Methods of Combinatorial Analysis in Synthesis of Efficient Recognition Algorithms[1]

E. V. Djukova*, A. S. Inyakin, and N. V. Peskov*****

** Computer Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia*
e-mail: djukova@ccas.ru
*** Moscow State Pedagogical University, ul. Malaya Pirogovskaya 1, Moscow, 119992 Russia*
e-mail: andre_w@mail.ru
**** Scientific Council on Cybernetics, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia*
e-mail: nick@motor.ru

Abstract—This paper is a survey of the research into discrete analysis of feature descriptions of objects in recognition and classifications problems of large dimensions. New models of discrete recognition procedures and new approaches to searching for informative fragments of feature descriptions of objects are described. New results concerning the metric (quantitative) properties of informative fragments are stated.

INTRODUCTION

We consider the recognition problem in its standard setting for the case where the objects are described by a set of n integer-valued features $\{x_1, \ldots, x_n\}$ each of which can take finitely many admissible values [1]. Suppose that the set M of objects under examination can be represented as a union of subsets (classes) $K_1, \ldots, K_l$. There is a finite set $\{S_1, \ldots, S_m\}$ of objects from M such that it is known to which classes they belong (this set is a training sample). The training objects are represented by their descriptions. Given a set of feature values which describe some object S from M (generally, it is not known to which class it belongs), it is required to determine the class containing this object.

The discrete approach to recognition and classification problems is based on a combinatorial analysis of feature descriptions of objects. Its foundation has been laid by Yu.I. Zhuravlev, S.V. Yablonskii, and M.N. Vaintsvaig. The main advantage of the discrete (logic) recognition procedures is the possibility of obtaining a result when there is no information about the distribution functions and the training samples are small. However, it is often hard to apply the apparatus and methods of discrete mathematics because of the purely computational difficulties related to search, which arise at the stage of determining informative fragments of object descriptions. As a rule, a fragment is considered informative if it distinguishes between objects from different classes. Because of the necessity of a large-scale exhaustive search and the initially low performance of computing facilities, for many years main efforts were directed towards the development of a general complexity theory for discrete data analysis in problems of recognition and synthesis of asymptotically optimal algorithms for searching for informative fragments. In this connection, the works of V.A. Slepyan, V.N. Noskov, E.V. Djukova, and A.A. Andreev should be mentioned.

Construction of discrete procedures often uses the apparatus of logic functions, and informative fragments in feature descriptions of objects are determined by constructing admissible and maximal conjunctions of logic functions. These problems are reduced to constructing coverings and irredundant coverings, respectively, 1 of special integer matrices.

The notion of an irredundant covering of an integer 1 matrix plays a fundamental role and generalizes the notion of an irreducible covering of a Boolean matrix, which is well known in discrete mathematics. For the first time, it was introduced in [2] for the purpose of improving Kora-type algorithms and obtaining related asymptotic estimates.

In [2, 7], an asymptotically optimal algorithm for finding irredundant coverings in the case where the 1 number of rows in the matrix is small in comparison with the number of columns was constructed. In substantiating the asymptotic optimality of this algorithm, 2 asymptotic formulas for the typical values of the number of irredundant coverings and the length of an irredundant 1 1 covering were obtained. The technical basis for these estimates was first developed by Slepyan and Noskov in studying the metric (quantitative) properties of sets of irredundant tests of binary tables. The most 1 complete expositions of the complexity theory for problems of discrete data analysis and synthesis of

[1] This work was supported by the Russian Foundation for Basic Research, project no. 01-01-00575, and by the Presidium of the Russian Academy of Sciences, program "Intelligent Computer Systems," project no. 1.1.

Received February 27, 2003

Pattern Recognition and Image Analysis, Vol. 13, No. 3, 2003, pp. 426–432.

asymptotically optimal algorithms for searching for informative fragments are contained in [8, 9].

At present, some new approaches to constructing
3 discrete recognition procedures have arizen and new results related to studying the metric properties of sets of coverings of integer matrices have been obtained.

1. DISCRETE RECOGNITION PROCEDURES BASED ON CONSTRUCTING COVERINGS OF CLASSES

Let N_j be the set of admissible values of the feature x_j, where $j = 1, 2, \dots, n$. Suppose that H is a set of r different features of the form $\{x_{j_1}, \dots, x_{j_r}\}$ and $\sigma = (\sigma_1,$
4 $\dots, \sigma_r)$, where $\sigma_i \in N_{j_i}$ for $i = 1, 2, \dots, r$. The r-tuple set σ is called an elementary classifier generated by the features from H.

Let $S' = (a_1, \dots, a_n)$ be an object from the training sample. We denote the fragment $(a_{j_1}, \dots, a_{j_r})$ of the description of the object S' by (S', H).

An elementary classifier $(\sigma_1, \dots, \sigma_r)$ generated by the features from H can satisfy one of the following three conditions:

(i) each fragment of the form (S', K), where $S' \in K$, coincides with $(\sigma_1, \dots, \sigma_r)$;

(ii) several (not all) fragments of the form (S', H), where $S' \in K$, coincide with $(\sigma_1, \dots, \sigma_r)$;

(iii) none of the fragments of the form (S', H), where $S' \in K$, coincides with $(\sigma_1, \dots, \sigma_r)$.

The first case occurs very rarely, which makes it hardly possible that the sets of feature values to be processes have property (i). The fundamental difference in the informativeness between the remaining two conditions is that condition (ii) characterizes only some subset of the set of training objects from K, while condition (iii) characterizes all objects from K. Therefore, when it is important to consider the class K separately from the others classes, the sets of feature values satisfying (iii) seem to be more informative. In this case, it is most natural to assign an object S to a class K if neither the descriptions of all objects from the class K nor that of the object S to be recognized contain the given set of feature values.

We denote the set of all elementary classifiers generated by sets of features from $\{x_1, \dots, x_n\}$ by C. Each recognition algorithm is determined by some subset C^A of the set C. To be more precise, for each class $K \in \{K_1, \dots, K_l\}$, a subset $C^A(K)$ of C is constructed, and $C^A = \bigcup_{j=1}^{l} C^A(K_j)$.

Consider two objects $S' = (a'_1, a'_2, \dots, a'_n)$ and $S'' = (a''_1, a''_2, \dots, a''_n)$. We estimate the closeness of the objects S' and S'' with respect to a set of features $H = \{x_{j_1}, \dots, x_{j_r}\}$ by the value

$$B(S', S'', H) = \begin{cases} 0 & \text{if } a'_{j_t} = a''_{j_t} \text{ for } t = 1, 2, \dots, r, \\ 1 & \text{otherwise}. \end{cases}$$

The closeness between an object S' and an elementary classifier $\sigma = (\sigma_1, \dots, \sigma_r)$ generated by a set of features $H = \{x_{j_1}, \dots, x_{j_r}\}$ is estimated by the value

$$B(\sigma, S', H) = \begin{cases} 0 & \text{if } a'_{j_t} = \sigma_t \text{ for } t = 1, 2, \dots, r, \\ 1 & \text{otherwise}. \end{cases}$$

Suppose that $K \in \{K_1, \dots, K_l\}$ and $\overline{K} = \{K_1, \dots, K_l\} \backslash K$.

A particular model A of a recognition algorithm is determined by a principle for constructing the set C^A and by an estimate $\Gamma(S, K)$ of the membership of the object S in the class K, which is evaluated by voting over elementary classifiers from $C^A(K)$. For instance, it is assumed that an elementary classifier σ from $C^A(K)$ generated by a set of features H votes for the membership of an object S in the class K if $B(\sigma, S, H) = 0$. An object S is assigned to the class with maximum membership estimate $\Gamma(S, K)$ (if there are several such classes, then the algorithm refuses to recognize the object).

A fragment (S', H), where $S' \in K$, is called a representative set for K if $B(S', S'', H) = 1$ for any training object S'' not belonging to the class K. A fragment $(S'$
$H)$, where $S' \in K$, is called an irredundant representa- 1
tive set for K if (i) $B(S', S'', H) = 1$ for any training object S'' from $\overline{K}$ and (ii) for any set $H' \subset H$, $\overline{K}$ contains a training object S'' for which $B(S', S'', H') = 0$.

In the classical model of the algorithm of voting
over (irredundant) representative sets, the set $C^A(K)$ 1
consists of all (irredundant) representative sets for K. 1
Its simplest modification estimates the likelihood that an object S belongs to a class K by the value

$$\Gamma_1(S, K) = \frac{1}{|C^A(K)|} \sum_{(S', H) \in C^A(K)} (1 - B(S, S', H)),$$

here and in what follows, $|N|$ denotes the cardinality of 5
the set N.

It is accepted that short representative sets are more informative; for this reason, in order to improve the quality of recognition and reduce computational burden, only short representative sets are usually considered. These may be representative sets of bounded
length or irredundant representative sets. 1

We use the following notation: M^k_{mn}, where $k \geq 2$, is the set of all $m \times n$ matrices with elements from $\{0,$
$1, \dots, k-1\}$ and E^r_k is the set of all k-ary r-tuples. 6 7

Suppose that $L \in M^k_{mn}$ and $\sigma \in E^r_k$. A set H of r different
columns of the matrix L is called a σ-covering if
8 the submatrix L^H of L formed by the columns from H
does not contain the row σ. We call a set H of r different
1 columns of the matrix L an irredundant σ-covering if (i)
8 the submatrix L^H does not contain the row $\sigma = (\sigma_1, \dots,$
$\sigma_r)$ and (ii) if $p \in \{1, 2, \dots, r\}$, then L^H contains at least
one row of the form $(\sigma_1, \dots, \sigma_{p-1}, \beta_p, \sigma_{p+1}, \dots, \sigma_r)$,
where $\beta_p \neq \sigma_p$.

Note that, if $\sigma = (\sigma_1, \dots, \sigma_r)$, then a set H of columns
1 of the matrix L is an irredundant σ-covering if and only
if the following two conditions hold:

(i) L^H does not contain the row $(\sigma_1, \dots, \sigma_r)$;

8 (ii) L^H contains a submatrix of form

$$\begin{bmatrix} \beta_1 & \sigma_2 & \sigma_3 & \dots & \sigma_{r-1} & \sigma_r \\ \sigma_1 & \beta_2 & \sigma_3 & \dots & \sigma_{r-1} & \sigma_r \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_1 & \sigma_2 & \sigma_3 & \dots & \sigma_{r-1} & \beta_r \end{bmatrix}$$

(with an accuracy up to some rows permuted), where β_p
8 $\neq \sigma_p$ for $p = 1, 2, \dots, r$. We call such a submatrix a σ-
8 submatrix.

1 In particular, an irredundant $(0, \dots, 0)$-covering of a
Boolean matrix is an irreducible covering [15].

Let $K \in \{K_1, \dots, K_l\}$. A training table can be treated as a pair of matrices L_1 and L_2, where L_1 is the matrix consisting of the descriptions of the training objects from the class K and L_2 is the matrix consisting of the descriptions of the remaining training objects. Obviously, the elementary classifier of form $(\sigma_1, \dots, \sigma_r)$
determined by a pair (S_i, H), where $S_i \in K$ and $H = \{x_{j_1},$
1 $\dots, x_{j_r}\}$, is an (irredundant) representative set for K if
and only if the set of the columns with numbers $j_1, \dots,$
1 j_r in the matrix L_2 is an (irredundant) $(\sigma_1, \dots, \sigma_r)$-covering.

Now, consider discrete models based on voting over σ-coverings of a class, namely, the model of voting
over the coverings of a class and the model of voting
9 over the antirepresentative sets of a class [10, 12]. The
application of these models somewhat reduces the computational burden if $|K| < |\bar{K}|$ (for example, when the number of classes is large). Below, we describe these models.

1 In the model of voting over the (irredundant) coverings of a class, the set $C^A(K)$ consists of the elementary
1 classifiers generated by (irredundant) coverings of the
matrix L_1. The likelihood that an object S belongs to a class K is estimated (in the simplest modification) by the value

$$\Gamma_2(S, K) = \frac{1}{|C^A(K)|} \sum_{\sigma \in C^A(K)} B(\sigma, S, H).$$

Now, consider the model of voting over the antirep- 9
resentative sets. The elementary classifier σ generated
by an (irredundant) σ-covering of the class K is an (irre- 1 1
dundant) antirepresentative set if it coincides with at 9
least one fragment of form (S', H), where S' is a training object from $\bar{K}$. The likelihood that an object S belongs to the class K is estimated (in the simplest modification) by the value

$$\Gamma_3(S, K) = \frac{1}{|C^A(K)|} \sum_{\sigma \in C^A(K)} B(\sigma, D, H).$$

Note that a representative set for a class K is antire- 9
presentative for $\bar{K}$. It is easy to show that, at $l = 2$, both described models assign the object S to the same class.

The suggested models were tested for real-life problems of medical prediction. Two problems were considered, the predicting of survivability of patients with
osteogenic sarcoma during one year after a course of 10
treatment and predicting of pathomorphosis (the degree 11
of destruction of cancerous cells after a course of chemotherapy). The preceding study showed that the first problem is more difficult. To evaluate the efficiency of the recognition procedures, the cross-validation method was used. The efficiency of the classical model of voting over representative sets was 61% for the survival problem and 83% for the pathomorphosis prob- 11
lem, while the efficiency of the algorithm of voting over coverings of classes for the survival problem was 75%. To improve the quality of the classical algorithm, the approach described in the next section was applied. As a result, the efficiency of this algorithm increased to 75% and 94%, respectively.

2. DETERMINATION OF TYPICAL OBJECTS IN CLASSES AND CONSTRUCTION OF MOST "WEIGHTY" ELEMENTARY CLASSIFIERS

In solving an applied recognition problem, it is interesting to try to evaluate the efficiency of the constructed algorithm in recognizing objects not included in the training sample. For instance, the well-known cross-validation method can be used. Unfortunately, for a number of applied problems, cross-validation does not always indicate a high efficiency of the algorithms described in Section 1. Such a situation occurs when the classes are poorly separated from each other (i.e., each class contains many objects whose descriptions are similar to the descriptions of objects not belonging to this class). In this case, although the constructed algorithms correctly recognize the objects which they "know" (i.e., those used in constructing the algorithms), they poorly recognize "new" objects. This section describes an approach making it possible to substantially improve the quality of the recognition algorithms [10–13].

We suggest to partition the training sample into two 12 subsamples, the first (base) to be used for constructing the set of representative sets and the second (test) for evaluating their weights. The sample should be partitioned so that the objects at the interface between 13 classes be included in the test subsample and all the 13 remaining (typical) objects, in the base subsample. Practical experiments on applied problems show that such a partitioning increases the number of short representative sets and, thereby, makes it possible to improve the quality of the recognition algorithms.

The typical objects can be determined with the use of the cross-validation method. The training objects correctly recognized in a cross-validation test are 13 included in the base subsample, and all the remaining 13 objects are included in the test subsample. This approach is fairly effective, but it is rather time-consuming when applied to problems of large dimensions.

The computational burden can be reduced by the method of determination of typical object suggested in [10–13]; it is based on evaluating the informativeness of separate feature values. The method is as follows.

Suppose that $S' \in K_i$, $i \in \{1, 2, \ldots, l\}$, and $j \in \{1, 2, \ldots, n\}$. We set

$$\bar{K}_i = \bigcup_{q=1}^{l} K_q \backslash K_i,$$

$$\mu_{ij}^{(1)}(S') = \frac{1}{K_i} \sum_{S'' \in K_i} (1 - B(S', S'', \{x_j\})),$$

and

$$\mu_{ij}^{(2)}(S') = \frac{1}{\bar{K}_i} \sum_{S'' \in \bar{K}_i} (1 - B(S', S'', \{x_j\})).$$

The values $\mu_{ij}^{(1)}(S')$ and $\mu_{ij}^{(2)}(S')$ characterize the proximity of the object S' to its class and to the other classes, respectively. We call

$$\mu_{ij}(S') = \mu_{ij}^{(1)} - \mu_{ij}^{(2)}$$

the weight of the value of the feature x_j for the object S'.

Suppose that a real number μ such that $-1 \le \mu \le 1$, the threshold minimal informativeness of feature values, is given. We say that the value of a feature x_j for S' is typical if $\mu_{ij}(S') > \mu$.

Let p be an integer such that $1 \le p \le n$. We consider an object S' typical for a class K_i with respect to the threshold p if the inequality $\mu_{ij}(S') > \mu$ holds for at least p features.

Note that the thresholds μ and p can be chosen from heuristic considerations; e.g., we can set $\mu = 0$ and $p = [n/2]$. Then, the value of a feature x_j for S' is typical for a class K_i if it is encountered more frequently for objects from K_i than for objects from K_i, and an object S' is typical for K_i if at least half feature values in its description are typical for K_i.

Suppose that the training sample is partitioned into base and test subsamples by one of the methods 12 described above. For the base sample, we construct a set of representative sets. To each of the constructed representative sets, we assign a weight computed on the test subsample. 13

Let ω be the representative set of a class $K \in \{K_1, \ldots K_l\}$ generated by a pair (S', H), where S' is an object from the base sample; by $\delta(K, \omega)$, we denote the number of objects in the test sample for which the representative tuple ω votes "correctly" and by $\delta(\bar{K}, \omega)$, the 4 number of objects in the test sample for which it votes "incorrectly." Let us define a function $\nu_{(S', H)}$ by one of the following formulas:

(i) $\nu_{(S', H)} = \delta(K, \omega)$;

(ii) $\nu_{(S', H)} = \begin{cases} \delta(K, \omega) - \delta(\bar{K}, \omega) \\ \text{if } \ \delta(K, \omega) > \delta(\bar{K}, \omega), \\ 0 \ \text{ if } \ \delta(K, \omega) < \delta(\bar{K}, \omega); \end{cases}$

(iii) $\nu_{(S', H)} = \dfrac{1 + \delta(K, \omega)}{1 + \delta(\bar{K}, \omega)}$.

The membership of an object S in a class K is estimated by the value

$$\Gamma_4(S, K) = \frac{1}{|C^A(K)|}$$

$$\times \sum_{(S', H) \in C^A(K)} \nu_{(S', H)}(1 - B(S, S', H)),$$

We define the informative weight of a feature x_j as

$$I_j = \frac{\sum\limits_{(S', H) \in C^A(K),\, x_j \in H} \nu_{(S', H)}}{\sum\limits_{(S', H) \in C^A} \nu_{(S', H)}}.$$

3. THE METRIC PROPERTIES OF THE SET OF COVERINGS

Traditionally, analyzing an improvement in the speed of algorithms based on construction of coverings of Boolean and integer matrices involves evaluating asymptotic estimates of the typical values of the most important quantitative characteristics of this set. Such characteristics are the number of coverings and the length of a covering. Technically, evaluating these estimates is very difficult. Of most importance and com-

1 plexity is analyzing the metric properties of irredundant coverings.

We use the following notation:

Ψ_0 is the interval $(\log_k mn, n)$,

Ψ_1 is the interval

$$\left(\frac{1}{2}\log_k mn - \frac{1}{2}\log_k\log_k mn - \log_k\log_k\log_k n, \text{ and}\right.$$

$$\left.\frac{1}{2}\log_k mn - \frac{1}{2}\log_k\log_k mn + \log_k\log_k\log_k n\right);$$

$a_n \approx b_n$ means that $\lim(a_n/b_n) = 1$ as $n \longrightarrow \infty$.

Suppose that $L \in M^k_{mn}$; $\sigma \in E^r_k$; $C(L, \sigma)$ is the set of all pairs of form (H, σ), where H is a σ-covering of the matrix L; $B(L, \sigma)$ is the set of all pairs of form (H, σ), 1 where H is an irredundant σ-covering of the matrix L; and $S(L, \sigma)$ is the set of all σ-submatrices of the matrix L. We set

$$C(L) = \bigcup_{r=1}^{n}\bigcup_{\sigma\in E^r_k} C(L,\sigma), \quad B(L) = \bigcup_{r=1}^{n}\bigcup_{\sigma\in E^r_k} B(L,\sigma),$$

and

$$S(L) = \bigcup_{r=1}^{n}\bigcup_{\sigma\in E^r_k} S(L,\sigma).$$

Earlier, in [2, 7–9], the case was studied where the number of rows in a matrix is by an order of magnitude smaller that the numbers of columns, i.e., $m^\alpha \le n \le k^{m^\beta}$, where $\sigma > 1$ and $\beta > 1$. It was shown that, in this case, the value $|B(L)|$ almost always (for almost all matrices from M^k_{mn}) asymptotically coincides with $|S(L)|$ as $n \longrightarrow \infty$ and is by an order of magnitude smaller than the number of coverings. On the basis of this observation, an asymptotically optimal algorithm for searching for coverings from $B(L)$ was constructed.

In [11], the completely opposite case where $n^\alpha \le m \le k^{n^\beta}$ for $\alpha > 1$ and $\beta < 1$ was considered and asymptotic expressions for the typical values of the number of 8 σ-submatrices and the order of a σ-submatrix were obtained. These results are contained in Theorem 1.

Theorem 1. If $n^\alpha \le m \le k^{n^\beta}$, where $\alpha > 1$ and $\beta < 1$, then

$$|S(L)| \approx \sum_{r\in\Psi_1} C^r_n C^r_m r!(k-1)^r k^{r-k^2},$$

as $n \longrightarrow \infty$ for almost all matrices L from M^k_{mn}, and the orders of almost all submatrices from $S(L)$ belong to the interval Ψ_1.

It was shown that almost all matrices have no σ-submatrices of orders larger than $\log_k mn$. Let

$$S_1(L) = \bigcup_{r\le\log_k mn}^{n}\bigcup_{\sigma\in E^r_k} S(L,\sigma),$$

The following theorem is valid.

Theorem 2. For almost all matrices $L \in M^k_{mn}$,

$$|S_1(L)| = 0.$$

as $n \longrightarrow \infty$.

In addition, for a practically general case, asymptotic expressions for the typical values of $|C(L)|$ and of the length of a σ-covering were obtained; namely, the following theorem was proved.

Theorem 3. If $m \le k^{n^\beta}$, where $\beta < 1$, then

$$|C(L)| \approx \sum_{r\in\Psi_0}^{n} C^r_n k^r$$

as $n \longrightarrow \infty$ for almost all matrices L from M^k_{mn}, and the lengths of almost all coverings from $C(L)$ belong to the interval Ψ_0.

It was also shown that, at $r \le \log_k m - \log_k(\log_k m \times \ln kn)$, almost all matrices have no coverings of length r. We set $r_0 = \log_k m - \log_k(\log_k m \times \ln kn)$ and

$$C_1(L) = \bigcup_{r\le r_0}\bigcup_{\sigma\in E^r_k} C(L,\sigma)$$

Theorem 4. For almost all matrices $L \in M^k_{mn}$,

$$|C_1(L)| = 0.$$

as $n \longrightarrow \infty$.

It was shown by comparing the found estimates that, if $n^\alpha \le m \le k^{n^\beta}$, where $\alpha > 1$ and $\beta < 1/2$, then the number of σ-submatrices is almost always by an order of magnitude larger than the number of irredundant 1 σ-coverings.

Theorem 5. If $n^\alpha \le m \le k^{n^\beta}$, where $\alpha > 1$ and $\beta < 1/2$, then $|S(L)|/|B(L)| \longrightarrow \infty$ as $n \longrightarrow \infty$ for almost all matrices L from M^k_{mn}.

4. THE TAXONOMY PROBLEM AND IRREDUNDANT COVERINGS OF INTEGER MATRICES 1

In [16], application of cluster analysis methods to processing data sets where it is difficult to define distance functions was considered and an approach to solving taxonomy problems with integer data by constructing σ-coverings of classes was developed. This approach is based on the following considerations.

Consider the situation where it is required to determine the degree of membership of an object S in a group of objects M. If the description of the object S 4 contains a tuple of feature values which is not contained in the descriptions of the objects from M, then we can say that uniting S with M destroys the intrinsic structure of the set M. Considering various combinations of feature values not contained in the descriptions of objects from M, we can estimate the proximity of the object S to the set M. Thus, in determining the degree of proximity of an object to a set M, the informative sets of feature values are those missing in the descriptions of all objects from the class M.

In [16], an algorithm based on these considerations was constructed; in many cases, it significantly improved clustering in model and real-life problems in comparison with well-known algorithms, such as the algorithms based on the nearest and farthest neighbor principles and on the choice of a central element, which use the Hamming distance as a distance function.

In the constructed clustering algorithm, estimating the proximity of an object S to a set M is reduced to 1 searching for irredundant coverings of the integer matrix formed from the feature descriptions of the 1 objects from M. The problem of searching for irredundant coverings of an integer matrix is solved on the basis of constructing irreducible coverings of a Boolean matrix obtained in a special way from the initial matrix.

CONCLUSION

This paper is concerned with studying discrete recognition and classification procedures. The most important stage in construction of these procedures is searching for informative fragments of feature descriptions of objects. The paper suggests new approaches to searching for such fragments.

(i) General principles of construction of discrete (logic) recognition procedures are considered and new models are described that make it possible to extend the domain of applicability of discrete analysis methods to recognition problems.

(ii) An approach to improving the efficiency of recognition algorithms based on determining typical training objects in every class is described.

(iii) Results related to the development of methods for solving problems of cluster analysis with integer data obtained by constructing coverings of integer matrices are presented.

REFERENCES

1. Zhuravlev, Yu.I., On an Algebraic Approach to Recognition or Classification Problems, in *Problemy kibernetiki* (Problems of Cybernetics), Moscow: Nauka, 1978, issue 33, pp. 5–68.
2. Djukova, E.V., On the Complexity of Implementation of Some Recognition Procedures, *Zh. Vychisl. Mat. Mat. Fiz.*, 1987, vol. 27, no. 1, pp. 114–127.
3. Dmitriev, A.I., Zhuravlev, Yu.I., and Krendelev, F.P., On Mathematical Principles of Classification of Objects or Phenomena, in *Diskretnyi analiz* (Discrete Analysis), Novosibirsk: Inst. Mat., Sibirsk. Otd., Akad. Nauk SSSR, 1966, issue 7, pp. 3–17.
4. Djukova, E.V., Asymptotically Optimal Test Algorithms in Recognition Problems, in *Problemy kibernetiki* (Problems of Cybernetics), Moscow: Nauka, 1982, issue 39, pp. 165–199.
5. Baskakova, L.V. and Zhuravlev Yu.I., A Model of Recognition Algorithms with Representative Sets and Systems of Support Sets, *Kibernetika*, 1978, no. 4, pp. 131–137.
6. Vaintsvaig, M.N., Pattern Recognition Learning Algorithm “Kora,” in *Algoritmy obucheniya raspoznavaniyu obrazov* (Pattern Recognition Learning Algorithms), Moscow: Sovetskoe Radio, 1973, pp. 82–91.
7. Djukova, E.V., Kora-Type Recognition Algorithms: Complexity of Implementation and Metric Properties, in *Raspoznavanie, klassifikatsiya, prognoz (matematicheskie metody i ikh primenenie)* (Recognition, Classification, Forecasting: Mathematical Methods and Their Application), Moscow: Nauka, 1989, issue 2, pp. 99–125.
8. Djukova, E.V. and Zhuravlev, Yu.I., Discrete Analysis of Feature Descriptions in Recognition Problems of High Dimension, *Zh. Vychisl. Mat. Mat. Fiz.*, 2000, vol. 40, no. 8, pp. 1264–1278.
9. Djukova, E.V. and Zhuravlev, Yu.I., Discrete Methods of Information Analysis in Recognition and Algorithm Synthesis, *Pattern Recogn. Image Anal.*, 1997, vol. 7, no. 2, pp. 192–207.
10. Djukova, E.V. and Peskov, N.V., On Discrete Recognition Procedures Based on Constructing Coverings of Classes, *Dokl. 10 Vserossiiskoi konferentsii po matematicheskim metodam raspoznavaniya obrazov* (Proc. 10th All-Russia Conf. on Math. Methods of Pattern Recognition), Moscow, 2001, pp. 48–51.
11. Djukova, E.V. and Peskov, N.V., A Search for Informative Fragments of Object Descriptions in Discrete Recognition Procedures, *Zh. Vychisl. Mat. Mat. Fiz.*, 2002, vol. 42, no. 5, pp. 743–755.
12. Djukova, E.V. and Peskov, N.V., Selection of Typical Objects in Classes for Recognition Problems, *Pattern Recogn. Image Anal.*, 2002, vol. 12, no. 3, pp. 243–249.
13. Djukova, E.V. and Peskov, N.V., On Some Approaches to Evaluating Informative Characteristics of Training Samples, *Dokl. 9 Vserossiiskoi konferentsii po matematicheskim metodam raspoznavaniya obrazov* (Proc. 9th All-Russia Conf. on Math. Methods of Pattern Recognition), Moscow, 1999, pp. 181–183.
14. Djukova, E.V. and Peskov, N.V., Informativeness of Features, Separate Feature Values, and Fragments of Object Descriptions, *Dokl. 10 Vserossiiskoi Konferentsii po matematicheskim metodam raspoznavaniya obrazov*

(Proc. 10th All-Russia Conf. on Math. Methods of Pattern Recognition), Moscow, 2001, pp. 201–204.

15. *Diskretnaya matematika i matematicheskie voprosy kibernetiki* (Discrete Mathematics and Mathematical Questions of Cybernetics), Yablonskii, S.V. and Lupanov, O.B., Eds., Moscow: Nauka, 1974.

16. Djukova, E.V. and Inyakin, A.S., Taxonomy Problem
1 and Irredundant Coverings of Integer Matrix, in *Soobshcheniya po prikladnoi matematike* (Communications in Applied Mathematics), Moscow: Vychisl. Tsentr, Ros. Akad. Nauk, 2002.

Elena V. Djukova. Born 1945. Graduated from Moscow State University in 1967. Received her PhD (Kandidat Nauk) degree in physics and mathematics in 1979 and doctoral degree in physics and mathematics in 1997. Leading Researcher at the Computer Center, Russian Academy of Sciences. Scientific interests: discrete mathematics and mathematical methods of pattern recognition and image processing. Author of about 60 publications.

Andrei S. Inyakin. Born 1978. Graduated from Moscow State University in 2000. Postgraduate student at Moscow State Pedagogical University. Scientific interests: discrete mathematics and mathematical methods of pattern recognition and image processing. Author of four publications.

Nikolai V. Peskov. Born 1978. Graduated from Moscow State University in 2000. Postgraduate student at the Scientific Council on Cybernetics, Russian Academy of Sciences. Scientific interests: discrete mathematics and mathematical methods of pattern recognition and image processing. Author of six publications.

SPELL: 1. irredundant, 2. optimality, 3. arizen, 4. tuple, 5. cardinality, 6. ary, 7. tuples, 8. submatrix, 9. antirepresentative, 10. osteogenic, 11. pathomorphosis, 12. subsamples, 13. subsample