], [Djukova, Zhuravlev, 1997] [Djukova, Zhuravlev, 2000], [Djukova, Inyakin, 2008].

By the moment the series of asymptotically optimal algorithms has been constructed for the condition $\log_d u \le (1-\varepsilon)\log_d n$, $\varepsilon > 0$, where $d = 2$ for the case of a monotone Boolean function, $d = k$ for the case when F is determined by a perfect CNF and $d = (k+1)/k$ for the case when F is determined by a CNF which is not perfect. In these algorithms there are unnecessary steps of two types. There are unnecessary steps of two types in these algorithms. Each step of the first type yields a conjunction which is not maximal whereas each step of the second type yields a maximal conjunction constructed before. Among the constructed algorithms there are both the algorithms that make only one type unnecessary steps and the algorithms containing the unnecessary steps of both types.

The experimental comparison of the constructed algorithm has been carried out. The best result has been shown by the algorithm OPT and by its modification, the algorithm OPT+, assigned for constructing the maximal conjunctions of a monotone Boolean function and for constructing the maximal conjunctions of a double-valued logical function defined by the perfect CNF, respectively.

In this paper an algorithm OPT++ is built which is a modification of the algorithm OPT for the case when a double-valued logical function F is defined by an arbitrary CNF. The algorithm OPT++ is asymptotically optimal as well, for the case $\log u \le (1-\varepsilon)\log n$, $\varepsilon > 0$, $d = (k+1)/k$ (this fact follows from works [Djukova, Zhuravlev, 1997], [Djukova, Zhuravlev, 2000]). In the following section it is shown that for the case $n \le u$ the logarithm of the number of the algorithm steps to the base d is equal to the logarithm of the number of all maximal conjunctions of the function F "almost always" as $n \to \infty$. The justification of the algorithm effectiveness is based upon the analysis of metric (quantitative) characteristics of the set of all maximal conjunction and of the so-called irreducible conjunctions of the function F. Показано, что эти асимптотики совпадают. Asymptotic estimates of typical values of the logarithm to the base d of the number of the maximal conjunctions of F and the logarithm to the base d of the number of the irreducible conjunctions of F are obtained. These asymptotic estimates are shown to be equal. When obtaining these estimates we used the technique from the papers [Dem'yanov, Djukova, 2007], [Djukova, 2007]. The effectiveness of the algorithm OPT++ is approved experimentally, particularly the experimental estimates of the number of "unnecessary" steps are obtained.

Major results

Let E_k^n be the set arrays $(\sigma_1,\dots,\sigma_n)$, where $\sigma_i \in \{0,1,\dots,k-1\}$, $i \in \{1,\dots,n\}$, and let a function $F(x_1,\dots,x_n)$ be determined on E_k^n and possesses values 1 and 0 on subsets N_F and $N_{\neg F}$, respectively.

Let

$$x^{\sigma} = \begin{cases} 1, & \text{if} \quad x = \sigma, \\ \text{otherwise}, & \end{cases}$$

$x, \sigma \in \{0,1,\dots,k-1\}$. The notions of an elementary conjunction and elementary disjunction are defined routinely. An elementary conjunction (EC) over variables $x_1,\dots,x_n$ is the formula of the type $x_{j_1}^{\sigma_1}\cdot\ldots\cdot x_{j_r}^{\sigma_r}$, where $x_{j_i} \in \{x_1,\dots,x_n\}$, when $i = 1,2,\dots,r$, and $x_{j_q} \neq x_{j_t}$ when $t,q \in \{1,2,\dots,r\}$, $t \neq q$. The value of the EC is equal to 1 if and only if all its factors are equal to 1. The number r is called the rank of the elementary conjunction.

An elementary disjunction (ED) over variables $x_1,\dots,x_n$ is the formula of the type $x_{j_1}^{\sigma_1} \vee \ldots \vee x_{j_r}^{\sigma_r}$, where $x_{j_i} \in \{x_1,\dots,x_n\}$, when $i = 1,2,\dots,r$, and $x_{j_q} \neq x_{j_r}$, when $t,q \in \{1,2,\dots,r\}$, $t \neq q$.

Let B be an elementary conjunction over variables $x_1,\dots,x_n$ and let $M(B,R)$ be the number of the disjunctions not containing variables from B in the CNF R. Let N_B be in interval of verity of the EC B. The EC B is called *admissible* for F if $N_B \cap N_{\neg F} = \varnothing$; i.e., when $M(B,R) = 0$. The EC B is called *irreducible* for F if there does not exist an admissible EC B' such that $N_{B'} \supset N_B$ and $M(B',R) = M(B,R)$. The EC B is called *maximal* for F if it is both admissible and irreducible. The EC B is called *irredundant* for F if it is irreducible for F and there does not exist an irreducible for F conjunction B' such that $N_B \supset N_{B'}$.

Statement 1. If an EC B is maximal for F then it is irredundant for F.

The converse statement if false(an irredundant conjunction may not be admissible).

Let the function F by determined by a CNF R of the type $D_1 \,\&\ldots\&\, D_u$, where D_i, $i = 1,2,\dots,u$, - ED over variables $x_1,\dots,x_n$. Denote by $D_C(F)$ the reduced DNF of the function F, i.e. the DNF consisting of all maximal conjunctions of the function F.

Let us consider the problem of the transformation of R into $D_C(F)$. The required statements 2 and 3 are presented below.

Statement 2 [Djukova, Nefedov, 2009]**.** An EC B is admissible for F if and only if each disjunction D_i, $i = \{1,2,\dots,u\}$, contains at least one factor of B.

Statement 3 [Djukova, Nefedov, 2009]**.** An EC $B = x_{j_1}^{\sigma_1} \cdot \ldots \cdot x_{j_r}^{\sigma_r}$ is irreducible for F if and only if there exist r disjunctions in the CNF K, each disjunction containing exactly one factor of B and, if $r > 1$, $p,q \in \{i_1,\ldots,i_r\}$, $p \neq q$, the disjunctions D_p and D_q contain different factors of B.

Statements 2 and 3 are respectively the criteria of admissibility and irreducibility of the EC of a function determined by a CNF.

Let us describe the algorithm ОПТ++ constructing DNF $D_C(F)$. The algorithm is based on constructing an irredundant for F conjunction B and checking the condition $N_B \cap N_{\neg F} = \varnothing$ on each step. If this condition is true then according to statement 1 the constructed on this step conjunction is maximal. Else the performed step is "empty" (unnecessary).

It is convenient to present the work of the algorithm OPT++ as the process of constructing a decision tree $Д_F$. Each inner vertex of this tree corresponds to the pair (B,R'), where B is the irreducible for F conjunction and R' is the CNF derived from R by deleting some disjunctions and variables. A dangling vertex corresponds to an irredundant conjunction. The set of the dangling vertices corresponds to a subset of all irredundant conjunctions. On each step the algorithm constructs one dangling vertex of the tree $Д_F$ at the time $O(kunq(u+q))$, where $q = \min(u,kn)$. The important characteristic of the algorithm is that different vertices of the dicision tree $Д_F$ correspond to different irreducible conjunctions. Thus the algorithm does not perform "repeated" steps.

Denote by $S_r(F)$ the set of all irreducible conjunctions of the rank r of the function F; denote by $B_r(F)$ the set of all maximal conjunctions of the rank r of the function F;

$$S(F) = \bigcup_{r=1}^{n} S_r(F),\ B(F) = \bigcup_{r=1}^{n} B_r(F).$$

Let $n \le u \le k^{n^\beta}$, $\beta < \frac{1}{2}$, $r_1 = [\log_d u - \log_d \ln \log_d u - 1]$, $d = (k+1)/k$. For the considered case the asymptotic form of the logarithm of the typical number of the conjunctions in $S(F)$ with ranks no lower than r_1 is obtained (when $n \to \infty$). It is shown that this asymptotic form is congruent with the asymptotic form of the logarithm of the typical number of the conjunctions in $B(F)$ and is also congruent with the asymptotic form of the typical number of the conjunctions in $B(F)$ with ranks no lower than r_1. The estimate of the typical rank value of a conjunction in $B(F)$ is obtained. The mention estimates are stated in Theorems 1-3 formulated below.

Theorem 1. If $n \le u \le k^{n^\beta}$, $\beta < 1/2$, then as $n \to \infty$ for almost all CNFs of type (1) the logarithm to the base d of the number of all conjunctions in $S(F)$ with ranks no lower than r_1 is asymptotically equal to the logarithm to the base d of the number of all conjunctions in $B(F)$ with ranks no lower than r_1 and is asymptotically equal to $\log_d C_n^{r_1} + r_1$.

Theorem 2. If $n \le u \le k^{n^\beta}$, $\beta < 1/2$, then as $n \to \infty$ for almost all CNFs of type (1) the logarithm to the base d of the number of all conjunctions in $B(F)$ with ranks no lower than r_1 is asymptotically equal to the logarithm to the base d of the number of all conjunctions in $B(F)$ and is asymptotically equal to $\log_d C_n^{r_1} + r_1$.

Theorem 3. If $n \le u \le k^{n^\beta}$, $\beta < 1/2$, then as $n \to \infty$ for almost all functions determined by a CNF of type (1) the ranks of almost all conjunctions from $B(F)$ are in $[r_1, \log_d un]$.

The proofs of Theorems 1-3 are based on Lemmas 1-6 stated below. In the proofs of Lemmas 1-6 the results from [Djukova, Peskov, 2002] and [Djukova, Inyakin, 2003] are used.

Let $N_{un}^k = \{F\}$ be the space of simple events in which each event F happens with a probability of $1/\left|N_{un}^k\right|$. The mathematical expectation of a random variable $X(F)$ defined on N_{un}^k is denoted by $\mathbf{M}X(F)$.

The following result is easy to prove

Lemma 1. Let $X(F) \geq 0$, $\theta > 0$, and $\nu_\theta(n)$ be the fraction of functions F from N_{un}^k for which $X(F) \geq \theta \mathbf{M}X(F)$. Then $\nu_\theta(n) \leq 1/\theta$.

In what follows, the notation $a_n \approx b_n$ as $n \to \infty$ and $a_n \leq_n b_n$ as $n \to \infty$ means $\lim \dfrac{a_n}{b_n} = 1$ as $n \to \infty$ and $\lim \dfrac{a_n}{b_n} \leq 1$ as $n \to \infty$, respectively.

Let $a_r = C_n^r C_u^r r!(k-1)^{r^2 - r} k^{r - r^2}$. Since $a_{r+1} = o(a_r)$ for $n \leq u$ and $r \geq r_1$, we have the following result.

Lemma 2. For $n \leq u$

$$\sum_{r \geq r_1} C_n^r C_u^r r! d^{r - r^2} \approx C_n^{r_1} C_u^{r_1} r_1! d^{r_1 - r_1^2}, \; n \to \infty .$$

Lemma 3. For $u \leq k^{n^\beta}$ and $\beta < 1/2$

$$\log_d \left(C_u^{r_1} r_1! d^{-r_1^2} \right) = \overline{o}\left(\log_d \left(C_n^{r_1} d^{r_1} \right) \right), \; n \to \infty .$$

Proof. The lemma can be proved by direct verification. Indeed, the obvious inequality $C_n^r \geq \left(\dfrac{n-r}{r} \right)^r$ implies that

$$\log_d C_n^{r_1} \geq_n (1-\beta) r_1 \log_d n, \; n \to \infty . \quad (2)$$

On the other hand,

$$b = C_u^{r_1} r_1! d^{-r_1^2} \leq \left(k^2 \ln \log_d u \right)^{r_1} .$$

Consequently, we have

$$\log_d b \leq_n r_1 \log_d \ln n, \; n \to \infty . \quad (3)$$

The assertion of the lemma follows from (2) and (3).

On $N_{un}^k = \{F\}$ consider the random variables $\eta_r(F) = \left|B_r(F)\right|$ and $\xi_r(F) = \left|S_r(F)\right|$. It is easy to calculate

$$\mathbf{M}\xi_r(F) \leq C_n^r C_u^r r! k^{r^2 - r} (k+1)^{r - r^2}$$

(see [Djukova, Peskov, 2002]).

Lemma 4. For $n \leq u \leq k^{n^\beta}$ and $\beta < 1/2$, for almost all F from N_{un}^k

$$\log_d \sum_{r \geq r_1} \eta_r(F) \leq \log_d \sum_{r \geq r_1} \xi_r(F) \leq_n \log_d C_n^{r_1} + r_1, \; n \to \infty .$$

Proof. Lemmas 2 and 3 straightforwardly imply that

$$\log_d \sum_{r \geq r_1} \mathbf{M}\eta_r(F) \leq \log_d \sum_{r \geq r_1} \mathbf{M}\xi_r(F) \approx \log_d C_n^{r_1} + r_1,\ n \to \infty.$$

Applying Lemma 1 with $\theta = \log_d \log_d n$, we derive the assertion of Lemma 4.

Lemma 5. For $u \leq k^{n^\beta}$ and $\beta < 1/2$, for almost all F from N_{un}^k

$$\log_d \sum_{r \geq r_1} \xi_r(F) \geq \log_d \sum_{r \geq r_1} \eta_r(F) \geq_n \log_d C_n^{r_1} + r_1,\ n \to \infty.$$

Proof. It was shown in [Djukova, Inyakin, 2003] that if F is determined by a perfect CNF and the condition of the lemma is fulfilled than

$$\eta_{r_1}(F) \approx C_n^{r_1} k^{r_1} \left(1 - k^{-r_1}\right)^u,\ n \to \infty.$$

Using the same technique we derive a similar estimate for the case when F is determined by an arbitrary CNF. This estimate is

$$\eta_{r_1}(F) \approx C_n^{r_1} k^{r_1} \left(1 - d^{-r_1}\right)^u,\ n \to \infty.$$

Estimating $\left(1 - d^{-r_1}\right)^u$, we have

$$\left(1 - d^{-r_1}\right)^u \approx e^{\frac{-u}{d^{r_1}}} \geq \left(\log_d u\right)^{-2d}.$$

Combining this with $\log_d \log_d u = o\left(\log_d C_n^{r_1}\right)$, we derive the assertion of the lemma.

Theorem 1 follows from Lemmas 4 and 5.

Lemma 6. For $n \leq u \leq k^{n^\beta}$ and $\beta < 1/2$, for almost all F from N_{un}^k

$$\sum_{r < r_1} \eta_r(F) = o\left(\sum_{r \geq r_1} \eta_r(F)\right),\ n \to \infty.$$

Proof. It was shown in [Djukova, Inyakin, 2003] that if F is determined by a perfect CNF and the condition of the lemma is fulfilled than

$$\sum_{r < r_1} \eta_r(F) = o\left(C_n^{r_1} k^{r_1}\right),\ n \to \infty.$$

Using the same technique we derive a similar estimate for the case when F is determined by an arbitrary CNF. This estimate for $u \leq k^{n^\beta}$ and $\beta < 1/2$ is

$$\sum_{r < r_1} \eta_r(F) = o\left(C_n^{r_1} d^{r_1}\right),\ n \to \infty.$$

On the other hand, according to Theorem 1, we have

$$\sum_{r \geq r_1} \eta_r(F) = \left(C_n^{r_1} d^{r_1}\right)^{1+\delta(n)}, \text{ where } \delta(n) \to 0 \text{ as } n \to \infty.$$

Consequently,

$$\sum_{r < r_1} \eta_r(F) = o\left(\sum_{r \geq r_1} \eta_r(F)\right),\ n \to \infty,$$

which proves the lemma.

Theorem 2 follows from Theorem 1 and Lemma 6. Theorem 3 follows from Lemma 6.

Remark 1. If a CNF has two identical conjunctions, then the set $B(F)$ obviously does not change after removing one of them. Consequently, when the number of maximal conjunctions of F is calculated, it is reasonable to consider the case when the CNF does not have identical conjunctions. For $u \le k^{n^{\beta}}$ and $\beta < 1/2$, this property is possessed by almost all CNFs of type (1).

Remark 2. The algorithm OPT++ is tested on random data. Experimental estimates of the number of maximal conjunctions and of the number of the "unnecessary" steps of the algorithm are obtained. The experimental results show that OPT++ works faster and performs considerably less "unnecessary" steps than other asymptotically optimal algorithms developed earlier. The part of "unnecessary" steps of the algorithm OPT++ increases as the part of missed feature values in the learning data increases; and it decreases as the fraction u/n increases.

Conclusion

New results are obtained in logical data analysis and in constructing recognition procedures based on the transformation of the normal forms of the characteristic functions of classes. An approach is developed, which allows to solve the problem of the transformation of the normal forms almost always with as asymptotical accuracy both for complete data and incomplete data. An algorithm OPT++ is developed for solving the problem of constructing the maximal conjunctions of a double-valued function F defined on k-ary n-dimensional arrays and determined by a CNF of u elementary disjunctions. The algorithm is justified both theoretically and experimentally. The theoretical justification is based on obtaining new asymptotic estimates of the typical values of the number of the maximal conjunctions and irreducible conjunctions of the function F. It is shown that OPT++ on experimental characteristics surpasses other similar algorithms.

Bibliography

[Chegis, Yablonskii, 1958] I. A. Chegis and S. V. Yablonskii, "Logical Methods for Control for Electric Circuits," Tr. Mat. Inst. im. V. A. Steklova, Akad. Nauk SSSR, 1958, Vol. 51, pp. 270 360 [in Russian].

[Dem'yanov, Djukova, 2007] E. A. Dem'yanov and E. V. Djukova, "On the Construction of Irredundant Coverings of an Integer Matrix," Zh. Vychisl. Mat. Mat. Fiz. 47 (3), 539-547 (2007) [Comput. Math. Math. Phys. 47 (3), 518-526 (2007)].

[Djukova, 1977] E. V. Djukova, "About Asymptotically Optimal Algorithm for Building Irreducible Tests," Dokl. Akad. Nauk SSSR 233 (4), 527-530 (1977) [in Russian].

[Djukova, 1982] E. V. Djukova, "Asymptotically Optimal Test Algorithms in Problems of Recognition," in Problems of Cybernetics, (Nauka, Moscow, 1982), Issue 39, pp. 165-199 [in Russian].

[Djukova, 1987] E. V. Djukova, "The Complexity of Realization of Some Recognition Procedures," Zh. Vychisl. Mat. Mat. Fiz. 27 (1), 114-127 (1987) [in Russian].

[Djukova, 1989] E. V. Djukova, "Recognition Algorithms of the Kora Type: Complexity of Implementation and Metric Properties," in Recognition, Classification, Forecast: Mathematical Methods and Their Applications (Nauka, Moscow, 1989), Vol. 2, pp. 99-125 [in Russian].

[Djukova, 2003] E.V. Djukova. Discrete (Logical) Recognition Procedures: Principles of Construction, Complexity of Realization and Basic Models // Pattern Recognition and Image Analysis. 2003. Vol. 13. No. 3. P. 417-425.

[Djukova, 2004] E. V. Djukova, "On the Implementation Complexity of Discrete (Logical) Recognition Procedures," Zh. Vychisl Mat. Mat. Fiz. 44 (3), 562-572 (2004) [Comput. Math. Math. Phys. 44 (3), 532-541 (2004)].

[Djukova, 2005] E. V. Djukova, "On the Number of Irreducible Coverings of an Integer Matrix," Zh. Vychisl. Mat. Mat. Fiz. 45 (5), 935-940 (2005) [Comput. Math. Math. Phys. 45 (5), 903-908 (2005)].

[Djukova, 2007] E. V. Djukova, "On Constructing the Irredundant Coverings of a Boolean Matrix," Doklady Akademii Nauk. 412 (1), 15-17 (2007) [in Russian].

[Djukova, Inyakin, 2003] E. V. Djukova and A. S. Inyakin, "Classification Procedures Based on the Construction of Class Coverings," Zh. Vychisl. Mat. Mat. Fiz. 43 (12), 1884-1895 (2003) [Comput. Math. Math. Phys. 43 (12), 1812-1822 (2003)].

[Djukova, Inyakin, 2008] E. V. Djukova and A. S. Inyakin, "On Asymptotically Optimal Constructing the Irredundant Coverings of an Integer Matrix," in Matematicheskie Voprosy Kibernetiki (Fizmatlit, Moscow, 2008), Issue 17, pp. 135-146 [in Russian].

[Djukova, Nefedov, 2009] E.V. Djukova, V.Y. Nefedov. The Complexity of Transformation of Normal Forms for Characteristic Functions of Classes // Pattern Recognition and Image Analysis. 2009. Vol. 19. No. 3. P. 435-440.

[Djukova, Peskov, 2002] E. V. Djukova and N. V. Peskov, "Search for Informative Fragments in Descriptions of Objects in Discrete Recognition Procedures," Zh. Vychisl. Mat. Mat. Fiz. 42, 741-753 (2002) [Comput. Math. Math. Phys. 42, 711-723 (2002)].

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

[Gurvich, Khachiyan, 1999] Gurvich V., Khachiyan L. On generating the irredundant conjunctive and disjunctive normal forms of monotone Boolean functions // Discrete Appl. Math. 1999. V. 96-97. № 1-3. P. 363-373.

[Jonson, 1988] Jonson D.S., Yannakakis M., Papadimitriou C.H. On General All Maximal Independent Sets // Information processing Letters. 1988. V. 27. P. 119-123.

[Yablonskii, 1986] S. V. Yablonskii, An Introduction to Discrete Mathematics (Nauka, Moscow, 1986) [in Russian].

[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].

Authors' Information

Elena Djukova *– Doktor Nauk, Major scientist, Dorodnicyn Computing Centre of RAS, Vavilova st. 40, 119333 Moscow, Russia; e-mail: edjukova@mail.ru*

Major Fields of Scientific Research: Pattern recognition, Discrete mathematics, Logical recognition procedures, Computational complexity of discrete problems

Vladimir Nefedov *– post-graduate student, Dorodnicyn Computing Centre of RAS, Vavilova st. 40, 119333 Moscow, Russia; e-mail: nefedov85@mail.ru*

Major Fields of Scientific Research: Discrete mathematics, Mathematical logic, Pattern Recognition