# Discrete Methods of Information Analysis in Recognition and Algorithm Synthesis[1]

**E. V. Djukova and Yu. I. Zhuravlev**

*Computer Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 117967 Russia*
*e-mail: djukova@ccas.ru, zhuravlev@ccas.ru*

**Abstract**—The paper sets out the general principles that underlie discrete methods of information analysis in recognition problems. An approach is proposed whereby recognition procedures can be constructed using logical functions. Basic models are described and matters related to an assessment of complexity in their realization are discussed.

## INTRODUCTION

The paper deals with an inquiry into a crucial issue of classical pattern recognition—that of a search for informative fragments in the descriptions of objects being recognized. Such fragments act as elementary classifiers and enable one to distinguish between objects falling in different classes. This issue remains topical because in the statement of a practical recognition problem, original descriptions of objects usually contain all observable and measurable characteristics or parameters. The objects are thus described by several tens or even hundreds of variables. In particular, this is true of medical diagnosis, geological, engineering, and sociological forecasting, and elsewhere. It is this circumstance, that is, the high dimension of descriptions, that largely make impossible for procedures of traditional computational mathematics to be directly applied to recognition problems and makes it necessary to resort to original mathematical constructions in their own right.

Note that in an analysis of complex descriptions by statistical methods one has to accept on trust additional hypotheses of probabilistic character, i.e., to impose rather strong requirements for spaces of objects under study. In addition, for the results obtained by statistical procedures to be reliable, extremely large arrays of precedents (learning information) must be available. That is why special emphasis is currently being placed on direct methods for the synthesis of optimal (proper) algorithms (the algebraic approach [28, 29, 32, 44–49]) and the combinatorial analysis of description space structure [1–7, 11, 12, 14–16, 18, 19, 21–27, 30, 33, 34, 42, 43, 53, 54] based on the findings and ideas of discrete analysis, which provides the framework for the present work.

The tools and procedures of discrete mathematics have several advantages to offer. Above all, it is possible to obtain a result even if there is no information about the applicable distribution functions and the available learning samples are small. In many cases, however, it is difficult, if feasible at all, to apply the discrete approach because of the purely computational difficulties associated with the need to carry out an exhaustive search. In particular, there arises the "canonically difficult" problem of building all irreducible (irredundant) coverings for a Boolean matrix, which may be formulated as the problem of transforming the conjunctive normal form of a logical function into its disjunctive form—the problem known from classical works of S.V. Yablonskii [55], and others.

Since the 1960s, many researchers have made attempts to find algorithms and procedures effective in a sense. In the mid-1970s, the first truly efficient procedures were developed for the cases of practical importance that are encountered in the synthesis of recognition algorithms, and a theoretical justification thereof was given. In particular, an approach was proposed whereby recognition problems could be handled by discrete analysis techniques. In a sense, it was asymptotically optimal from the viewpoint of computational difficulty. Its basic idea was to drop some of the conditions that the sought informative fragments of an object description should meet and to develop an approximate solution. This would result in a set of fragments rather close in properties to the sought one. Within the framework of this idea, a new method was proposed for the search of irreducible coverings for a Boolean matrix [15, 16, 18, 19, 21, 22, 25].

The method was tested in quite a number of problems, such as the search for irredundant representative descriptors, the search for irredundant tests, and the search for maximal conjunctions of logical functions. In addition, it was demonstrated by modifying some well-known concepts (e.g., the concept of an irreducible covering for a Boolean matrix) that the above problems have much in common in solution methodology [21, 22, 25].

A most important thing about the results thus obtained was that they led to efficient algorithms useful

[1] The study was supported by the Russian Foundation for Basic Research, project nos. 96-01-00543 and 96-01-00552).

Received January 26, 1997

in applied recognition problems [8, 17, 20]. This is noteworthy, since the number of fragments increases exponentially with the dimension of a description, it would be unreal to look for a way out solely in ever faster and more powerful computers.

Sections 1 through 4 of this paper characterize briefly the mathematical methods used in pattern recognition, describe the discrete approach, and define the principles on which logical recognition procedures are built. Also, a brief historical overview is given of the results achieved in developing asymptotically optimal discrete methods of information analysis in recognition problems. These sections give a fairly full idea about the subject of the study.

Sections 5 through 7 sum up the results reported in [15, 16, 18, 19, 21, 22, 24, 25].

Section 5 introduces the concept of a irredundant covering for an integer-valued matrix, which is basic to all of the technique described. An asymptotically optimal algorithm for the search of two types of irredundant coverings is proposed and substantiated. The first type arises when we build informative fragments generated by irredundant representative descriptors and irredundant tests. The second is encountered when we must transform the normal forms of logical functions. Asymptotic estimates are given for the typical numbers of irredundant coverings and the typical length of a irredundant covering. Taken together, the ideas and results stated in this section provide a methodological basis common to all the subsequent sections.

Section 6 deals with what are known as the metric properties of sets of elementary classifiers with the proviso that the number of features is significantly greater than that of objects in the learning sample. Asymptotic estimates are given for typical numbers of representative descriptors, for the typical length of a representative descriptor, for the same typical characteristics of a set of irredundant representative descriptors and a set of irredundant tests. The asymptotic estimates given in Section 5 are used there.

Using the same technical basis (the construction of irredundant coverings), Section 7 develops asymptotically optimal methods of construction and examines the metric properties of the abridged disjunctive normal forms (cnfs) of completely defined and partial two-valued functions. These constructions are primarily of interest to classical discrete mathematics.

## 1. MATHEMATICAL METHODS OF PATTERN RECOGNITION

The theory and practice of recognition have advance in several directions, and quite a number of approaches thereto have been accumulated by now. In most cases, the central problem is the synthesis of algorithms, extreme in accuracy in a particular sense, for a specified learning sample.

The mathematical methods used in pattern recognition may arbitrarily be divided into four classes: (1) statistical methods [9, 10, 38–40], (2) optimization methods for many-parameter recognition models [50–52], (3) methods based on the idea of joint application and adjustment of sets of recognition algorithms from heuristic families [26, 28, 29, 32, 35–37, 44–49], and (4) discrete methods of information analysis and algorithm synthesis [1–7, 11, 12, 14–16, 18, 19, 21–27, 30, 33, 34, 42, 43, 53, 54]. We note here that this study used the discrete approach, and the sections that follow examine it in more detail.

Statistical methods of data analysis and recognition algorithm construction use the tools of probability theory and mathematical statistics. They invoke additional statistical hypotheses about the probabilistic character of the learning sample and the distribution of objects by classes. The most important results of the statistical approach are the recognition algorithm reliability estimates formed for models of limited capacity. For such methods to be used accurately and well, large learning samples are needed.

Multiparameter recognition models are ordinarily optimized by classical methods, such as the least squares, the penalty functions, linear, quadratic and dynamic programming, etc. Difficulties mainly arise in the optimization, owing to the complex structure of algorithm models used as optimization domains (such models are many-parameter families of mappings, the parameters usually being very great in number and diverse in properties). This requires that special approaches should be developed to the formulation and solution of such problems.

As general considerations clearly suggest, algorithm models which are rather meager in parameters are bound to yield a solution of low quality simply because they would not contain appropriate algorithms. On the other hand, models with many of parameters and having therefore an elaborate arrangement would, as already noted, lead to optimization problems that are complicated from a practical point of view. They call for approximate optimization methods to develop locally extreme algorithms. In many cases, however, approximate optimization methods will not offer a way out, since locally extreme solutions obtained within a "plentiful" family might be worse than an optimal solution found in a "meager" family. This circumstance gave rise to a school of thought that seeks to use jointly several algorithms, not so good in quality individually, so that the resultant algorithm is of a substantially better quality than would be an algorithm developed by optimization using original heuristic models.

Among the methods that draw upon the idea of joint usage of sets of recognition algorithms, especially strong emphasis was placed on algebraic adjustment methods. Lying at their roots are mathematical concepts, such as closure, bases in operator spaces, completeness, and other such concepts. The starting point for the algebraic approach was the idea that, rather than using heuristic algorithm families as fixed domains where a solution ought to be sought, we can select in a certain manner some algorithms from the available families and, by performing appropriate operations (adjustments) on them, build optimal algorithms targeted at particular problems. At first, the adjustments were some operations on real matrices and the original algorithm families were separation and estimation algorithms. Later, many other specific families of algorithms and adjustments were explored. It should be noted that the quality of a solution seems to be largely dependent on what algorithms were chosen as the basis, and this issue has yet to be clarified. Nevertheless, the algebraic approach has now become a general theoretical basis for research into recognition problems.

## 2. DISCRETE METHODS OF FEATURE SPACE STRUCTURE ANALYSIS

It was in the mid-1960s that the discrete approach to the recognition problem made especially great strides [11, 14, 27].

Recognition (learning by precedents) is, in fact, an extrapolation problem. In contrast to traditional mathematical formulations, however, in pattern recognition this problem is often tackled in a space of heterogeneous features of a very high dimension often running into several tens or even hundreds (although the objects in the learning sample described by these features may be relatively small in number). It is this fact that gave rise to what came to be known as discrete methods of information analysis (based on methods of discrete mathematics). These methods do not call for strong assumptions about the properties of the test object (such as its metrizability, compliance with probabilistic laws, etc.).

In discrete methods, the key object is the aggregate of all subsets of a formal set of features, from which we must select special sets of features. These sets of features should contain certain information about classes; e.g., they should enable us to distinguish between objects from different classes or a given object from the objects of the other classes. Other and more complex requirements as to the information content may be imposed. Sometimes, it would be of interest, on the contrary, to have sets of features that contain indiscriminating information.

It should be noted here that since the original feature space has a high dimension and the cardinality of the aggregate of subsets involved increases exponentially with the number of features, it is very difficult to explore all of its subsets. To tackle this problem, a specific discrete technique was developed and became quite common. It is used in optimization methods, in adjustment methods (in developing both basis algorithms and adjustment operations), etc. With informative sets of features thus brought to light, it is possible to analyze the source information qualitatively in order, say, to classify the features according to the information they carry and to reduce their number.

Discrete methods also gave rise to a number of elaborate heuristics called logical recognition procedures (their construction involves logical functions). These are, above all, test algorithm models, algorithms of voting by representative descriptors, the exhaustive search of conjunctions, etc. In organization, these algorithms are exact (proper) on the learning material and were initially intended solely for use in the processing of integer-valued, mainly binary, information. The discrete analysis methods have been used in recent years to process real-valued information, thus providing a basis on which more universal models of logical recognition procedures are being developed (models with recording).

As already noted, when this approach is applied to recognition problems, there arise computational difficulties as it involves a large-scale exhaustive search. As a rule, the need to generate informative sets of features entails the need to solve the laborious problem of constructing irreducible (irredundant) coverings for a Boolean matrix. It may be also stated as the problem of transforming normal forms of Boolean functions (multiplication of logical brackets). So matters related to an analysis of the complexity of this problem and associated problems were very important from the start.

Rather interesting results were obtained in this respect. In particular, asymptotically optimal methods for generating informative sets of features were developed so that efficient implementations for logical recognition procedures could be constructed and substantiated. These methods are asymptotically optimal in the following sense: they involve discarding some of the conditions that are to be met by the sought constructions (sets of features, elementary conjunctions, etc.). This leaves an approximate solution to the problem, which means that the set of constructions thus built is nearly always the same in cardinality as the sought set (and, of course, includes it as a subset). The principal results were obtained from a study into the statistical properties of irreducible (irredundant) coverings for a Boolean matrix [15, 16, 18, 19, 21, 22]. The concept of a irredundant covering for an integer Boolean matrix was introduced and investigated in [25] to improve and generalize the technique for constructing informative sets of features and for forming the necessary asymptotic estimates.

## 3. RECOGNITION PROCEDURES BASED ON LOGICAL FUNCTIONS: PRINCIPLES OF CONSTRUCTION AND BASIC MODELS

We now describe the general principles that underlie the construction of logical recognition algorithms.

These algorithms are mainly intended to process integer-valued information. The descriptions of objects are given as sets of values of $t$-valued features $x_1, \ldots, x_n$, $t \geq 2$.

Objects of a class $K$ are separated from objects of other classes by constructing so-called admissible or almost admissible conjunctions of a partial (incompletely defined) two-valued function $f_K(x_1, \ldots, x_n)$ suitably constructed from the learning sample. The function $f_K$ has value 1 on the sets describing the learning objects in the class $K$, and value 0 on the sets that describe other learning objects. That is, $f_K$ is the characteristic function of the class $K$ on the learning sample.

The sought conjunctions are constructed on the basis of $n$ families of predicates.

Suppose a finite family of predicates, $n$, is specified for the $j$th feature, $j \in \{1, 2, \ldots, n\}$. The predicates in $\mathfrak{P}^j$ are functions of value 0 or 1, which are defined on the set of admissible values of the $j$th feature. If, e.g., the set of admissible values of the $j$th feature is the set $\{0, 1, \ldots, t-1\}$, then as $\mathfrak{P}^j$ one may use the set $\{x^0, \ldots, x^{t-1}\}$, where $x^\sigma = 1$ if and only if $x = \sigma$. By the same token, one may consider that $x^\sigma = 1$ if and only if $x \geq \sigma$.

Let $E_t^n$ be the set of all $t$-ary sets of length $n$, $A_K$ be the set of sets in $E_t^n$ on which the function $f_K$ has value 1, and $B_K$ be the set of sets in $E_t^n$ on which $f_K$ has value 0.

We introduce below several definitions. Some are modifications of the concepts known from the theory of disjunctive normal forms of Boolean functions [13] for the more general case at hand. Others (such as the definition of the almost admissible conjunction of the function $f_K$ and its associates) have no counterparts in the theory of disjunctive normal forms of Boolean functions. They were first introduced in [25] in order to obtain the constructions described in what follows.

By *an elementary conjunction (e.c.) over the variables* $x_1, \ldots, x_n$ is meant a formula of the form

$$P_1 \,\&\, \ldots \,\&\, P_r,$$

where $P_i \in \mathfrak{P}^{j_i}$, $j_i \in \{1, 2, \ldots, n\}$ for $i = 1, 2, \ldots, r$ and $j_{i_1} \neq j_{i_2}$ for $i_1 \neq i_2$, $i_1, i_2 = 1, 2, \ldots, r$. The symbol $P_i$ is a *conjunctive multiplier* and $r$ is the *rank*. We omit henceforth the conjunction symbol & for brevity.

An e.c. specifies a function that is defined on sets out of $E_t^n$ and takes on value 0 or 1. This function is equal to 1 if and only if both conjuncts are equal to 1.

By *an elementary disjunction (e.d.) over the variables* $x_1, \ldots, x_n$ is meant a formula of the form

$$P_1 \vee \ldots \vee P_r,$$

where $P_i \in \mathfrak{P}^{j_i}$, $j_i \in \{1, 2, \ldots, n\}$ for $i = 1, 2 \ldots, r$, and $j_{i_1} \neq j_{i_2}$, for $i_1 \neq i_2$, $i_1, i_2 = 1, 2, \ldots, r$. The symbol $P_i$ is called a *disjunctive multiplier*.

The e.d. specifies a function which is defined on sets in $E_t^n$ and takes on value 0 or 1. This function is equal to 0 if and only if both disjuncts are equal to 0.

The set of sets in $E_t^n$ on which the e.c. $\mathfrak{B}$ is equal to 1 is called *an interval* and is denoted by $N_{\mathfrak{B}}$. The cardinalities of the sets $N_{\mathfrak{B}} \cap A_K$ and $N_{\mathfrak{B}} \cap B_K$ are denoted by $q_1^K(\mathfrak{B})$ and $q_2^K(\mathfrak{B})$, respectively.

The e.c. $\mathfrak{B}$ over the variables $x_1, \ldots, x_n$ is called *admissible* for $f_K$ if $q_1^K(\mathfrak{B}) \neq 0$ and $q_2^K(\mathfrak{B}) = 0$.

The e.c. $\mathfrak{B}$ over the variables $x_1, \ldots, x_n$ is called *maximal* for $f_K$ if $\mathfrak{B}$ is an admissible conjunction and there exists no admissible e.c. $\mathfrak{B}'$ such that $N_{\mathfrak{B}} \subset N_{\mathfrak{B}'}$.

The e.c. $\mathfrak{B}$ over the variables $x_1, \ldots, x_n$ is called *almost admissible* for $f_K$ if $q_1^K(\mathfrak{B}) \neq 0$.

The e.c. $\mathfrak{B}$ over the variables $x_1, \ldots, x_n$ is called *almost maximal* for $f_K$ if $\mathfrak{B}$ is almost admissible and there exists no e.c. $\mathfrak{B}'$ such that $N_{\mathfrak{B}} \subset N_{\mathfrak{B}'}$ and $q_2^K(\mathfrak{B}') = q_2^K(\mathfrak{B})$.

A formula of the form

$$\mathfrak{B}_1 \vee \ldots \vee \mathfrak{B}_s,$$

where $\mathfrak{B}_i$, $i \in \{1, 2, \ldots, s\}$ is an e.c., is called a *disjunctive normal form (cnf)*. The cnf realizes the function which is defined on sets in $E_t^n$ and takes on value 0 or 1. This function is equal to 0 if and only if each of its constituent conjunctions is equal to 0.

A formula of the form

$$\mathfrak{C}_1 \,\&\, \ldots \,\&\, \mathfrak{C}_s,$$

where $\mathfrak{C}_i$, $i \in \{1, 2, \ldots, s\}$ is an e.d., is called a *conjunctive normal form (cnf)*. The cnf realizes a function which is defined on sets out of $E_t^n$ and takes on value 0 or 1. This function is equal to 1 if and only if each of its constituent disjunctions is equal to 1. The cnf is called *perfect* if each of its e.d.s has $n$ disjuncts.

The cnf composed of all maximal conjunctions of the function $f_K$ is called its *abridged cnf*.

Let $\mathfrak{B}$ be an almost admissible conjunction of the function $f_K$ of the form

$$P_1 \,\&\, \ldots \,\&\, P_r,$$

where $P_i \in \mathfrak{P}^{j_i}$, $j_i \in \{1, 2, \ldots, n\}$ for $i = 1, 2, \ldots, r$, and let $S \in N_{\mathfrak{B}} \cap A_K$, $S = (a_1, \ldots, a_n)$. The subset $(a_{j_1}, \ldots, a_{j_r})$ out of the set $S$ is called an *elementary classifier generated by the conjunction* $\mathfrak{B}$ *for the class* $K$.

In the general case, an elementary classifier is a fragment of the description of a learning object having specified properties. These properties are defined by the predicates $P_1, \ldots, P_r$. The elementary classifier generated by an admissible conjunction and by the description of a learning object in the class $K$ allows this object to be distinguished from any learning objects that do not belong to the class $K$. Thus, to the admissible conjunctions of the function $f_K$ there correspond certain fragments from the description of learning objects that contain discriminating information.

Every recognition algorithm $A$ in a given class is defined by a subset $P_A$ of the set of elementary classifiers. The most crucial step in the learning process is to construct the set $P_A$. Then a voting procedure is applied for every element $p$ in $P_A$. In the simplest modification, it is assumed that the elementary classifier $p$ votes in favor of assigning the object being recognized to the class $K$ if its description belongs to the truth domain of the conjunction of the function $f_K$ that has generated $p$.

Once informative fragments of descriptions of learning objects are constructed, we can assess the parameters that tell us how informative certain features and their combinations are and how representative each of the learning objects is.

As a rule, we use a particular part of the set $\mathfrak{M}_K$ of all admissible conjunctions of the function $f_K$ in generating the set of elementary classifiers $P_A$. For instance, we can take all maximal conjunctions or impose a restriction on the rank of a conjunction (it is presumed that the rank of the conjunction should not exceed a specified number). Other constraints may be imposed as well. Instead of $\mathfrak{M}_K$, we may consider the set $\mathfrak{M}_K^*$ that consists of all almost admissible conjunctions of the function $f_K$. We are interested, of course, in almost admissible conjunctions $\mathfrak{B}$ in $\mathfrak{M}_K^*$ such that $q_1^K(\mathfrak{B})$ is significantly greater than $q_2^K(\mathfrak{B})$.

Typical examples of the algorithms in the above class are:

(1) test models (a test algorithm was first described in [14]);

(2) algorithms of the Kora type or models using representative descriptors (the Kora algorithm is described in [11] and the first model using representative descriptors was proposed in [6]).

The learning material is ordinarily presented in tabular form (a learning table), in which every column corresponds to a particular feature and every row is a set of feature values describing one of the objects. The rows of the table are partitioned into disjoint classes such that every class contains only the rows that describe objects in the same class.

Let $T$ be a learning table in which the rows are partitioned into $s$ classes $K_1, \ldots, K_s$, and $\mathfrak{M}^A \subseteq \mathfrak{M}_{K_1}^* \cup \mathfrak{M}_{K_2}^* \cup \ldots \cup \mathfrak{M}_{K_s}^*$ is the set of conjunctions, which corresponds to the recognition algorithm $A$. The first to be explored was the case, where $T$ is a binary table, $f_{K_i}$ is a Boolean function, $i = 1, 2, \ldots, s$, and the conjunctions from $\mathfrak{M}^A$ take the form $x_{j_1}^{\sigma_1}, \ldots, x_{j_r}^{\sigma_r}$, where the symbol $x^\sigma$ for $x$ and $\sigma$ in $\{0, 1\}$ specifies a predicate equal to 1 if and only if $x = \sigma$.

Each conjunction $\mathfrak{B} = x_{j_1}^{\sigma_1}, \ldots, x_{j_r}^{\sigma_r}$ in $\mathfrak{M}_K^*$, $K \in \{K_1, \ldots, K_s\}$, defines two following subtables $T_{\mathfrak{B}}^K$ and $\bar{T}_{\mathfrak{B}}^K$ of the table $T$. The subtable $T_{\mathfrak{B}}^K$, is formed by columns numbered $j_1, \ldots, j_r$ and by the rows that describe objects in $K$. It is easy to see that this subtable contains at least one row of the form $(\sigma_1, \ldots, \sigma_r)$. The subtable $T_{\mathfrak{B}}^K$, is formed by columns numbered $j_1, \ldots, j_r$ and by the rows that do not describe objects in $K$. If $q_2^K(\mathfrak{B}) = 0$, that is if $\mathfrak{B}$ is an admissible conjunction, then, by virtue of the foregoing, the subtable $\bar{T}_{\mathfrak{B}}^K$, does not contain a single row of the form $(\sigma_1, \ldots, \sigma_r)$. This condition is not satisfied for $q_2^K(\mathfrak{B}) \neq 0$. In the former case, the set $(\sigma_1, \ldots, \sigma_r)$ is called *representative for the class K with respect to the reference set* $(j_1, \ldots, j_r)$ and in the latter, *almost representative*. Every representative descriptor is assumed to be also an almost representative one.

The set $(\sigma_1, \ldots, \sigma_r)$, which is representative (almost representative) for $K$ with respect to the reference set $(j_1, \ldots, j_r)$, is called a *irredundant* one, if the e.c. $x_{j_1}^{\sigma_1}, \ldots, x_{j_r}^{\sigma_r}$ is maximal (almost maximal) for $f_K$.

To a irredundant representative descriptor there corresponds an incompressible informative fragment from the description of a learning object, i.e., a fragment which, on being compressed, loses its ability to distinguish a given object from an objects in other classes.

We call the set of the columns numbered $j_1, \ldots, j_r$ in the table $T$ a *test* if for every function $f_{K_i}$, $i \in \{1, 2, \ldots, s\}$, each almost admissible conjunction of the form $x_{j_1}^{\sigma_1}, \ldots, x_{j_r}^{\sigma_r}$ is admissible. In terms of its content, the test is a feature set that selects in the description of every learning object a representative descriptor and contains therefore a sufficient amount of information for the learning material to be partitioned into classes.

The test is a *irredundant* one if no one of its own subsets is a test.

The concepts of the (irredundant) test and of the (irredundant) representative descriptor can be readily generalized to the case, where $T$ is an integer-valued table.

In test recognition algorithms, the set $P_A$ is most often generated by irredundant tests, and algorithms of the Kora type use irredundant representative descriptors or irredundant almost representative descriptors as elements of the set $P_A$. All information about the set of elementary classifiers for these models is embedded in a special Boolean matrix, $L_T$, called the *comparison matrix of the table T*. In the case at hand, i.e., where $T$ is a binary table and the conjunctions that generate elementary classifiers are defined in a standard manner, this matrix is formed by *mod 2* pairwise addition of the table $T$ rows that belong to different classes.

In [7], a class of recognition algorithms is constructed, which contains both types of the above algorithms as its elements. The set of elementary classifiers is defined for an algorithm $A$ in this class by selecting a group of objects, $Q(A)$, in the learning sample. In this case, a set of features is taken to be informative if it allows every object $G$ in the group $Q(A)$ to be distinguished from objects that do not belong to the same class as does the object $G$.

Lastly, note that the construction of various families of elementary classifiers is used in estimation models [29, 31].

## 4. THE CONSTRUCTION OF ELEMENTARY CLASSIFIERS. IRREDUNDANT COVERINGS FOR BOOLEAN AND INTEGER-VALUED MATRICES

The recognition algorithm models described above have a certain advantage over statistical methods because they permit recognition even with a sample of learning objects smaller than the number of features. The high computational costs at the learning step (construction of elementary classifiers), however, make them difficult to implement. For instance, in searching for irredundant tests, we must construct a set of all irredundant coverings for the comparison matrix $L_T$ of the learning table $T$ and in constructing irredundant representative descriptors, a similar problem has to be solved for several submatrices of the matrix $L_T$.

Let $L$ be an arbitrary Boolean matrix. We call the set $H$ of columns in the matrix $L$ the covering if the intersection of every row of the matrix $L$ with at least one column in $H$ yields 1 (that is, the submatrix of the matrix $L$ formed by columns in $H$ should not contain rows of the form $(0, \ldots, 0)$). The covering is called irreducible (irredundant) if none of its own subsets is a covering. It is easy to see that the irredundant property implies that a submatrix of the matrix $L$ formed by columns in $H$ should contain each of the rows $(1, 0, 0, \ldots, 0, 0)$, $(0, 1, 0, \ldots, 0, 0)$, and $(0, 0, \ldots, 0, 1)$, that is the set of columns $H$ should contain a submatrix in every row and every column of which one element is exactly 1. This is called a *unit* matrix.

The foundation for asymptotically optimal methods of discrete information analysis in recognition problems was laid as early as the 1970s as part of an inquiry into the complexity of implementing test procedures.

The early procedures for constructing irredundant tests actually reduced to decoding a monotone function and were found before long to be rather ineffectual [33, 54]. Significantly better results were achieved with stochastic test algorithms [34] where the construction of a set of all irredundant tests was replaced by the construction of a fairly representative random sample from that set. But, similar to their deterministic predecessors, these algorithms were tailored to problems where the number of features used to describe learning objects was relatively small compared to that of the objects themselves. Even a slight increase in the number of features would sharply increase the learning time. We often see in practice, however, that the dimension of feature space is significantly greater than the number of learning objects.

In addition to procedures whereby all irredundant tests could be found, attention was given at the same time to estimation of the number of irredundant tests and to the length of a irredundant test, i.e., to what are called metric (quantitative) properties of the set of irredundant tests.

Asymptotic estimates for the number of irredundant tests and the length of a irredundant test for the case where $T$ is a binary table were first reported in [42, 53]. The key results were obtained for the case of practical importance where the number of rows in $T$ was small compared to that of features, $n$. The above estimation, a technically challenging problem in itself, was later taken up also in [1, 2, 16, 18]. The final results were reported in [2]. That was how technical foundations were actually laid on which asymptotic estimates could be obtained for the typical numbers of irreducible coverings and for the typical length of an irreducible covering.

Later [15, 16, 18] proposed a different approach to developing asymptotically optimal recognition procedures of the test type. The approach was based on an analysis of the complexity of constructing a set of all irreducible coverings for the comparison matrix $L_T$ of the learning table $T$. Let us have a closer look at it, taking as an example the construction of irreducible coverings for an arbitrary $u \times n$ Boolean matrix $L$. A case of practical interest is where the matrix $L$ has no identical columns. It is, therefore, natural to suppose that $u \geq \log_2 n$.

For the case $\log_2 n \leq u \leq n^{1-\varepsilon}$ $(\varepsilon > 0)$, an algorithm was developed, which could keep the exhaustive search to a minimum in a sense in a typical situation involving the construction of a set of all irreducible coverings for the matrix $L$.

The exhaustive search is reduced as follows. The original problem is replaced by the logically easier problem of constructing all sets of columns from the matrix $L$ that satisfy only the "irredundant" condition as stated in the above definition of the irreducible covering for a Boolean matrix. Actually, all unit submatrices of the matrix $L$ are to be constructed. In the case at hand, the number of unit submatrices of the matrix $L$

nearly always (for almost all $u \times n$ matrices $L$) proved to be the same asymptotically as the number of irreducible coverings with $n \longrightarrow \infty$. Hence, the set of columns containing a unit submatrix is nearly always an irreducible covering.

The method was tested in computer experiments. Its software described in detail in [15] develops unit submatrices, maximal in a certain sense, and checks every such submatrix to see if the corresponding column set is a covering. The algorithm saved the computer memory resources quite well. It solved quite efficiently the problems of the deterministic and stochastic construction of irredundant tests and of developing the appropriate recognition procedures. Realizations of the procedures were incorporated in the PARK package of application programs [17], the DISARO dialog pattern analysis and recognition system [20], and the OBRAZ system [18]. All of these packages were developed by the Computer Center of the Russian Academy of Sciences.

In [1], one more algorithm was proposed, asymptotically best in finding irredundant tests in the cases where $\log_2 h = o(\log_2^2 n)$, with $n$ tending to infinity ($h$ is the number of rows in the matrix $L_T$). The algorithm was developed and substantiated using the same principles as in [15, 16, 18].

Later on, basic studies in the field were spearheaded at matters related to the development of efficient realizations for procedures of the Kora type (models with representative descriptors). The language used was that of disjunctive normal forms. The researchers were concerned at first with the case of binary learning information. The construction of the set of all irredundant representative descriptors for the class $K$ was formulated as the problem of constructing an abridged cnf for the partial Boolean function $f_K$ (a cnf which consists of all maximal conjunctions of the function $f_K$). It was a problem of constructing an abridged cnf for a completely defined Boolean function in which the set of zeros is the same as the set of zeros of the function $f_K$, or transforming a perfect cnf into an abridged cnf [12, 29, 42).

A modification of the algorithm from [15, 16, 18] was used in [19, 21, 22] to produce an asymptotically optimal solution to the problem of constructing an abridged cnf $D_{\mathfrak{K}}$ for the Boolean function $F_{\mathfrak{K}}$ of $n$ variables defined by the cnf $\mathfrak{K}$. Consideration was given there to the cases where the cnf $\mathfrak{K}$ was perfect, did not contain negations of variables, or, lastly, was not necessarily perfect and could contain negations of variables. In each of the above cases, it was assumed that the number of elementary disjunctions in the original cnf was not greater than $n^{1-\varepsilon}$ ($\varepsilon > 0$). The complexity of a solution was estimated by the number of & operations that had to be performed in transforming normal forms. It was shown that nearly always (for almost all cnfs of the type in question) and with $n \longrightarrow \infty$, this count was not asymptotically greater than the length of the sought abridged cnf. As a matter of record, the problem was solved by constructing an auxiliary cnf that consisted of all almost maximal conjunctions of the function $F_{\mathfrak{K}}$. Subject to the same constraints, asymptotic estimates were developed for the typical length of the cnf $D_{\mathfrak{K}}$ and for the typical rank of its conjunctions.

Asymptotic estimates for the typical length of the abridged cnf for a completely defined Boolean function of $n$ variables defined by enumerating zero points, with $n \longrightarrow \infty$, were reported in [3].

It was shown in [19, 21] that a set of all irredundant representative descriptors of the learning table $T$ with $m$ rows could be built efficiently (with an asymptotically minimal complexity) by constructing irreducible coverings for the series of Boolean matrices $L^{(1)}, \ldots, L^{(m)}$ (the submatrices of the comparison matrix $L_T$ of the table $T$). The matrix $L^{(i)}$, $i \in \{1, 2, \ldots, m\}$, is formed as a result of mod 2 summation of the $i$th row from the table $T$ with all rows not included in the same class as the row in question. The irreducible coverings were constructed by the algorithm from [15, 16, 18]. In [19, 21], a similar approach (involving the construction of what are known as weighted irreducible coverings for the matrices $L^{(1)}, \ldots, L^{(m)}$ was applied to constructing almost representative descriptors. With this approach, both binary and integer-valued information can be processed. The efficiency of the method was verified on a computer. The experiments were conducted for the case where the number of learning objects was not greater than that of features. Appropriate recognition algorithms were developed, and their computer realizations were likewise incorporated in the DISARO and OBRAZ systems mentioned above.

In [21, 24], the asymptotic estimates for the typical numbers of irredundant representative descriptors and for the typical length of a irredundant representative descriptor for the class $K$ of the (first binary and then integer) learning table were derived subject to the condition that the number of learning objects not belonging to the class $K$ should not be greater than $n^{1-\varepsilon}$ ($\varepsilon > 0$). Asymptotic estimates were also derived in [24] for the same typical numerical characteristics of the set of representative descriptors applicable to a fairly broad class of binary tables. These characteristics included typical values of, first, the length of the abridged cnf for a partial two-valued function and the rank of its constituent conjunctions and, second, the number of admissible conjunctions and the rank of the admissible conjunction of a partial Boolean function.

Note that the first asymptotic estimates for the typical length of the minimal cnf that realizes a partial Boolean function were derived in [27].

Asymptotically optimal methods for the construction of irreducible coverings were employed in developing new and better models of recognition procedures which used irredundant representative descriptors and irredundant almost representative descriptors as elementary classifiers [21, 23, 26]. Such models can be useful in processing arbitrary numerical information.

## 5. IRREDUNDANT COVERINGS FOR INTEGERVALUED MATRICES. STATISTICAL PROPERTIES OF IRREDUNDANT COVERINGS

The concept of a irredundant (irreducible) covering for a Boolean matrix may be generalized to the case of an integer-valued matrix.

Consider $\sigma \in E_k^r$, $k \geq 2$, and $\sigma = (\sigma_1, \ldots, \sigma_r)$.

We introduce the following notation:

$R(\sigma)$ is the set of all sets $(\beta_1, \ldots, \beta_r)$ in $E_k^r$ such that $\beta_j \neq \sigma_j$ for $j \in \{1, 2, \ldots, r\}$.

$Q_p(\sigma)$, $p \in \{1, 2, \ldots, r\}$ is the set of all sets $(\beta_1, \ldots, \beta_r)$ in $E_k^r$ such that $\beta_p \neq \sigma_p$ and $\beta_j = \sigma_j$ for $j \in \{1, 2, \ldots, r\} \backslash \{p\}$.

Suppose $E \subseteq E_k^r$ and $E = \{\sigma^{(1)}, \ldots, \sigma^{(q)}\}$. We set

$$Q_p(E) = \left(\bigcup_{j=1}^{q} Q_p(\sigma^{(j)})\right) \backslash E.$$

Let $L = (a_{ij})$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, be a matrix with elements from $\{0, 1, \ldots, k-1\}$, $k \geq 2$, $H$ be a set of $r$ different columns from the matrix $L$, and $L^H$ be a submatrix of the matrix $L$ formed by the columns of the set $H$.

The set of rows in the submatrix $L^H$ can be treated as a subset $E^H$ of sets in $E_k^r$. The column set $H$ is called the *$E$ covering* if $E^H \cap E = \varnothing$. The column set $H$ which is an $E$ covering is called *the irredundant $E$ covering* if for any $p$ from $\{1, 2, \ldots, r\}$ we can indicate a pair of sets $(\sigma', \sigma'')$ such that $\sigma' \in E^H$, $\sigma'' \in E$, and the sets $\sigma'$ and $\sigma''$ differ in the $p$th place but are the same in all other places.

Thus, the set $H$ of columns from the matrix $L$ is a irredundant $E$ covering if the submatrix $L^H$ of the matrix $L$ formed by the columns of the set $H$ has the following two properties:

(1) $L^H$ does not contain a single row from $E$;

(2) if $p \in \{1, 2, \ldots, r\}$, then $L^H$ does contain at least one row from the set $Q_p(E)$.

Note that if $L$ is a Boolean matrix and $\sigma \in E_2^r$, then the set $H$ of columns from the matrix $L$ is a irredundant $\{\sigma_1, \ldots, \sigma_r\}$ covering if and only if the following two conditions are satisfied:

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

(2) $L^H$ contains each of the rows

$$(\bar{\sigma}_1, \sigma_2, \sigma_3, \ldots, \sigma_{r-1}, \sigma_r),$$
$$(\sigma_1, \bar{\sigma}_2, \sigma_3, \ldots, \sigma_{r-1}, \sigma_r),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$(\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{r-1}, \bar{\sigma}_r).$$

In particular, $H$ is a irredundant $\{0, \ldots, 0\}$ covering (or, as is customary to say, an irreducible covering) of the matrix $L$ if the following conditions are satisfied:

(1) $L^H$ does not contain the row $(0, \ldots, 0)$;

(2) $L^H$ contains each of the rows

$$(1, 0, 0, \ldots, 0, 0),$$
$$(0, 1, 0, \ldots, 0, 0),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$(0, 0, 0, \ldots, 0, 1).$$

Thus, the concept of a irredundant $E$ covering for $E = \{\sigma\}$ introduced above is a generalization of the well-known concept of the irreducible covering for a Boolean matrix.

If, on the other hand, $L$ is a Boolean matrix, $\sigma \in E_2^r$, and $E = R(\sigma)$, then the set $H$ of columns from the matrix $L$ is a irredundant $E$ covering if and only if the following two conditions are satisfied:

(1) $L^H$ does not contain the row $(\bar{\sigma}_1, \ldots, \bar{\sigma}_r)$;

(2) $L^H$ contains each of the rows

$$(\sigma_1, \bar{\sigma}_2, \bar{\sigma}_3, \ldots, \bar{\sigma}_{r-1}, \bar{\sigma}_r),$$
$$(\bar{\sigma}_1, \sigma_2, \bar{\sigma}_3, \ldots, \bar{\sigma}_{r-1}, \bar{\sigma}_r),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$(\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \ldots, \bar{\sigma}_{r-1}, \sigma_r).$$

Consequently, if $\sigma = (1, \ldots, 1)$, the concept of an irreducible covering is the same as that of a irredundant $R(\sigma)$ covering for a Boolean matrix.

From a practical point of view, two cases are as important as the general case. They are $E = \{\sigma\}$, $\sigma \in E_k^r$, and $E = R(\sigma)$, $\sigma \in E_k^r$. Irredundant coverings of the general type and, notably, irredundant $\{(0, \ldots, 0)\}$ coverings are encountered in studies of the metric properties of a set of irredundant tests and in constructing irredundant test search algorithms. The case of $E = \{\sigma\}$, $\sigma \in E_k^r$ is needed in order to obtain similar results for a set of irredundant representative descriptors. The latter case is related to the transformation of normal forms of logical functions.

Let $B(L, E)$ denote the set of all pairs of the form $(H, E)$, where $H$ is a irredundant $E$ covering for the matrix $L$ (in the case when $E = \{\sigma\}$, we use the symbolism $B_\sigma(L)$).

Let the set $H$ consist of the columns numbered $j_1, \ldots, j_r$, $j_1 < \ldots < j_r$, of the matrix $L$.

The set of the elements $\{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$ of the matrix $L$ is called *consistent in $E$* if the $i_p$-th row of the submatrix $L^H$ belongs to $Q_p(E)$ for $p = 1, 2, \ldots, r$. Let $S(L, E)$ denote the totality of all pairs of the form $(Q, E)$, where $Q$ is the set of elements from the matrix $L$, consistent in $E$. (Where $E = \{\sigma\}$, $\sigma \in E_k^r$, we also use the symbolism $S_\sigma(L)$.

It is easy to see that the set $H$ of the columns numbered $j_1, \ldots, j_r$ of the matrix $L$ is a irredundant covering if and only if we can indicate in $L$ a set of elements of the type $\{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$ consistent in $E$, and the submatrix $L^H$ does not contain any of the rows from $E$.

We introduce the following notation:

$|A|$ is the cardinality of the set $A$;

$a_n \sim b_n$, $n \longrightarrow \infty$ implies that $\lim\limits_{n \to \infty} a_n/b_n = 1$;

$\mathfrak{M}_{mn}^k$ is the totality of all $m \times n$ matrices with elements from $\{0, 1, \ldots, k-1\}$, $k \geq 2$;

$E_1^r$ is a set consisting of one set of the form $(1, \ldots, 1)$ whose length is $r$;

$\Phi_d(m)$ is the interval

$$\left(\frac{1}{2}\log_d mn - \frac{1}{2}\log_d\log_d mn - \log_d\log_d\log_d n,\right.$$

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

In what follows, we give asymptotic estimates for the typical numbers of irredundant $E$ coverings and the typical length of a irredundant $E$ covering for matrices of $\mathfrak{M}_{mn}^k$ in the following two cases: (a) $E = \{\sigma\}$, $\sigma \in E_k^r$ and (b) $E = R(\sigma)$, $\sigma \in E_t^r$, $k - 1 \leq t \leq k$. Compute in each case the same numerical estimates for the corresponding set of consistent sets. Typical situations are identified on the basis of propositions like "For almost all matrices $L$ in $\mathfrak{M}_{mn}^k$, with $n \longrightarrow \infty$, the property $\mathfrak{P}$ is satisfied," and this property may also be of a limiting character. It means that the proportion of the matrices in $\mathfrak{M}_{mn}^k$ for which the property $\mathfrak{P}$ is satisfied up to $\varepsilon$ tends to 1 and, at the same time, $\varepsilon$ tends to 0 with $n$ tending to infinity. For example, if two functionals, $F(L)$ and $G(L)$, are defined on the matrices in $\mathfrak{M}_{mn}^k$, we say that for almost all matrices $L$ in $\mathfrak{M}_{mn}^k$, with $n$ tending to infinity, it is true that $F \sim G$ (that is, $F$ is asymptotically equal to $G$) if there exist two positive infinitely diminishing functions, $\alpha(n)$ and $\beta(n)$, such that for sufficiently large $n$ we have $1 - |\mathfrak{M}|/|\mathfrak{M}_{mn}^k| \leq \alpha(n)$, where $\mathfrak{M}$ is the set of matrices $L$ in $\mathfrak{M}_{mn}^k$, for which $1 - \beta(n) \leq F(L)/G(L) \leq 1 + \beta(n)$.

For $L \in \mathfrak{M}_{mn}^k$, $k - 1 \leq t \leq k$, we set

$$B_1(L) = \bigcup_{r=1}^{n} \bigcup_{\sigma \in E_k^r} B_\sigma(L),$$

$$B_2^t(L) = \bigcup_{r=1}^{n} \bigcup_{\sigma \in E_t^r} B(L, R(\sigma)),$$

$$S_1(L) = \bigcup_{r=1}^{n} \bigcup_{\sigma \in E_k^r} S_\sigma(L),$$

$$S_2^t(L) = \bigcup_{r=1}^{n} \bigcup_{\sigma \in E_t^r} S(L, R(\sigma)).$$

There is the following

**Theorem 5.1.** [25]. (1) *If* $m^\alpha \leq n \leq k^m$, $\alpha > 1$, $k \geq 2$, *then for almost all matrices* $L$ *in* $\mathfrak{M}_{mn}^k$, *with* $n \longrightarrow \infty$ *tending to infinity, it is true that*

$$|B_1(L)| \sim |S_1(L)| \sim \sum_{r \in \Phi_k(m)} C_n^r C_m^r r!(k-1)^r k^{r-r^2}$$

*and for almost all pairs in* $B_1(L)$ *the lengths of irredundant coverings fall in the interval* $\Phi_k(m)$.

(2) *If* $m^\alpha \leq n \leq (k/(k-1))^m$, $\alpha > 1$, $k \geq 2$, $k - 1 \leq t \leq k$, *then for almost all matrices* $L$ *in* $\mathfrak{M}_{mn}^k$, *with* $n \longrightarrow \infty$ *tending to infinity, it is true that*

$$\left|B_2^t(L)\right| \sim \left|S_2^t(L)\right| \sim \sum_{r \in \Phi_d(m)} C_n^r C_m^r r! t^r (k-1)^{r^2 - r} k^{-r^2}$$

*and for almost all pairs in* $B_2^t(L)$ *the lengths of irredundant coverings fall in the interval* $\Phi_d(m)$, *where* $d = k/(k-1)$.

irredundant $\{\sigma\}$ coverings and $R(\sigma)$ coverings can be constructed by a modified algorithm from [15, 16, 18]. It is developed on the same principles as used in the asymptotically optimal construction of irreducible coverings. The problem reduces to developing the sets $S_1(L)$ and $S_2^t(L)$, respectively.

Moreover, in constructing irredundant $\{\sigma\}$ coverings, the counterpart of a unit submatrix is the submatrix of the matrix $L$ which, except the permutation of rows, takes the form

$$\begin{pmatrix} \beta_1 & \sigma_2 & \sigma_3 & \ldots & \sigma_{r-1} & \sigma_r \\ \sigma_1 & \beta_2 & \sigma_3 & \ldots & \sigma_{r-1} & \sigma_r \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \sigma_1 & \sigma_2 & \sigma_3 & \ldots & \sigma_{r-1} & \beta_r \end{pmatrix},$$

where $\beta_p \neq \sigma_p$ for $p = 1, 2, \ldots, r$.

In constructing irredundant $\{R(\sigma)\}$ coverings, the counterpart of a unit submatrix is the submatrix of the matrix $L$ which, except the permutation of rows, takes the form

$$\begin{pmatrix} \sigma_1 & \beta_{12} & \beta_{13} & \dots & \beta_{1r-1} & \beta_{1r} \\ \beta_{21} & \sigma_2 & \beta_{23} & \dots & \beta_{2r-1} & \beta_{2r} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \beta_{r1} & \beta_{r2} & \beta_{r3} & \dots & \beta_{rr-1} & \sigma_r \end{pmatrix},$$

where $\beta_{ij} \neq \sigma_j$ for $i \neq j$, $i, j = 1, 2, \dots, r$.

## 6. IRREDUNDANT COVERINGS FOR INTEGER-VALUED MATRICES IN THE CONSTRUCTION OF ELEMENTARY CLASSIFIERS. METRIC PROPERTIES OF ELEMENTARY CLASSIFIERS WHEN THE NUMBER OF FEATURES IS GREATER THAN THE NUMBER OF PRECEDENTS

We denote by $W_r^n$, $r \leq n$, the set of all sets of the form $(j_1, \dots, j_r)$, where $j_u \in \{1, 2, \dots, n\}$, for $u = 1, 2, \dots, r$ and $j_1 < \dots < j_r$.

Let $\sigma \in E_k^r$, $\sigma = (\sigma_1, \dots, \sigma_r)$, $w \in W_r^n$, $w = (j_1, \dots, j_r)$ and $T$ be a learning table with elements from $\{0, 1, \dots, k-1\}$, $k \geq 2$, the rows of which are partitioned into the classes $K_1, \dots, K_s$.

We further use $T_K$ and $\bar{T}_K$ to denote the subtables of the table $T$, which are formed, respectively, by rows from the class $K$, $K \in \{K_1, \dots, K_s\}$ of the table $T$, and by rows not contained in the class $K$.

The set $(\alpha_1, \dots, \alpha_n)$ of $E_k^n$ is called *a w extension of the set* $\sigma$ if $\alpha_{j_t} = \sigma_t$ for $t = 1, 2, \dots, r$.

It is easy to see that *the set* $\sigma$ *is a (irredundant) representative descriptor for the class K of the table T with respect to w* if the set of the columns numbered $j_1, \dots, j_r$ of the table $\bar{T}_K$ table is a (irredundant) $\{\sigma\}$ covering for $\bar{T}_K$ and at least one row contained in $T_K$ is a $w$ extension of the set $\sigma$.

Thus, the construction of irredundant representative descriptors for the class $K$ reduces to constructing irredundant $\{\sigma\}$ coverings for that part of the table $T$ which does not contain descriptions of objects in the class $K$.

The asymptotic estimates obtained in Theorem 5.1 for irredundant $\{\sigma\}$ coverings are used in studying the metric properties of a set of irredundant representative descriptors.

Simple additional constructions lead to asymptotic estimates of typical values for the number of irredundant representative descriptors and the length of a irredundant representative descriptor for the class $K$ of the learning table $T$.

Let $m_1$ be the number of rows in $T$ that describe objects in the class $K$, $m_2$ be the number of rows in $T$ that describe objects in the remaining classes, and $P_K(T)$ be the set of irredundant representative descriptors for the class $K$.

The table $T$ may be treated as a pair of matrices of the type $(L_1, L_2)$, where $L_1 \in \mathfrak{M}_{m_1 n}^k$, $L_2 \in \mathfrak{M}_{m_2 n}^k$. Let $\mathfrak{N}_{m_1 m_2 n}^k$ be the set of all pairs of matrices of the said type.

It was proved as true

**Theorem 6.1.** [24]. *If* $m_2^{\alpha} \leq n \leq k^{m_2}$, $\alpha > 1$, $k \geq 2$, *then for almost all tables T in* $\mathfrak{N}_{m_1 m_2 n}^k$, *with* $n \longrightarrow \infty$ *tending to infinity, it is true that*

$$|P_K(T)| \sim \sum_{r \in \Phi_k(m_2)} (1 - (1 - k^{-r})^{m_1})$$
$$\times C_n^r C_{m_2}^r r! (k-1)^r k^{r - r^2}$$

and the lengths of almost all sets of $P_K(T)$ fall in the interval $\Phi_k(m_2)$.

We denote by $P_K^0(T, r)$ the set of all representative descriptors for the class $K$ of the table $T$ whose lengths are not greater than $r$.

Let $r_0 = ]\log_2 m_2 n[$ (where $]x[$ is the least integer which is not smaller than $x$).

Suppose that $\log_2 m_2 \leq n / \log_2^{\alpha} n$, $\alpha > 1$, $m_1 \leq m_2$. Then there holds

**Theorem 6.2.** [24]. *If* $r_0 \leq r \leq n$, *then for almost all tables T in* $\mathfrak{N}_{m_1 m_2 n}^2$, *with* $n \longrightarrow \infty$ *tending to infinity, it is true that*

$$|P_K^0(T, r)| \sim \sum_{r = r_0}^{r} C_n^r 2^r (1 - (1 - 2^{-r}))^{m_1}$$

*and the lengths of almost all sets of* $P_K^0(T, r)$ *fall in the interval* $[r_0, r]$.

We now define a test and a irredundant test using the concepts of an $E$ covering and of a irredundant $E$ covering, respectively.

Let $H$ be a set of $r$ different columns numbered $j_1, \dots, j_r$ of the table $T$, and $E_K^H$ be a set composed of different rows of the subtable of the table $T$ formed by the columns numbered $j_1, \dots, j_r$.

It is easy to see that the column set $H$ is a test for $T$ if and only if for each class $K$ of the table $T$ it is true that the set of columns from the table $T_K$ is an $E_K^H$ covering.

The test composed of the columns numbered $j_1, \ldots, j_r$ is a irredundant one if for some class $K$, the set of columns from the table $T_K$ is a irredundant $E_K^H$ covering.

When test algorithms are used in practice, it is usual to apply the constructions given below. With them, the search for irredundant tests reduces to a search for irreducible coverings of a Boolean matrix.

Let us call as the sum of rows $S_{i_1}$ and $S_{i_2}$ of the learning table $T$ an $n$-ary Boolean vector in which the $j$th place is 0 if it is the same in both $S_{i_1}$ and $S_{i_2}$, and 1 otherwise. Let $L_T$ be a Boolean matrix, each row of which is formed by adding together two rows of $T$ belonging to different classes. Obviously, the set of columns numbered $j_1, \ldots, j_r$ is a irredundant test if and only if the set of columns numbered $j_1, \ldots, j_r$ from the matrix $L_T$ is an irreducible covering.

It is convenient to assess the metric properties of a set of irredundant tests $\mathfrak{T}(T)$ for an integer-valued table $T$ using the statistical properties of irreducible coverings or irredundant $(0, \ldots, 0)$ coverings for the matrix $L_T$. Let $T^k_{m_1m_2n}$ be the set of all $K$-valued tables in which the rows are partitioned into two classes containing $m_1$ and $m_2$ rows, respectively. In [18], the proof is given for

**Theorem 6.3.** *If* $(m_1m_2)^\alpha \le n \le k^{(m_1m_2)^\beta}$, $\alpha > 1$ and $\beta < 1/2$, *then for almost all tables T in* $T^k_{m_1m_2n}$ *it is true that*

$$|\mathfrak{T}(T)| \sim \sum_{r \in \Phi_k(m_1m_2)} C_n^r C_{m_1m_2}^r r!(k-1)^r k^{-r^2},$$

$$n \longrightarrow \infty,$$

*and the lengths of almost all irredundant tests in* $\mathfrak{T}(T)$ *fall in the interval* $\Phi_k(m_1m_2)$.

## 7. IRREDUNDANT COVERINGS IN THE SIMPLIFICATION OF NORMAL FORMS OF SPECIAL K-VALUED LOGICAL FUNCTIONS. THE COMPLEXITY OF CONSTRUCTING AN ABRIDGED CNF

The same technical basis (an inquiry into the statistical properties of irredundant coverings) was used to develop asymptotically optimal construction methods and to study the metric properties of abridged cnfs for completely defined and partial two-valued logical functions defined on $K$-ary $n$-dimensional sets, $k \ge 2$ [19, 21, 22, 25].

Let $\mathfrak{K}$ be a conjunctive normal form (cnf) of the type

$$D_1 \,\&\, \ldots \,\&\, D_m, \tag{7.1}$$

where $D_i$, $i = 1, 2, \ldots, m$, is an e.d. over the variables $x_1, \ldots, x_n$, $D_i = x_{i1}^{\sigma_{i1}} \vee \ldots \vee x_{ip_i}^{\sigma_{ip_i}}$.

Recall that the cnf $\mathfrak{K}$ is *perfect* if $p_i = n$ for $i = 1, 2, \ldots, m$.

We set

$$x^\sigma = \begin{cases} 0, & \text{if } x = \sigma \\ 1, & \text{otherwise,} \end{cases} \tag{7.2}$$

$x, \sigma \in \{0, 1, \ldots, k-1\}$, $k \ge 2$.

The cnf $\mathfrak{K}$ specifies the function $F_{\mathfrak{K}}(x_1, \ldots, x_n)$, which is defined on the set $E_k^n$ of $K$-ary $n$-dimensional sets and can take on value 0 or 1. The set of sets in $E_k^n$, on which the function $F_{\mathfrak{K}}$ takes on value 1 is denoted by $N_{F_{\mathfrak{K}}}$.

The e.c. $\mathfrak{B}$ is called *maximal* for $F_{\mathfrak{K}}$ if $N_{\mathfrak{B}} \subseteq N_{F_{\mathfrak{K}}}$ and there exists no interval $N_{\mathfrak{B}'}$ such that $N_{\mathfrak{B}} \subset N_{\mathfrak{B}'} \subseteq N_{F_{\mathfrak{K}}}$. The cnf composed of all maximum conjunctions of the function $F_{\mathfrak{K}}$ is called *abridged*.

The e.c. $\mathfrak{B}$ of rank $r$ is called *almost maximal* for $F_{\mathfrak{K}}$ if either of the following two conditions is satisfied: (1) $r = 1$ and a number $i$ can be indicated in $\{1, 2, \ldots, m\}$ such that the disjunction $D_i$ contains $\mathfrak{B}$ as a disjunct; (2) $r > 1$ and $r$ numbers $i_1, i_2, \ldots, i_r$ can be indicated in $\{1, 2, \ldots, m\}$ such that each of the disjunctions $D_{i_1}, \ldots, D_{i_r}$ contains exactly one conjunct from $\mathfrak{B}$ and, if $p \ne q$, $p, q \in \{1, 2, \ldots, r\}$, then the disjunctions $D_{i_p}$ and $D_{i_q}$ contain different conjuncts from $\mathfrak{B}$.

It is easy to show that the e.c. $\mathfrak{B}$ is maximal for $F_{\mathfrak{K}}$ if and only if the following two conditions are satisfied:

(1) $\mathfrak{B}$ is almost maximal for $F_{\mathfrak{K}}$;

(2) each disjunction $D_i$, $i = 1, 2, \ldots, m$, contains at least one conjunct from $\mathfrak{B}$.

Consider the construction of an abridged cnf $D_{\mathfrak{K}}$ for the function $F_{\mathfrak{K}}$ when for every symbol of the type $x^\sigma$ included into $\mathfrak{K}$, the following condition is satisfied: $\sigma \in \mathfrak{S}$, $\mathfrak{S} \subseteq \{0, 1, \ldots, k-1\}$.

Subject to the constraints $m^\alpha \le n \le d^m$, where $\alpha > 1$ and $d = |\mathfrak{S}|/(|\mathfrak{S}| - 1)$, when $\mathfrak{K}$ is a perfect cnf and $d = (|\mathfrak{S}| + 1)/|\mathfrak{S}|$ otherwise, the following results were obtained.

It was shown that this problem reduced to constructing a set of all irredundant $R(\sigma)$ coverings for an integer-valued matrix. With the results obtained for irredundant coverings, it is possible, when transforming $\mathfrak{K}$ to $D_{\mathfrak{K}}$, to keep the "exhaustive search" to a minimum in a certain sense.

The exhaustive search is kept to a minimum in the following manner. The original problem is replaced by a logically simpler one whose objective is to construct a cnf $D^1_{\mathfrak{K}}$, composed of all almost maximum conjunctions of the function $F_{\mathfrak{R}}$. The cnf $D^1_{\mathfrak{R}}$ is fairly close in its properties to $D_{\mathfrak{R}}$. This cnf contains all maximal conjunctions of the function $F_{\mathfrak{R}}$, and its length is nearly always (for almost all cnfs of the kind in question) asymptotically the same as that of the cnf $D^1_{\mathfrak{R}}$. Once $D^1_{\mathfrak{R}}$ is constructed, it is possible nearly always to "almost solve" the original problem. The complexity of the solution is assessed in terms of the count of & operations required to construct $D^1_{\mathfrak{R}}$. It is shown that, with $n \longrightarrow \infty$, this count is nearly always not greater than the length of the cnf $D_{\mathfrak{R}}$ (Assertion (2) of Theorem 5.1 is used to prove this). So, the proposed approach to constructing an abridged cnf of the function $F_{\mathfrak{R}}$ offers a way to solve the problem with an asymptotically minimal complexity in a certain sense.

When $k = 2$, $\mathfrak{S} = \{1\}$, and $\mathfrak{R}$ is not a perfect cnf, we actually construct a set of all irreducible coverings for a Boolean matrix. In the general case, if $\mathfrak{R}$ is a perfect cnf, one constructs a set of irredundant coverings of $B^t_2(L)$, where $L$ is a matrix with elements from $\{0, 1, \ldots, t-1\}$, $t = |\mathfrak{S}|$. If $\{0, 1, \ldots, t\}$ is not a perfect cnf, we solve a similar problem for a matrix with elements from $\{0, 1, \ldots, T\}$.

The estimate derived in Theorem 5.1 for $|B^t_2(L)|$ directly leads to asymptotic estimates of typical values for the length $l(\mathfrak{R})$ of the cnf $D_{\mathfrak{R}}$ and the rank of its constituent conjunctions, and also to similar estimates for $D^1_{\mathfrak{R}}$ whose length is denoted by $l_1(\mathfrak{R})$ below.

Let $M^1_{mn}$ be a set of all perfect cnfs of the above type, and $M^2_{mn}$ be a set of all cnfs of the same type. There holds

**Theorem 7.1.** [25]. (1) *If $m^\alpha \le n \le d^m$, $\alpha > 1$ and $d = t/(t-1)$, then for almost all cnfs $\mathfrak{R}$ in $M^1_{mn}$, with $n \longrightarrow \infty$ tending to infinity, it is true that*

$$l(\mathfrak{R}) \sim l_1(\mathfrak{R}) \sim \left|S^t_2(L_{\mathfrak{R}})\right| \sim \sum_{r \in \Phi_d(m)} C^r_n C^r_m r! d^{r-r^2},$$

*and the ranks of almost all conjunctions in $D_{\mathfrak{R}}$ fall in the interval $\Phi_d(m)$.*

(2) *If $m^\alpha \le n \le d^m$, $\alpha > 1$, $d = (t+1)/t$, then for almost all cnfs $\mathfrak{R}$ in $M^2_{mn}$, with $n \longrightarrow \infty$ tending to infinity, it is true that*

$$l(\mathfrak{R}) \sim l_1(\mathfrak{R}) \sim \left|S^t_2(L_{\mathfrak{R}})\right| \sim \sum_{r \in \Phi_d(m)} C^r_n C^r_m r! d^{-r^2},$$

*and the ranks of almost all conjunctions in $D_{\mathfrak{R}}$ fall in the interval $\Phi_d(m)$.*

Of special interest is the case, where $k = 2$.

Let $N_{mn}$ be a set of all cnfs of the form defined in (7.1), which implement Boolean functions. We introduce the notation:

$N^1_{mn}$ is a subset in $N_{mn}$ composed of cnfs that do not contain negations of variables;

$N^2_{mn}$ is a subset in $N_{mn}$, composed of all perfect cnfs;

$N^3_{mn}$ is a subset in $N_{mn}$, composed of arbitrary cnfs.

By virtue of Theorem 7.1, we have assertions formulated in the three theorems below [19, 21, 22].

**Theorem 7.2.** *If $m^\alpha \le n \le 2^m$, $\alpha > 1$, then for almost all cnfs $\mathfrak{R}$ in $N^1_{mn}$, with $n \longrightarrow \infty$ tending to infinity, it is true that*

$$l(\mathfrak{R}) \sim l_1(\mathfrak{R}) \sim \left|S^1_2(L_{\mathfrak{R}})\right| \sim \sum_{r \in \Phi_2(m)} C^r_n C^r_m r! 2^{-r^2}$$

*and the ranks of almost all conjunctions in $D_{\mathfrak{R}}$ fall in the interval $\Phi_2(m)$.*

**Theorem 7.3.** *If $m^\alpha \le n \le 2^m$, $\alpha > 1$, then for almost all cnfs $\mathfrak{R}$ in $N^2_{mn}$, with $n \longrightarrow \infty$ tending to infinity, it is true that*

$$l(\mathfrak{R}) \sim l_1(\mathfrak{R}) \sim \left|S^2_2(L)\right| \sim \sum_{r \in \Phi_2(m)} C^r_n C^r_m r! 2^{r-r^2},$$

*and the ranks of almost all conjunctions in $D_{\mathfrak{R}}$ fall in the interval $\Phi_2(m)$.*

**Theorem 7.4.** *If $m^\alpha \le n \le (3/2)^m$, $\alpha > 1$, then for almost all cnfs $\mathfrak{R}$ in $N^2_{mn}$, with $n \longrightarrow \infty$ tending to infinity, it is true that*

$$l(\mathfrak{R}) \sim l_1(\mathfrak{R}) \sim \left|S^2_2(L)\right| \sim \sum_{r \in \Phi_{3/2}(m)} C^r_n C^r_m r! (2/3)^{r^2},$$

*and the ranks of almost all conjunctions in $D_{\mathfrak{R}}$ fall in the interval $\Phi_{3/2}(m)$.*

When $\mathfrak{S} = \{0, 1, \ldots, k-1\}$ and $\mathfrak{R}$ is a perfect cnf, the asymptotic estimate of $l(\mathfrak{R})$ yields an asymptotic estimate for the length of an abridged cnf for a two-valued function which is defined on sets in $E^n_k$ and takes on value 0 exactly on $m$ sets.

Note that the problem formulated in this section is also applicable to the case where $x^\sigma$ is defined differently, e.g.,

$$x^\sigma = \begin{cases} 1, & \text{if } x = \sigma, \\ 0, & \text{otherwise,} \end{cases} \tag{7.3}$$

$x, \sigma \in \{1, 2, \ldots, k-1\}$.

Then, by similar reasoning, it is possible to obtain a result fully analogous to the one formulated in Theorem 7.1.

Consideration is given in [25] to the construction of partial two-valued functions defined on $K$-ary $n$-dimensional sets. It is shown that this problem can be reduced to constructing both irredundant $R(\sigma)$ coverings and irredundant $\{\sigma\}$ coverings.

Let $f(x_1, \ldots, x_n)$ be a two-valued function which is defined on some subset of sets in $E_k^n$. Also let $A = \{\alpha_1, \ldots, \alpha_v\}$ be the set of sets in $E_k^n$ on which the function $f$ takes on value 1, and $B = \{\beta_1, \ldots, \beta_u\}$ be the set of sets in $E_k^n$ on which the function $f$ takes on value 0.

Let $L_1$ be a matrix whose rows are the sets $\alpha_1, \ldots, \alpha_v$, and $L_2$, be a matrix whose rows are the sets $\beta_1, \ldots, \beta_u$.

The following criterion of maximality in respect of the conjunction of a partial function $f$ applies:

**Theorem 7.5.** *The e.c. of the form* $x_{j_1}^{\sigma_1}, \ldots, x_{j_r}^{\sigma_r}$, *where* $x_{j_t}^{\sigma_t}$ *is defined for* $t = 1, 2, \ldots, r$ *by the rule (7.2) and* $j_1 < \ldots < j_r$, *is maximal for f if and only if the set of columns numbered* $j_1, \ldots, j_r$ *in the matrix* $L_2$ *is a irredundant* $R(\sigma)$ *covering,* $\sigma = \{\sigma_1, \ldots, \sigma_r\}$, *and the submatrix of the matrix* $L_1$ *composed of columns numbered* $j_1, \ldots, j_r$ *contains at least one row of the form* $(\gamma_1, \ldots, \gamma_r)$, *where* $\gamma_t \neq \sigma_t$ for $t = 1, 2, \ldots, r$.

The maximality criterion for the conjunction of the partial function $f$ is also provided by

**Theorem 7.6.** *The e.c. of the form* $x_{j_1}^{\sigma_1}, \ldots, x_{j_r}^{\sigma_r}$, *where* $x_{j_t}^{\sigma_t}$ *is defined for* $t = 1, 2, \ldots, r$ *by the rule (7.3) and* $j_1 < \ldots < j_r$, *is maximal for f if and only if the set of columns numbered* $j_1, \ldots, j_r$ *in the matrix* $L_2$ *is a irredundant* $\{\sigma\}$ *covering,* $\sigma = (\sigma_1, \ldots, \sigma_r)$, *and the submatrix of* $L_1$, *which is composed of the columns numbered* $j_1, \ldots, j_r$ *contains at least one row of the form* $(\sigma_1, \ldots, \sigma_r)$.

The problem of constructing the abridged cnf of the partial function $f$ can be reduced to the problem of constructing all irreducible coverings for a number of Boolean matrices. To demonstrate, we associate the set $\alpha_i$, $i \in \{1, 2, \ldots, v\}$ with the Boolean matrix $L^{(i)}$ where the intersection of the $p$th row and the $q$th column yield value 1 if the sets $\alpha_i$ and $\beta_p$ differ in the $q$th coordinate, and value 0 otherwise.

There holds

**Theorem 7.7.** *The conjunction* $x_{j_1}^{\sigma_1}, \ldots, x_{j_r}^{\sigma_r}$, *where* $x_{j_t}^{\sigma_t}$ *is defined for* $t = 1, 2, \ldots, r$ *by the rule* (7.3) *and* $j_1 < \ldots < j_r$, *is maximal for f if and only if we can indicate* $i \in \{1, 2, \ldots, m\}$ *such that the set of the columns numbered* $j_1, \ldots, j_r$ *in the matrix* $L^{(i)}$ *is a irredundant* $\{(0, \ldots, 0)\}$ *covering, and for* $t \in \{1, 2, \ldots, r\}$, *the* $j_t$ *coordinate of the set* $\alpha_t$ *is equal* to $\sigma_t$.

## CONCLUSION

The objective of this paper was to provide an overview of the results achieved in the development of asymptotically optimal methods for the construction of irredundant coverings for Boolean and integer matrices, and in the use of these methods in the solution and analysis of recognition problems. The results are as follows:

(1) A new class of algorithms was proposed whereby a search can be performed for irreducible (irredundant) coverings of a Boolean matrix. These algorithms are based on the construction of unit submatrices of the original matrix, and they make a better use of its internal structure as compared with previous methods.

(2) The concept of an asymptotically optimal algorithm is introduced to construct sets of all irreducible coverings for Boolean matrices by discarding some conditions that define the coverings. It was shown that the algorithms of the class indicated in (1) above are asymptotically optimal in a case of practical importance, namely, where the number of rows of a Boolean matrix is smaller than that of its columns.

(3) In substantiating the asymptotic optimality of the algorithms, asymptotic estimates of typical values were derived for the number of irreducible coverings and the length of an irreducible covering for a $u \times n$ Boolean matrix subject to the condition that $\log_2 n \leq u \leq n^{1-\varepsilon}$ ($\varepsilon > 0$).

(4) The concept of a irredundant covering for an integer-valued matrix was introduced as a generalization of the concept of an irreducible covering for a Boolean matrix. Results that were fully analogous to those mentioned in (1) through (3) above for irreducible coverings of a Boolean matrix were obtained for the principal practically important types of irredundant coverings for an integer-valued matrix.

(5) A class of models of logical recognition algorithms was described, including most of the algorithms used in practice. In these models, the key element at the learning stage is the construction of elementary classifiers. Ordinarily, this reduces to a search for irredundant coverings of special matrices (construction of maximal conjunctions of two-valued functions). With an asymptotically optimal search for irredundant coverings serving as the basis, improvements were made in known models of logical recognition algorithms and qualitatively new models were built.

(6) The technique thus developed was drawn upon to obtain asymptotically optimal solutions to problems involving the construction of elementary classifiers for the principal models of logical recognition algorithms.

The same technique was used to derive typical asymptotic estimates for the key quantitative characteristics of a set of elementary classifiers for this type of models, such as the number of representative descriptors and the length of a representative descriptor, and similar characteristics for a set of irredundant representative descriptors and a set of irredundant tests.

(7) Asymptotically optimal methods were proposed whereby abridged cnfs can be constructed for two-valued (partial and completely defined) functions. With these, maximal conjunctions are synthesized on constructing and analyzing a set of all almost maximal conjunctions.

(8) The metric (quantitative) properties of abridged cnfs for two-valued functions were investigated. In particular, asymptotic estimates of typical values were obtained for the number of (almost) maximal conjunctions and for the rank of an (almost) maximal conjunction of a function of $n$ variables on condition that its zeros are not greater in number than $n^{1-\varepsilon}$ $(\varepsilon > 0)$.

(9) The results thus obtained were verified by experiment. The verification confirmed the efficiency of the proposed algorithms for the construction of irredundant coverings in practical and model problems.

## REFERENCES

1. Andreev, A.E., Some Topics of Test Pattern Recognition, *Dokl. Akad. Nauk SSSR*, 1981, vol. 255, no. 4, pp. 781–784.
2. Andreev, A.E., On the Asymptotic Behavior of the Number of Irredundant Tests and the Minimal Length of a Test for Almost all Tables, *Problemy Kibernetiki*, Moscow: Nauka, 1984, no. 41, pp. 117–141.
3. Andreev, A.E., On the Minimization of Disjunctive Normal Forms, *Dokl. Akad. Nauk SSSR*, 1984, vol. 274, no. 2, pp. 265–269.
4. Aslanyan, L.A., On a Recognition Method Based on Class Separation by Disjunctive Normal Forms, *Kibernetika*, 1975, no. 5, pp. 103–110.
5. Aslanyan, L.A., Recognition Algorithms with Logical Separators, *Sbornik Rabot po Matematicheskoi Kibernetike* (Set of Papers on Mathematical Cybernetics), Moscow: Computer Center, USSR Academy of Sciences, 1976, no. 1, pp. 116–131.
6. Baskakova, L.V. and Zhuravlev, Yu.I., A Model of Recognition Algorithms with Representative Descriptors and Systems of Reference Sets, *Zh. Vych. Matem. Matem. Fiz.*, 1981, vol. 21, no. 5, pp. 1264–1275.
7. Bushmanov, O.N., *Klass algoritmov raspoznavaniya, osnovannyi na poiske empiricheskikh zakonomernostei* (A Class of Recognition Algorithms Based on a Search for Empirical Relations), Moscow: Computer Center, Russian Academy of Sciences, 1992.
8. Bushmanov, O.N., Djukova, E.V., Zhuravlev, Yu.I., Kochetkov, D.V., and Ryazanov, V.V., A Pattern Analysis and Recognition System, in *Raspoznavanie, Klassifikatsiya, Prognoz (Matematicheskie Metody i Ikh Primenenie)* (Recognition, Classification, Forecasting: Mathematical Methods and Their Application), Moscow: Nauka, 1989, no. 2, pp. 250–273.
9. Vapnik, V.N. and Chervonenkis, A.Ya., *Teoriya raspoznavaniya obrazov* (Theory of Pattern Recognition), Moscow: Nauka, 1974.
10. Vapnik, V.N., *Vosstanovlenie zavisimostei po empiricheskim dannym* (Reconstruction of Dependences from Empirical Data), Moscow: Nauka, 1979.
11. Weinzweig, M.N., KORA—A Pattern Recognition Learning Algorithm, in *Algoritmy obucheniya raspoznavaniya obrazov* (Pattern Recognition Learning Algorithms), Moscow: Sov. Radio, 1973, pp. 82–91.
12. Denisova, R.A., *Metod sinteza tupikovykh predstavitel'nykh naborov dlya K-znachnykh tablits* (A Method to Synthesize Irredundant Representative Descriptors for $K$-ary Tables), Moscow: Computer Center, USSR Academy of Sciences, 1984.
13. *Diskretnaya matematika i matematicheskie voprosy kibernetiki* (Discrete Mathematics and Mathematical Aspects of Cybernetics), Yablonskii, S.V. and Lupanov, O.B., Eds., Moscow: Nauka, 1974.
14. Dmitriev, A.I., Zhuravlev, Yu.I., and Krendelev, F.P., On Mathematical Principles of Object or Phenomenon Classification, in *Diskretnyi analiz* (Discrete Analysis), Novosibirsk: Institute of Mathematics, Siberian Division, USSR Academy of Sciences, 1966, no. 7, pp. 3–17.
15. Djukova, E.V., On an Algorithm for Constructing Irredundant Tests for Binary Tables, *Sbornik rabot po diskretnoi matematike* (Set of Papers on Discrete Mathematics), Moscow: Computer Center, USSR Academy of Sciences, 1976, no. 1, pp. 167–185.
16. Djukova, E.V., On an Asymptotically Optimal Algorithm for Constructing Irredundant Tests, *Dokl. Akad. Nauk SSSR*, 1977, vol. 233, no. 4, pp. 527–530.
17. Djukova, E.V., Zhuravlev, Yu.I., Zenkin, A.A., *et al.*, *Paket prikladnykh programm dlya resheniya zadach raspoznavaniya i klassifikatsii (PARK)* (PARK—A Package of Application Programs for Recognition and Classification Problems), Moscow: Computer Center, USSR Academy of Sciences, 1981.
18. Djukova, E.V., Asymptotically Optimal Test Algorithms for Recognition problems, *Problemy Kibernetiki*, Moscow: Nauka, 1982, no. 39, pp. 165–199.
19. Djukova, E.V., On Complexity of Realization of Some Recognition Procedures, *Zh. Vych. Matem. Matem Fiz.*, 1987, vol. 27, no. 1, pp. 114–127.
20. Djukova, E.V., Zhuravlev, Yu.I., and Ryazanov, V.V., *Dialogovaya sistema analiza i raspoznavaniya obrazov* (An Interactive System of Pattern Analysis and Recognition), Moscow: Computer Center, USSR Academy of Sciences, 1988.
21. Djukova, E.V., KORA-type Recognition Algorithms: Complexity of Realization and Metric Properties, *Raspoznavanie, Klassifikaziya, Prognoz: Matematicheskie Metody i Ikh Primenenie* (Recognition, Classification, Forecasting: Mathematical Methods and Their Application), Moscow: Nauka, 1989, no. 2, pp. 99–125.
22. Djukova, E.V., On Solution of Systems of Boolean Equations of the Quasi-Nelson Type, *Voprosy kibernetiki, diskretnaya matematika, metody i primenenie*

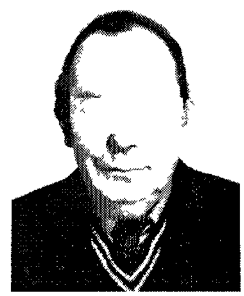
(Topics in Cybernetics. Discrete Mathematics, Methods and Applications), Moscow: USSR Academy of Sciences, Scientific Council for the Complex Program "Cybernetics," 1989, pp. 5–19.

23. Djukova, E.V. and Karneeva, I.L. Learning Algorithm Models Based on Various Methods of Recording Original Information, in *Matematicheskie metody v raspoznavanii obrazov i diskretnoi optimizatsii* (Mathematical Methods in Pattern Recognition and Discrete Optimization), Moscow: Computer Center, USSR Academy of Sciences, 1990, pp. 43–56.
24. Djukova, E.V., Asymptotic Estimates of Some Characteristics of a Set of Representative Descriptors and the Stability Problem, *Zh. Vych. Matem. Matem Fiz.*, 1995, vol. 35, no. 1, pp. 122–134.
25. Djukova, E.V., Asymptotically Optimal Methods of Discrete Information Analysis in Recognition problems, *Doctoral (Phys.&Math.) Dissertation*, Moscow: Computer Center, Russian Academy of Sciences, 1996.
26. Djukova, E.V., Zhuravlev, Yu.I., and Rudakov, K.V., On Algebraic Synthesis of Correct Recognition Procedures on the Basis of Elementary Algorithms, *Zh. Vych. Matem. Matem. Fiz.*, 1996, vol. 36, no. 8, pp. 217–225.
27. Zhuravlev. Yu.I., On a Class of Incompletely Defined Functions of Boolean Algebra, *Diskretnyi Analiz*, Novosibirsk: Institute of Mathematics, Siberian Division, USSR Academy of Sciences, 1964, no. 2, pp. 23–27.
28. Zhuravlev, Yu.I., Correct Algebras over Sets of Noncorrect (Heuristic) Algorithms, *Kibernetika*, 1977, no. 4, pp. 14–21; 1977, no. 6, pp. 21–27; 1978, no. 2, pp. 35–43.
29. Zhuravlev, Yu.I., On an Algebraic Approach to Recognition or Classification Problems, *Problemy Kibernetiki*, Moscow: Nauka, 1978, no. 33, pp. 5–68.
30. Zhuravlev, Yu.I. and Platonenko, I.M., On an Economical Multiplication of Boolean Equations, *Zh. Vych. Matem. Matem. Fiz.*, 1984, vol. 20, no. 5, pp. 1294–1309.
31. Zhuravlev, Yu.I., and Nikiforov, V.V. Estimation-Based Recognition Algorithms, *Kibernetika*, 1971, no. 3, pp. 1–11.
32. Zhuravlev, Yu.I. and Rudakov, K.V., On Algebraic Adjustment of Information Processing (Transformation) Procedures, in *Problemy prikladnoi matematiki i informatiki* (Topics in Applied Mathematics and Information Technology), Moscow: Nauka, 1987, pp. 187–198.
33. Konstantinov, R.M., Koroleva, Z.E., and Kudryavtsev, V.B., On a Combinatorial-Logical Approach to Forecasting of Ore Presence, *Problemy Kibernetiki*, Moscow: Nauka, 1975, no. 30, pp. 5–33.
34. Kuznetsov, V.E., On a Stochastic Algorithm to Compute Information Characteristics of Tables by the Test Method, *Diskretnyi Analiz*, Novosibirsk: Institute of Mathematics, Siberian Division, USSR Academy of Sciences, 1973, no. 23, pp. 8–23.
35. Mazurov, V.D., On a Recognition Learning Method, *Kibernetika*, 1970, no. 2, pp. 92–94.
36. Mazurov, V.D., Committees of Inequality Systems and the Recognition problem, *Kibernetika*, 1971, no. 3, pp. 140–146.
37. Mazurov, V.D., Kazantsev, V.S., Beletskii, N.G., Krivonogov, A.I., and Smirnov, A.I., Topics in Justification and Application of Committee Recognition Algorithms, *Raspoznavanie, klassifikaziya, prognoz* (Recognition, Classification, Forecasting), Moscow: Nauka, 1989, pp. 114–148.
38. Matrosov, V.L., Limited-capacity Correct Algebras over Sets of Noncorrect Algorithms, *Dokl. Akad. Nauk SSSR*, 1980, vol. 253, no. 1, pp. 25–30.
39. Matrosov, V.L., On the Competeness Criteria in the Model of Estimates Evaluation Algorithms and Its Algebraic Closure, *Dokl. Akad. Nauk SSSR*, 1981, vol. 258, no. 4, pp. 791–796.
40. Matrosov, V.L., Synthesis of Optimal Algorithms in Algebraic Closures of Recognition Algorithm Models, *Raspoznavanie, klassifikaziya, prognoz* (Recognition, Classification, Forecasting), Moscow: Nauka, 1989, pp. 149–176.
41. *Metod komitetov v raspoznavanii obrazov* (The Committee Method in Pattern Recognition), Sverdlovsk: Institute of Mathematics, Ural Scientific Center, USSR Academy of Sciences, 1984.
42. Noskov, V.N. and Slepyan, V.A., On the Number of Irredundant Tests for One Class of Tables, *Kibernetika*, 1972, no. 1, pp. 60–65.
43. Platonenko, I.M., *O realizatsii algoritmov tipa "Kora" s pomoshch'yu resheniya sistem bulevykh uravnenii spetsial'nogo vida* (On Implementation of KORA-type Algorithms Through Solving Systems of Special Boolean Equations), Moscow: Computer Center, USSR Academy of Sciences, 1983.
44. Rudakov, K.V., *On Symmetric and Functional Restrictions on Classification Algorithms, Dokl. Akad. Nauk SSSR*, 1987, vol. 297, no. 1, pp. 43–46.
45. Rudakov, K.V., Universal and Local Restrictions in the Heuristic Algorithm Correction Problem, *Kibernetika*, 1987, no. 2, pp. 30–35.
46. Rudakov, K.V., Completeness and Universal Restrictions in the Heuristic Classification Algorithm Correction Problem, *Kibernetika*, 1987, no. 3, pp. 106–109.
47. Rudakov, K.V., Symmetric and Functional Restrictions in the Heuristic Classification Algorithm Correction Problem, *Kibernetika*, 1987, no. 4, pp. 73–77.
48. Rudakov, K.V., On Application of Universal Restrictions in Analysis of Classification Algorithms, *Kibernetika*, 1988, no. 1, pp. 1–5.
49. Rudakov, K.V., On an Algebraic Theory of Universal and Local Restrictions for Classification Problems, *Raspoznavanie, klassifikaziya, prognoz* (Recognition, Classification, Forecasting), Moscow: Nauka, 1989, pp. 176–201.
50. Ryazanov, V.V., Optimization of Algorithms for Estimation by Parameters Characterizing Representativity of Reference Rows, *Zh. Vych. Matem. Matem. Fiz.*, 1976, vol. 16, no. 6, pp. 1559–1570.
51. Ryazanov, V.V., On Construction of Optimal Recognition and Taxonomy (Classification) Algorithms for Applied Problems, *Raspoznovanie, klassifikaziya, prognoz* (Recognition, Classification, Forecasting), Moscow: Nauka, 1989, pp. 229–279.
52. Ryazanov, V.V. and Sen'ko, O.V. On Some Recognition Models and Their Optimization Methods, *Raspoznavanie, klassifikaziya, prognoz: matematicheskie metody i ikh primenenie* (Recognition, Classification, Forecasting: Mathematical Methods and Their Application), Moscow: Nauka, 1990, no. 3, pp. 106–145.

53. Slepyan, V.A., On the Number of Irredundant Tests and Column Informative Content Measure for Almost all Binary Tables, *Dokl. Akad. Nauk SSSR*, 1987, vol. 297, no. 1, pp. 43–46.
54. Slutskaya, T.L. Algorithms to Compute Information Weights of Features, *Diskretnyi Analiz*, Novosibirsk: Institute of Mathematics, Siberian Division, USSR Academy of Sciences, 1966, no. 12, pp. 75–90.
55. Chegis, I.A. and Yablonskii, S.V., Logical Methods of Electrical Circuit Inspection, *Trudy Mat. in-ta im. V.A. Steklova AN SSSR*, 1958, vol. 51, pp. 270–360.
56. McCluskey, E.J., Minimization of Boolean Functions, *Bell Lab. Tech. J.*, 1956, vol. 35, Nov., pp. 1417–1444.


**Elena V. Djukova.** Born 1945. Graduated from Moscow State University in 1967. Received a PhD (Kandidat Nauk) degree in the physical and mathematical sciences in 1979. A senior researcher at the Computer Center, Russian Academy of Sciences. Author of about 40 publications.

Scientific interests: discrete mathematics, mathematical methods of pattern recognition, and image processing.



**Yurii I. Zhuravlev.** Born 1935. Graduated from the Faculty of Mechanics and Mathematics, Moscow State University, in 1957. Received a PhD (Kandidat Nauk) degree in 1959 and a postdoctoral (Doktor Nauk) degree in the physical and mathematical sciences in 1965. Since 1969, with the Computer Center of the Russian Academy of Sciences, Moscow, first as a laboratory head and then, as a deputy director. Professor of the Moscow Institute of Physics and Technology. A full member of the Russian Academy of Sciences (since 1992) and of the Academy of Sciences of Spain (since 1993).

Scientific interests: mathematical logic, algebra, discrete optimization, pattern recognition, and the use of mathematical and computational methods for solving applied problems of data processing and research automation in industry, medicine, geology, and sociology. Author of more than 170 publications on information technology and applied mathematics.

Editor-in-Chief of the journal *Pattern Recognition and Image Analysis* and the yearbook *Raspoznavanie, klassifikaziya, prognoz: matematicheskie metody i ikh primenenie* (Recognition, Classification, and Forecasting: Mathematical Methods and Their Application) and a member of the editorial boards of several international and Russian journals on information technology and applied mathematics.