Chapter 1 [1]

A COMMON LOGIC APPROACH TO DATA MINING AND PATTERN RECOGNITION

Arkadij D. Zakrevskij
United Institute of Informatics Problems
of the National Academy of Sciences of Belarus
Surganova Str. 6, 220012 Minsk, Belarus
E-mail: zakr@newman.bas-net.by

Abstract: In this chapter a common logical approach is suggested to solve both data mining and pattern recognition problems. It is based on using finite spaces of Boolean or multi-valued attributes for modeling of the natural subject areas. Inductive inference used for extracting knowledge from data is combined with deductive inference, which solves other pattern recognition problems. A set of efficient algorithms was developed to solve the regarded problems, dealing with Boolean functions and finite predicates represented by logical vectors and matrices.

An abstract world model for presentation of real subject areas is also introduced. The data are regarded as some information concerning individual objects and are obtained by the experiments. The knowledge, on the contrary, represents information about the qualities of the whole subject area and establishes some relationships between its attributes. The knowledge could be obtained by means of inductive inference from some data presenting information about elements of some reliable selection from the subject area. That inference consists of looking for empty (not containing elements of the selection) intervals of the space, putting forward corresponding hypotheses (suggesting emptiness of the intervals in the whole subject area), evaluating their plausibility and accepting the more plausible ones as *implicative regularities*, represented by elementary conjunctions.

These regularities serve as axioms in the deductive inference system used for solving the main recognition problem, which arises in a situation when an object is contemplated with known values of some attributes and unknown values of some others, including goal attributes.

Key Words: Data Mining, Data and Knowledge, Pattern Recognition, Inductive Inference, Implicative Regularity, Plausibility, Deductive Inference.

---

[1] Triantaphyllou, E. and G. Felici (Eds.), **Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques**, Massive Computing Series, Springer, Heidelberg, Germany, pp. 1-43, 2006.

1. INTRODUCTION

1.1 Using Decision Functions

There exist a great variety of approaches to data representation and data mining aimed at knowledge discovery [Frawley, Piatetsky-Shapiro, *et al.*, 1991], and only some of them are mentioned below. The most popular base for them is perhaps using the Boolean space M of binary attributes constituting some set $X = \{x_1, x_2, ..., x_n\}$.

When solving pattern recognition problems, the initial data are frequently represented by a set of points in the space M presenting positive and negative examples [Bongard, 1970], [Hunt, 1975], [Triantaphyllou, 1994]. Every point is regarded as a Boolean vector with components corresponding to the attributes and taking values from the set {0, 1}. The problem is considered as finding rules for recognizing other points, i.e. deciding which of them are positive and which are negative (in other words, guessing the binary value of one more attribute, called a goal attribute). To solve that problem, some methods were suggested that construct a Boolean function f separating the two given sets of points. This function is used as a *decision function* dividing the Boolean space into two classes, and so uniquely deciding for every element to which class does it belong. This function can be considered as the knowledge extracted from the two given sets of points.

It was suggested in some early works [Hunt, 1975], [Pospelov, 1990] to use threshold functions of attributes as classifiers. Unfortunately, only a small part of Boolean functions can be presented in such a form. That is why disjunctive normal forms (DNF) were used in subsequent papers to present arbitrary Boolean decision functions [Bongard, 1970], [Zakrevskij, 1988], [Triantaphyllou, 1994]. It was supposed that the simpler function f is (the shorter DNF it has), the better classifier it is.

For example, let the following Boolean matrix A show by its rows the positive examples, and matrix B – the negative ones (supposing that $X = (a, b, c, d)$).

$$\begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array}, \qquad \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \boldsymbol{B} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

No threshold function separating these two sets exists in that case. However, a corresponding decision function f can be easily found in DNF by the visual minimization method based on using the Karnaugh maps: rectangular tables with 2^n squares which represent different elements of the space M and are ordered by the Gray code [Karnaugh, 1953], [Zakrevskij, 1960]. This order is indicated by the lines on the top and left sides of the table. They show columns and rows where corresponding variables take value 1. Some of the table elements are marked with 1 or 0 – the known values of the represented Boolean function. For example, four top elements in Figure 1 (scanned from left to right) correspond to inputs (combinations of values of the arguments *a*, *b*, *c*, *d*) 0000, 0010, 0011, and 0001. Two of them are marked: the second with 0 (negative example) and the fourth with 1 (positive example).

It is rather evident from observing the table that all its elements which represent positive examples (marked with 1) are covered by two intervals of the space M over X, that do not contain zeros (negative examples).

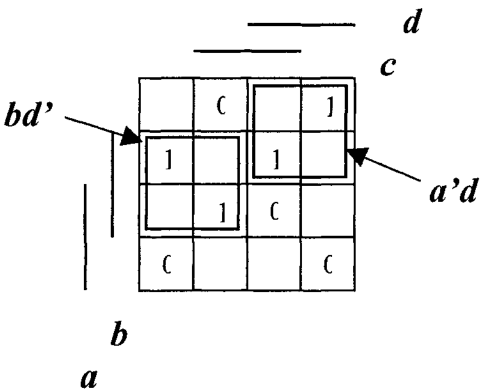


Figure 1. Using a Karnaugh Map to Find a Decision Boolean Function.

The characteristic functions of those intervals are bd' (*b* and not *d*) and $a'd$ (not *a* and *d*), hence the sought-for decision function could be

$$f = bd' \vee a'd.$$

In general, to find a decision Boolean function with minimum number of products is a well-known hard combinatorial problem of incompletely specified Boolean functions minimization. Nevertheless, several practically efficient methods were developed for its solution, exact and approximate ones [Zakrevskij, 1965], [Zakrevskij, 1988], some of them oriented towards large databases [Triantaphyllou, 1994].

It is worthwhile to note a weak point of recognition techniques aimed at binary decision functions. They produce too much categorical classification, when sometimes the available information is not sufficient for that and it would be more appropriate to answer: "I do not know". Generally speaking, for these techniques there appear to be some troubles connected with plausibility evaluation of the results of recognition. Because of it, new approaches have been developed, overcoming this drawback.

A special approach was suggested in [Piatetsky-Shapiro, 1991], [Agrawal, Imielinski, *et al.*, 1993], [Matheus, Chan, *et al.*, 1993], [Klösgen, 1995] for very big databases. The whole initial data are presented by one set of so called transactions (some subsets A_i from the set of all attributes A), and association rules are searched defined as condition statements "if V, then w", where $V \subset A$ and usually $w \in A$. They are regarded valid if only the number of transactions for which $V \cup \{w\} \subseteq A_i$ (called the *support*) is big enough, as well as the percentage of transactions where $V \cup \{w\} \subseteq A_i$ holds taken in the set of transactions where relation $V \subseteq A_i$ is satisfied (called the *confidence level*). The boundaries on the admissible values of these characteristics could be defined by users.

One more approach is suggested below. It is based on introducing a special symmetrical form of knowledge (called implicative regularities) extracted from the data. That form enables us to apply powerful methods of deductive inference, which was developed before for mechanical theorem proving [Chang and Lee, 1973], [Thayse, Gribomont, *et al.*, 1988] and now is used for solving pattern recognition problems.

1.2 Characteristic Features of the New Approach

The following main properties of the suggested common approach to data mining and pattern recognition should be mentioned next.

First, the concepts of data and knowledge are more strictly defined [Zakrevskij, 1988], [Zakrevskij, 2001]. The data are considered as some information about separate objects, while the knowledge is information about the subject area as a whole. According to this approach, we shall believe that the data present information about the existence of some objects with definite combinations of properties (attribute values), whereas the knowledge presents information about existing regular relationships between attributes, and these relationships are expressed by prohibiting some combinations of properties.

Second, no attributes are regarded *a priori* as goal ones. All attributes are included into the common set $X = \{x_1, x_2, ..., x_n\}$ and have equal rights there. Hence, the data are presented by only one set of selected points from the Boolean space over X, and there is no need to represent them, for instance, by two sets of examples (positive and negative).

Third, the knowledge consists of some known regularities. The key question is to choose a proper model for them. Starting from general assumptions the following statements are accepted. Any regarded regularity defines a logical connection between some attributes. This means that some combinations of attribute values are declared impossible (prohibited). In the simplest case such a regularity can be expressed by the logical equation $k_i = 0$ or by the equation $d_i = 1$, where k_i is a conjunct formed of some attributes (in direct or inverse mode) from the set X, d_i is a disjunct, and $d_i = \neg k_i$. For instance, the equations $ab'c = 0$ and $a' \vee b \vee c' = 1$ represent the same regularity, which prohibits the following combination: $a = 1, b = 0, c = 1$. A regularity of this kind is called *implicative regularity* (more general than functional one) [Zakrevskij, 1982, 1987]. It prohibits a set of attribute value combinations forming an interval in the Boolean space M over X – the characteristic set of the conjunct k_i. As it is shown below, regularities of the considered type could be rather easily extracted from the data, and it is not difficult to evaluate their strength and plausibility, which is very important for their further application.

Fourth, no decision functions for some separate attributes are found and used. Instead, the recognition problem is solved individually in every concrete situation, when an object with known values of some attributes is observed and the value of some other attribute (regarded as goal one in that situation) should be found. Different result types are foreseen by that [Zakrevskij, 1982]:

success – the value is found, since the knowledge is sufficient for that,

failure – the knowledge is not sufficient, and that is why the unique value of the attribute cannot be defined,

inconsistency – the object contradicts the knowledge, so either it does not belong to the subject area or the knowledge is not reliable.

Fifth, some results of the theory of Boolean functions were generalized for finite predicates [Zakrevskij, 1990]. That enabled us to extend the methods developed for dealing with data and knowledge in the case of binary attributes onto the case of multi-valued ones.

To characterize the suggested approach, it is necessary to define also the concept of the *world model*.

Not declining far from the tradition, we shall use an abstract artificial world into which many natural subject areas can be mapped without any essential detriment. Suppose this world is a set W of some objects. Objects can differ in values of attributes, where attributes compose the set $X = \{x_1, x_2, \ldots, x_n\}$. Each one of the attributes x_i is characterized by the corresponding finite set V_i of its alternative values, and the Cartesian product of these sets $V_1 \times V_2 \times ... \times V_n$ constitutes the space of multi-valued attributes M. Elements from W are identified with some elements of M and may be considered as abstract models of real objects of a natural subject area.

Hence, the world is represented by a relation $W \subseteq M$ or by a corresponding finite predicate $\varphi\ (x_1, x_2, ..., x_n)$ taking value 1 on the elements of the set W. In case of two-valued attributes this predicate degenerates into a Boolean function $f(x_1, x_2, ..., x_n)$. The world is trivial when $W = M$, and in this case the problems discussed below have no sense. However, it turns out in the majority of practical interpretations that the number of different world objects is essentially less than the number of all elements in the space M: $|W| << |M|$.

This chapter is organized as follows. The basic notions of data and knowledge are discussed in the second section, and some modes of data and knowledge representation by logical matrices are proposed, both for cases of Boolean and multi-valued attributes. The problem of extracting knowledge from data is regarded in the third section, where implicative regularities are introduced to present the knowledge, and the rules for estimating their plausibility are suggested. In the fourth section, a method for testing the knowledge matrices for consistency is proposed. Also, some algorithms of knowledge matrices equivalence transformations are described, leading to their useful simplification. The fifth section is devoted to the concluding stage of the pattern recognition process, of which the aim is to calculate the values of the goal attributes of an observed object. A method of deductive inference is suggested, which uses the previously found knowledge and the partial information about the object. Special attention is paid to the deductive inference in finite predicates. The last section contains a brief enumeration of some practical applications of the suggested approach.

2. DATA AND KNOWLEDGE

2.1 General Definitions

Any research in the pattern recognition problem is inevitably connected with data and knowledge processing. The question about decomposing information into data and knowledge appears when developing systems of artificial intelligence defined usually as knowledge based systems.

Both data and knowledge are basic concepts, and that is why it is difficult to define them strictly formally. A number of definitions were suggested reflecting different aspects of these concepts but touching rather forms of representation of data and knowledge and the rules of their using than their essence. For instance, knowledge was regarded as "a form of computer representation of information" [Pospelov, 1990], and was defined as "information which is necessary for a program for its intellectual behavior" [Waterman, 1986]. Such attempts suggest the idea of impossibility of differentiation between data and knowledge, strict and universal at the same time, for any situation. According to a more universal definition the knowledge is regarded as some useful information derived from the data [Frawler, Piatetsky-Shapiro, *et al.*, 1991].

In this chapter, a working definition is proposed, intended for use in logical inference. Proceeding from the general suppositions, it is natural to define the data as any information about individual objects, and the knowledge about the world W as a whole. According to this assumption, we shall consider the *data* presenting information about the existence of some objects with definite combinations of properties (P), and consider the *knowledge* presenting information about the existence of regular relationships between attributes, prohibiting some other combinations of properties (Q) by equations $k_i = 0$, where k_i is a conjunction over the set of attributes X. In other words, the knowledge is regarded as the information about the non-existence of objects with some definite (now prohibited) combinations of attribute values.

Reflecting availability of the mentioned combinations by the predicates P and Q, one can present the data by the affirmations

$$\exists w \in W\colon\ P(w),$$

with the existential quantifier $\exists$ (there exists) and the knowledge – by affirmations

$$\neg\,\exists w \in W\colon\ Q(w),$$

with its negation $\neg\,\exists$ (there does not exist). The latter ones could be easily transformed into affirmations

$$\forall w \in W\colon\ \neg\, Q(w),$$

with the generality quantifier $\forall$ (for every).

When natural subject areas are investigated, the data present initial information obtained by discovering some objects and revealing their qualities via attribute value measurements. The result of data processing, their generalization, could be regarded as the knowledge. The classical example of the data are the Tycho Brahe tables of how the planets of our solar system move across the sky, whereas the Kepler laws, induced from them, can serve as an excellent example of the knowledge.

Suppose that the data present a complete description of some objects where for each attribute its value for a considered object is shown. Usually not all the objects from some world W could be described but only a relatively small part of them which forms a random selection F: $|F| << |W|$. Selection F can be represented by a set of selected points in the space M.

The distribution of these points reflects the regularities inherent in the world: every prohibition generates some empty, i.e. free of selected points, region in the space M. The reverse affirmation suggests itself: empty regions correspond to some regularities. But such an affirmation is a hypothesis which could be accepted if only it is plausible enough. The matter is that an empty region can appear even if there are no regularities, for instance when $W = M$ (everything is possible) and elements of the set F are scattered in the space M quite at random obeying the law of uniform distribution of probabilities. Evidently, the probability of such an event depends only on the character and the size of the empty region as well as on parameters of the space M and the cardinality of the set F. It is pertinent to remember the Ramsey theorem which asserts that in each structure consisting of n elements some regular substructures having m elements are arising if m is considerably less than n. That means that for any m a corresponding n can be found for which this assertion holds [Boolos and Jeffrey, 1989], [Graham and Spencer, 1990].

The data are obtained usually experimentally, while the knowledge is obtained from an expert, or by means of inductive inference from some data presenting information about elements of some reliable selection from the subject area. The knowledge asserts the non-existence of objects with some definite combinations of properties, declaring bans on them. The inductive inference consists in looking for empty (not containing elements of the selection) intervals of the space M, putting forward corresponding hypotheses (suggesting emptiness of the intervals in the whole subject area), evaluating plausibility of these hypotheses and accepting the more plausible of them as implicative regularities.

A set of regularities forms the contents of a knowledge base and can be subjected to equivalence transformations in order to increase the efficiency of using it when solving various problems of recognition.

The main recognition problem relates to the situation when an object is contemplated with known values of some attributes and unknown values of some others, including goal attributes. The possible values of the latter ones are to be calculated on the base of the knowledge. Sufficiently high plausibility of the forecasting should be guaranteed by that. This problem could be solved by means of deductive inference of the theorem proving type. The problem of minimizing inference chains could be solved additionally, that arises when explanatory modules of expert systems are worked out.

2.2 Data and Knowledge Representation – the Case of Boolean Attributes

Using Boolean (two-valued) attributes to describe an object, we suppose that the value 1 means that the object has the corresponding property and the value 0 means that it has not.

Consider first the problem of data representation.

If the information about some object is complete (in as much as our world model allows), it can be represented by a point in the Boolean space or by a corresponding Boolean vector indicating with 1s in some positions which properties belong to the object. For example, if $X = \{x_1, x_2, x_3, x_4\}$, the vector 1001 means that the described object possesses the properties x_1 and x_4, but not x_2 or x_3.

When the information about the object is incomplete, a ternary vector could be used for its representation. For example, the vector 10-1 means that it is not known if the object has the property x_3.

A selection of elements from W can be presented by a Boolean matrix $\boldsymbol{K}$ (or a ternary one, in case of incomplete information about the regarded elements). Let us call it a *conjunctive matrix*, meaning that its rows are interpreted as products. It could be regarded as a *data matrix* and looks as follows:

$$\begin{array}{cc} & \begin{array}{cccccccc} a & b & c & d & e & f & g & h \end{array} \\ \boldsymbol{K} = & \left[\begin{array}{cccccccc} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ & & & \cdots & & & & \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ \cdot \\ m \end{array} \end{array}$$

Note that the number of rows m in this matrix should be large enough, otherwise it would be practically impossible to extract any useful knowledge from the data.

By contrast, when presenting knowledge it is more convenient for future operations to use *disjunctive ternary matrices* in which all rows are interpreted as elementary disjunctions, called also simply *disjuncts*.

Suppose that $X = \{a, b, c, d, e, f\}$ and consider the implicative regularity $ab'e = 0$ forbidding the combination 101 of values of the attributes $a, b, e,$ accordingly. The corresponding empty interval of the space M contains eight elements: 100010, 100011, 100110, 100111, 101010, 101011, 101110 and 101111. The equation $ab'e = 0$ may be changed for the equivalent equation $ab'e \rightarrow 0$ with the implication operator $\rightarrow$ (if... then...), known as the *sequent* (its left part is always a conjunction, and the right part is a disjunction). The latter equation may be subjected to equivalence transformations consisting of transferring arbitrary literals between the left part (conjunction) and the right one (disjunction), changing each time their type (positive for negative or *vice versa*). In such a way it is possible to obtain the following set of the equivalent equations $ae \to b$ (if $a = 1$ and $e = 1$, then $b = 1$), $ab' \to e'$, $a \to b \vee e'$, ..., $1 \to a' \vee b \vee e'$. The last one could be changed for the disjunctive equation $a' \vee b \vee e' = 1$.

A set of regularities given in such a form can be presented by a ternary disjunctive matrix $\boldsymbol{D}$, called below a *knowledge matrix*. For example, the knowledge matrix

$$\begin{array}{cc} & \begin{array}{cccccccc} a & b & c & d & e & f & g & h \end{array} \\ \boldsymbol{D} = & \begin{bmatrix} 1 & - & - & 0 & - & - & 0 & - \\ - & - & - & 1 & - & 1 & - & - \\ 0 & 1 & - & - & - & - & - & - \end{bmatrix} \end{array}$$

affirms that every object of the regarded area must satisfy the equations

$$a \vee d' \vee g' = 1,\ d \vee f = 1 \text{ and } a' \vee b = 1.$$

In other words, in the considered Boolean space there exists no object which has any of the following combinations of attribute values: ($a = 0$, $d = 1, g = 1$), ($d = 0, f = 0$) and ($a = 1, b = 0$).

The set of these equations can be reduced to one equation $D = 1$ where D is a CNF (conjunctive normal form) represented by matrix $\boldsymbol{D}$. In our case:

$$D = (a \vee d' \vee g')\,(d \vee f)\,(a' \vee b).$$

2.3 Data and Knowledge Representation – the Case of Multi-Valued Attributes

In the general case of multi-valued attributes, it is more convenient to use sectional Boolean vectors and matrices introduced for the representation of finite predicates [Zakrevskij, 1990], [Zakrevskij, 1993]. A *sectional Boolean vector* consists of some sections (domains) corresponding to attributes and each section has several binary digits corresponding to the attribute values indicating definite properties. For example, the section corresponding to the attribute *color*, which has the values *blue*, *red*, *green*, *yellow*, *brown*, *black* and *white*, should have 7 bits.

Suppose that $X = \{x, y, z\}$, and the attributes x, y, z select their values from the corresponding sets $V_1 = \{a, b, c\}$, $V_2 = \{a, e, f, g\}$, $V_3 = \{h, i\}$ (note that these sets may intersect). Then vector 010.1000.01 describes an object with the value b of the attribute x, the value a of the attribute y and the value i of the attribute z. If a vector represents some element of the space M of multi-valued attributes, it has the only 1 in each section. The situation is different in the case of having some fuzziness. Vector 011.1001.01 can be interpreted as presenting a partial information about the object, when we know only that $x \neq a$, $y \neq e$, $y \neq f$ and $z \neq h$. Note, that each of these inequalities serves as an *information quantum* and is marked by a zero in the corresponding component of the vector.

In the case of finite predicates, generalized conjuncts and disjuncts can be used to present the knowledge [Zakrevskij, 1994]. Any interval in the space of multi-valued attributes is defined as a direct product of non-empty subsets α_i taken by one from each set V_i. Its characteristic function is defined as a conjunct, and the negation of the latter is a disjunct.

Considering the previous example, suppose that $\alpha_1 = \{a\}$, $\alpha_2 = \{a, e, g\}$, and $\alpha_3 = \{h, i\}$. The interval $I = \alpha_1 \times \alpha_2 \times \alpha_3$ presented by the vector 100.1101.11 has the characteristic function (conjunct)

$$k = (x = a) \wedge ((y = a) \vee (y = e) \vee (y = g)) \wedge ((z = h) \vee (z = i)),$$

which could be simplified to:

$$k = (x = a) \wedge ((y = a) \vee (y = e) \vee (y = g)),$$

in as much as $(z = h) \vee (z = i) = 1$. If this product enters the equation $k = 0$ reflecting a regular connection between x and y, then $I \cap W = \varnothing$, i.e. interval I turns out to be empty. The regularity can be represented by the equation

$$(x = a) \wedge ((y = a) \vee (y = e) \vee (y = g)) = 0.$$

As it can be seen from the above example, the structure of a conjunctive term in the finite predicate algebra is more intricate compared with that of the binary case – the two-stage form of the type $\wedge \vee$ is inherent in it. One can avoid that complexity changing the equation $k = 0$ for the equivalent equation $\neg k = 1$ and transforming $\neg k$ into a one-stage disjunctive term d. Such transformation is based on the de-Morgan rule and changes expressions $\neg (x_i \in \alpha_i)$ for equivalent expressions $x_i \in V_i \backslash \alpha_i$. This is possible since all sets V_i are finite.

For the considered example we have:

$$\begin{aligned} d = \neg k &= (x \neq a) \vee ((y \neq a) \wedge (y \neq e) \wedge (y \neq g)) = \\ &= (x = b) \vee (x = c) \vee (y = f). \end{aligned}$$

Hence, the same regularity can be expressed in a more compact form, as follows:

$$(x = b) \vee (x = c) \vee (y = f) = 1.$$

Suppose that the knowledge about the world obtained either from experts or by induction from data is represented by a set of disjuncts d_1, d_2, ..., d_m. The corresponding equations $d_i = 1$ are interpreted as conditions which should be satisfied for any objects of the world, and it is possible to reduce these equations to a single equation $D = 1$ the left part of which is presented in the conjunctive normal form (CNF) $D = d_1 \wedge d_2 \wedge \ldots \wedge d_m$. It follows from here that in the finite predicate algebra the CNF has some advantage over the disjunctive normal form (DNF) $K = k_1 \vee k_2 \vee \ldots \vee k_m$ which is used in the equivalent equation $K = 0$. Indeed, DNF has three stages $(\vee \wedge \vee)$, whereas CNF has only two $(\wedge \vee)$.

Suppose for instance, that $X = \{a, b, c\}$, $V_1 = \{1, 2, 3\}$, $V_2 = \{1, 2, 3, 4\}$ and $V_3 = \{1, 2\}$. Then the knowledge matrix

$$\begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \boldsymbol{D} = & \begin{bmatrix} 0\,0\,1 & . & 0\,0\,1\,0 & . & 0\,0 \\ 1\,1\,0 & . & 0\,0\,1\,1 & . & 0\,1 \\ 0\,1\,0 & . & 1\,1\,0\,0 & . & 1\,0 \\ 0\,0\,1 & . & 0\,1\,0\,0 & . & 0\,1 \end{bmatrix} \end{array}$$

may be interpreted as a set of disjunctive equations as follows:

$$(a = 3) \vee (b = 3) = 1,$$
$$(a = 1) \vee (a = 2) \vee (b = 3) \vee (b = 4) \vee (c = 2) = 1,$$
$$(a = 2) \vee (b = 1) \vee (b = 2) \vee (c = 1) = 1,$$
$$(a = 3) \vee (b = 2) \vee (c = 2) = 1$$

or as one equation with a CNF in the left part:

$$((a = 3) \vee (b = 3)) \wedge ((a = 1) \vee (a = 2) \vee (b = 3) \vee (b = 4) \vee (c = 2)) \wedge$$
$$\wedge ((a = 2) \vee (b = 1) \vee (b = 2) \vee (c = 1)) \wedge ((a = 3) \vee (b = 2) \vee (c = 2)) = 1.$$

3. DATA MINING – INDUCTIVE INFERENCE

3.1 Extracting Knowledge from the Boolean Space of Attributes

A very important part of the pattern recognition problem is obtaining knowledge from data [Frawler, Piatetsky-Shapiro, *et al.*, 1991]. The data could be represented by a *sampling population* F – a set of some randomly selected elements from the regarded world W.

As it was formulated above, we solve that problem by analyzing the distribution of elements of set F in the Boolean space M and revealing implicative regularities which are reflected by empty intervals (not intersecting with F). That operation can be reduced to observing a Boolean data matrix $\boldsymbol{K}$ and looking for such combinations of attribute values which do not occur in the matrix.

The number of attributes coming into an implicative regularity is called its rank. It coincides with the rank of the corresponding interval. The less attributes are tied with a regularity, the stronger is the tie, as will be shown below. So, it is worthwhile to look for regularities of smaller rank. Consider, for example, the following data matrix:

$$\begin{array}{c} \begin{matrix} a & b & c & d & e & f \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

There are no empty intervals of rank 1, because each column contains 1s and 0s. So we look further for empty intervals of rank 2 and find five of them, corresponding to the following combinations: ($a = 0$, $f = 1$), ($b = 1$, $d = 0$), ($b = 0$, $e = 0$), ($c = 1$, $f = 1$), ($d = 0$, $e = 0$). In a more compact form these intervals may be represented by conjuncts $a'f$, bd', $b'e'$, cf, $d'e'$. Can we consider that these found empty intervals reflect real regularities inherent in the world from which the data were extracted? Such conclusions could be accepted if only they are plausible enough.

Consider the general case of n binary attributes and m elements in the sampling population (selection) F. Suppose, we have found an empty interval of rank r (comprising 2^{n-r} elements of the Boolean space M) and put forward the corresponding hypothesis, affirming that this interval is free of any elements from the regarded world W. May we rely on it and derive with its help some logical conclusions when recognizing an object with the unknown value of the goal attribute? The problem is to estimate the *plausibility* of that hypothesis.

We should take into account that the regarded interval could be empty quite accidentally, as in reality the selection F is taken by random from the whole space M. In that case there could be no regularities in the disposition of the elements from F in M.

It would be useful to find the probability p of such an event as a function $p(n, m, r)$ of the parameters n, m, r. The hypothesis can be accepted and used further in procedures of deductive inference only if this probability is small enough. Its calculation is rather difficult, so it was proposed in [Zakrevskij, 1982] to approximate it by the mathematical expectation $E(n, m, r)$ of the number of empty intervals of rank r.

That value can be calculated by the formula

$$E(n, m, r) = C_n^r \, 2^r (1-2^{-r})^m , \tag{1}$$

where C_n^r is the number of r-element subsets of an n-element set, $C_n^r \, 2^r$ is the number of intervals of rank r in the space M, and $(1-2^{-r})^m$ is the probability of some definite interval of rank r being empty, not containing any elements from F.

Some empty intervals could intersect, hence $E(n, m, r) \geq p(n, m, r)$. The question is how big could be the difference $E(n, m, r) - p(n, m, r)$? It was shown, that it becomes negligible small for small values of $E(n, m, r)$. But that is just the case of interest for us.

It turns out that the value of the function $E(n, m, r)$ grows very rapidly with rising r. That is evident from Table 1 of the dependence E on r under some fixed values of the other parameters: $n = 100$ and $m = 200$.

Table 1. The Dependency of E on r Under Fixed n and m.

R	1	2	3	4	5	6
$E(100, 200, r)$	1.24×10^{-58}	2.04×10^{-21}	3.26×10^{-6}	1.56×10^{2}	4.21×10^{6}	3.27×10^{9}

It is clear that searching for empty intervals and putting forward corresponding hypotheses can be restricted in this case by the relation $r < 4$. If an empty interval of rank $r < 4$ is found, we have good reasons to formulate the corresponding regularity, but there are no grounds for that if $r \geq 4$. So, when $n = 100$ and $m = 200$, there is no sense in looking for empty intervals of ranks more than 3. The search for regularities could be strongly restricted in that case by checking for emptiness only intervals of rank 3, which number is $C_{100}^3 \times 2^3 = 1{,}293{,}600$. This is not much, compared to the number 3^{100} of all intervals in the Boolean space of 100 variables, approximately 5.15×10^{47}.

A threshold ω may be introduced to decide whether it is reasonable to regard an empty interval as presenting some regularity: the positive answer should be given when $E < \omega$. Its choice depends on the kind of problems to be solved on the base of found regularities.

Suppose $\omega = 0.01$. Then the maximum rank r_{max} of intervals which should be analyzed when looking for regularities could be found from Table 2, showing its dependence on n and m :

Table 2. The Dependency of the Maximum Rank r_{max} on the Parameters n and m.

$n \backslash m$	20	50	100	200	500	1,000
10	1	2	3	4	5	6
30	1	2	2	3	4	5
100	1	1	2	3	4	5

Two conclusions, justified for the regarded range of parameters, could follow from this table. First, in order to increase r_{max} by one it is necessary to double the size of the experiment, measured by the number m of elements in F. Second, approximately the same result could be achieved by reducing by a factor of 10 the number of attributes used for the description of the regarded objects.

Suppose $r_{max} = 2$ which is justified when the selection F is rather small. In that case we have to pay attention only to pairs of attributes, looking for some forbidden combinations of their values. This task can be executed by an incremental algorithm. Such an algorithm analyzes the elements of the selection F consecutively, one by one, and fixes such two-element combinations which have occurred. This is done by using a symmetrical square Boolean $2n \times 2n$ matrix $\boldsymbol{S}$ for that, with rows and columns corresponding to the values $x_1 = 0$, $x_1 = 1$, $x_2 = 0$, $x_2 = 1$, etc. This matrix is presented in a convenient form in Tables 3 and 4. Its elements corresponding to occurring combinations are marked with 1. The rest of the combinations (not occurring) are presented by zero (empty) elements and are accepted as forbidden. The regularities presented by them connect some attributes in pairs and are called *syllogistic regularities* [Zakrevskij, 1988].

For example, let us consider the following selection of data F from the world W (only to illustrate the algorithm, despite the fact that the selection is too small for $r_{max} = 2$):

$$
\begin{array}{c c}
\begin{array}{ccccc} a & b & c & d & e \end{array} & \\
\begin{bmatrix}
0 & 1 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 0 & 0
\end{bmatrix} &
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\end{array}
$$

Begin with its first element 01001 and fix occurring combinations of values for $C_5^2 = 10$ different pairs of attributes, marking with 1s corresponding elements of matrix $\boldsymbol{S}$ (Table 3). Note that they are presented there in symmetric pairs, besides, to facilitate further operations, additional formal pairs ($x_i = 1$, $x_i = 1$) and ($x_i = 0$, $x_i = 0$) are included into that set and have found their place on the main diagonal. So the whole set ($5^2 = 25$ pairs) is presented by a minor produced by the vector multiplication of definite rows and columns corresponding to argument values in the vector 01001.

By considering in the same way other elements of the selection F and by taking into account new pairs generated by them we can find all occurring pairs of attribute values obtained by the analysis of the selection F. The result is shown in Table 4.

Table 3. Finding All the Occurring Pairs of the Attribute Values Generated by the Element 01001.

	a		*b*		*c*		*d*		*e*	
a	1			1	1		1			1
b										
	1			1	1		1			1
c	1			1	1		1			1
d	1			1	1		1			1
e										
	1			1	1		1			1

Table 4. Finding All the Occurring Pairs of the Attribute Values Generated by the Selection *F*.

	a		*b*		*c*		*d*		*e*	
a	1			1	1	1	1		1	1
		1	1	1	1		1	1		1
b		1	1		1			1		1
	1	1		1	1	1	1	1	1	1
c	1	1	1	1	1		1	1	1	1
	1			1		1	1		1	
d	1	1		1	1	1	1		1	1
		1	1	1	1			1		1
e	1			1	1	1	1		1	
	1	1	1	1	1		1	1		1

The zero elements of the resulting matrix point to the found syllogistic regularities. These regularities can be presented in another form, by the following ternary knowledge matrix $\boldsymbol{D}$.

$$\boldsymbol{D} = \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} 0 & 0 & - & - & - \\ 0 & - & - & 1 & - \\ 1 & - & 1 & - & - \\ 1 & - & - & - & 0 \\ - & 0 & 1 & - & - \\ - & 0 & - & 0 & - \\ - & 0 & - & - & 0 \\ - & - & 1 & 1 & - \\ - & - & 1 & - & 1 \\ - & - & - & 1 & 0 \end{bmatrix} \end{array}$$

When the selection F is noticeably bigger compared with the number of attributes, the maximum rank r_{max} of implicative regularities could be 3, 4 or even more. The run-time for their finding swiftly increases. Nevertheless it is restricted, because the number of intervals to be checked could be approximated by $C_n^3\, 2^3$, $C_n^4\, 2^4$, etc.

3.2 The Screening Effect

Assume that we have found an empty (not intersecting with F) interval of the Boolean space M, and the mathematical expectation E corresponding to it is small enough. Then the logic of reasoning given above enables us to reject (according to the *modus tollens* rule, i.e. by means of finding contradictions) the supposition that this interval is empty accidentally, when in reality no regularity exists in the subject area W. In that case the probability of selecting elements from M for including them into W is evenly distributed, as well as the probability of using them as elements of the selection F and elements of intervals of the given rank. The hypothesis about the existence of an implicative regularity corresponding to a found empty interval is compared with the hypothesis confirming the absence of any regularities, and as a result of that comparison, the first hypothesis is accepted.

The character of probability distribution could be changed substantially when a set of existing regularities is known *a priori* and when as a result of the selection F analysis additional hypotheses are put forward, based on 'experimentally' proved emptiness of some intervals. In this case the known regularities forbid some s combinations of attribute values and that leads to reducing the number of possible combinations from 2^n (admitted by the hypothesis of the absence of any regularities) to $w = 2^n - s$.

Let us consider a conjunct of rank r corresponding to an interval of the space M which intersects with the subject area W. Suppose that a large part of this interval (comprised of u elements) belongs to the forbidden region formed by the union of intervals corresponding to known regularities. As a result the number of possible elements of the interval is reduced from 2^{n-r} to $q = 2^{n-r} - u$. It follows from here that the probability of a random hit of some definite (not forbidden) point of the space M into the considered interval changes from 2^{-r} to q/w, which could sometimes perceptibly reduce the probability of this interval intersection with the random selection F. Such an effect is called a *screen effect*. It is equivalent to a conventional rise of the rank of the analyzed conjunct and, consequently, could lead to a considerable increase of the value E. In turn, that increase can distort the results of inductive inference, and that leads to obtaining some fictitious regularities.

Consider for example the Boolean space of four attributes a, b, c, d. Suppose that the subject area is characterized by three conjunct-regularities $a'bd'$, $a'cd$ and acd'. The corresponding intervals are shown in Figure 2.

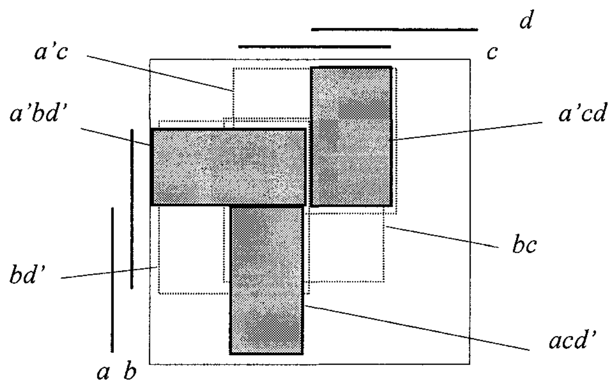


Figure 2. Illustrating the Screening Effect.

Look at the interval corresponding to conjunct $a'c$. It contains four elements presented by vectors 0010, 0011, 0110 and 0111. The three last elements belong to the forbidden region (they are screened by the given conjuncts), and only the first one does not contradict them. In this case

$r = 2, p = 10, q = 1$. As a result, the above probability decreases from 1/4 to 1/10, and that is equivalent to raising the rank of the regarded conjunction (in comparison with the case of absence of any regularities) to more than one. Similar conclusions follow from considering two other intervals which correspond to conjuncts bd' and bc.

Experimental research has confirmed that the effect of generating fictitious regularities during inductive inference could be rather big. One can compensate that screening effect by increasing the volume m of selection F or by putting an additional restriction on the rank r of the analyzed intervals.

3.3 Inductive Inference from Partial Data

It was supposed above that all objects of an experimental selection are described by their complete abstract models, with defined values of all attributes. A more general case of *partial data* presenting incomplete information about these objects was investigated in [Zakrevskij and Vasilkova, 1997]. For instance, we can know that some object has the attribute a and has not b, but we do not know if it has c.

Next suppose that the selection contains m objects which should be represented by some points of the Boolean space of n attributes, in other words by Boolean vectors with n components. However, as a result of a certain fuzziness of the data the values of several attributes (different for diverse objects) could remain unknown. It follows from here that the disposition of those points is defined to approximation of some intervals, and it is confirmed only that these intervals are not empty.

Suppose also that the *degree of fuzziness* is given by the parameter u, the same for all objects and attributes probability of the event "the value of a current attribute of a current object is unknown".

Such uncertainty can be marked by the symbol "-" in corresponding components of the vectors describing the regarded objects, changing in such a way the Boolean vectors for ternary ones. However, it is more convenient, from the computer point of view, to use sectional Boolean vectors with two-component sections in the case of binary attributes.

As a result, the information about the experimental data set is presented by a Boolean data matrix $\boldsymbol{K}$ of size $m \times 2n$, in which columns are divided into sections, two columns in each. The regularities are extracted just from this matrix.

During inductive inference, disjuncts presented by $2n$-component Boolean vectors are checked one by one, and those of them which are satisfied by every element of the given selection are accepted. The search begins with disjuncts of the minimum rank (1) and terminates by disjuncts of a certain "critical" rank. The current disjunct is compared with all rows of matrix $\boldsymbol{K}$ and, if it does not contradict with any of them (that is, if the component-wise conjunction of the compared vectors differs from the vector **0**), it is considered as satisfied. Then it can be accepted as a regularity on the condition that the probability of its accidental origin is small enough.

Computation of that probability is rather complicated, but in the case of small values it can be well approximated by the mathematical expectation $E(m, n, r, u)$ of the number of disjuncts of rank r, which are satisfied by every element of the data. Note that the latter is a random selection from M represented by the corresponding $m\times 2n$ Boolean matrix $\boldsymbol{K}$ eroded according to the parameter u.

The value of E is defined as the product of the number $C_n^r 2^r$ of different disjuncts of rank r in the space of n binary attributes and the probability $(1-\frac{(1-u)^r}{2^r})^m$ that one of them (arbitrarily chosen) does not contradict the data presented by the matrix $\boldsymbol{K}$:

$$E(m,n,r,u) = C_n^r 2^r (1-\frac{(1-u)^r}{2^r})^m. \qquad (2)$$

It is evident, that E swiftly increases with rising the rank r. The strong dependence of E on r facilitates finding the critical rank r^* for disjuncts checked during inductive inference. Knowing that value restricts the volume of computations performed while looking for regularities: it is defined by the number N of checked disjuncts for which ranks must not exceed r^*. This number can be obtained by the formula

$$N = \sum_{r=1}^{r^*} C_n^r 2^r. \qquad (3)$$

3.4 The Case of Multi-Valued Attributes

It is a little more difficult to extract knowledge from the space of n multi-valued attributes $x_1, x_2, \ldots, x_n$, see for example [Zakrevsky, 1994]. To begin with, define the probability p that a disjunct will be satisfied by an accidentally chosen element of the space. It could be calculated by the formula

$$p = 1 - \prod_{i=1}^{n} \frac{r_i}{s_i}, \qquad (4)$$

where s_i is the number of all values of the attribute x_i, and r_i is the number of those of them which do not enter this disjunct. For instance, for the disjunct 00.1000.101 $p = 1 - (2/2)\times(3/4)\times(1/3) = 3/4$. Let us divide all disjuncts into classes D_j, forming them from disjuncts with the same value of p. Next let us number these classes in order of increasing p and introduce the following conventional signs: q_j is the number of disjuncts in the class D_j, p_j is the value of p for elements from D_j.

Find now the mathematical expectation E_j of the number of disjuncts from the class D_j, which do not contradict the random m-element selection from the regarded space:

$$E_j = q_j\,(p_j)^m, \tag{5}$$

and introduce the analogous quantity E_k^+ for the union of classes D_1, D_2, …, D_k:

$$E_k^+ = \sum_{i=1}^{k} E_j. \tag{6}$$

Inductive inference is performed by consecutively regarding classes D_j in order of their numbers and summarizing corresponding values E_j until the sum surpasses a threshold t, which is introduced by taking into account the specific of the considered subject area. All disjuncts belonging to these classes are accepted as regularities if they do not contradict the data, i.e. if they are satisfied by any element of the selection F.

An expert may fix several thresholds and assign accordingly different levels of plausibility to the found regularities. For example, regularities obtained by thresholds 10^{-10}, 10^{-6}, 10^{-3} could be estimated as *absolutely plausible*, *usually*, *most likely*, respectively. This differentiation gives some flexibility to recognition procedures. Choosing a proper level of plausibility one can use only some of regularities contained in the knowledge base and vary in such a way the plausibility of the logical conclusions obtained during recognition. For example, using only the most plausible regularities can result in obtaining a small number of logical conclusions, but more reliable ones, while extending the used part of the knowledge base extends the set of obtained logical conclusions, at the expense of their plausibility.

4. KNOWLEDGE ANALYSIS AND TRANSFORMATIONS

4.1 Testing for Consistency

Any disjunctive knowledge matrix $\boldsymbol{D}$ is *consistent* if the corresponding conjunctive normal form (CNF) D is *satisfiable*, i.e. if there exists at least one solution of the equation $D = 1$. Checking $\boldsymbol{D}$ for consistency is a hard combinatorial problem [Cook, 1971]. In the general case it involves an unavoidable exhaustive procedure, which could be significantly reduced by a tree searching technique, taking into account the specific features of the regarded problem.

In the case of binary attributes the knowledge matrix $\boldsymbol{D}$ could be presented as a ternary matrix, and checking it for consistency is equivalent to looking for a Boolean vector, which satisfies every row $\boldsymbol{d}_j$ of matrix $\boldsymbol{D}$, either having 1 in some component where $\boldsymbol{d}_j$ has 1 or having 0 in some component where $\boldsymbol{d}_j$ has 0.

This task is equivalent to another one, over a ternary matrix $\boldsymbol{C}$ that represents the disjunctive normal form (DNF) of the characteristic Boolean function V for the prohibited area of the Boolean space M (taking value 1 on the area elements). That function V is called a *veto function*. Evidently, the equations $D = 1$ and $V = 0$ are equivalent, and matrix $\boldsymbol{C}$ could be easily obtained from matrix $\boldsymbol{D}$ by the component-wise inversion illustrated below.

$$\boldsymbol{D} = \begin{bmatrix} 0 & - & 1 & 1 & - \\ - & 1 & 0 & - & - \\ 1 & - & - & - & 0 \\ - & 1 & - & - & 1 \\ 0 & - & - & 0 & - \end{bmatrix}, \qquad \boldsymbol{C} = \begin{bmatrix} 1 & - & 0 & 0 & - \\ - & 0 & 1 & - & - \\ 0 & - & - & - & 1 \\ - & 0 & - & - & 0 \\ 1 & - & - & 1 & - \end{bmatrix}.$$

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 \end{matrix} \qquad\qquad \begin{matrix} 0 & 1 & 0 & 0 & 0 \end{matrix}$$

In this case the function V should be checked for identity (the relation $V \equiv 1$ is verified by that), and that corresponds to checking $\boldsymbol{C}$ for degeneration. Matrix $\boldsymbol{C}$ is called *degenerate ternary matrix* if and only if no Boolean vector exists orthogonal to every row of the matrix. Remind that two vectors are orthogonal if a component exists where one vector has the value 0 and the other has the value 1. For the regarded example we can find such a vector (it is shown under $\boldsymbol{C}$), and the same vector (shown under $\boldsymbol{D}$) satisfies matrix $\boldsymbol{D}$. That proves the consistency of the latter.

The discussed problem is well known; it is enough to say that it lies in the base of the theory of computational complexity. Many methods and algorithms were developed for its solution, for example [Davis, Longemann, *et al.*, 1962], [Zhang, 1997], so we shall not go here in detail.

A rather efficient method of checking a ternary matrix C for degeneration has been suggested in [Zakrevskij, 1988]. It realizes the idea "to try and find a Boolean vector w orthogonal to every row of matrix C" and uses the tree searching technique described in [Nilsson, 1971]. By that the tree vertices present current situations and edges proceeding from them point to possible choices of some component values. The sought-for vector w is constructed by assigning definite values to its components (when moving forward) one by one, and reassigning them (after it moves backward). So during the search process it is a variable ternary vector changing with time. A current situation is presented by a current value of the vector w and a minor $\boldsymbol{T}$ of matrix C obtained from C by deleting the satisfied rows (orthogonal to w) and columns which correspond to the variables having accepted some values. To reduce calculations some rules are used that allow avoiding alternative situations as much as possible.

That method implements the deductive inference of the *modus tollens* rule, illustrated with the following example of a matrix C and a search tree (see also Figure 3) which shows that matrix C is degenerate.

$$
\begin{array}{cc}
\begin{array}{cccccc} a & b & c & d & e & f \end{array} & \\
\left[\begin{array}{cccccc}
1 & - & - & - & - & 0 \\
1 & - & 1 & - & 1 & - \\
1 & 0 & - & - & 1 & 1 \\
1 & - & - & - & 0 & 1 \\
1 & 1 & 0 & - & - & 1 \\
0 & 1 & - & - & - & 1 \\
0 & - & - & - & 0 & 1 \\
0 & 0 & - & - & 1 & - \\
0 & - & - & 0 & 1 & 0 \\
0 & 1 & - & 1 & 1 & 0 \\
0 & 1 & - & - & 0 & 0 \\
0 & - & 1 & - & 0 & 0 \\
0 & 0 & 0 & - & 0 & 0
\end{array}\right] &
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \end{array}
\end{array}
$$

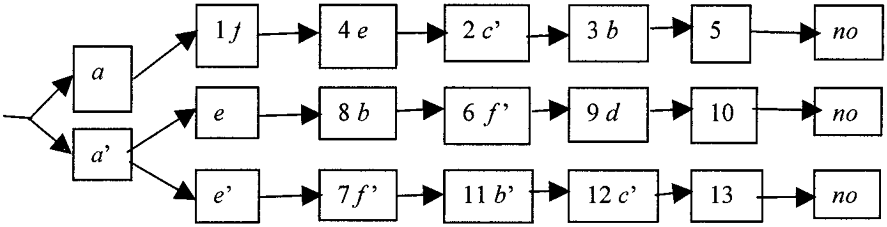


Figure 3. A Search Tree.

The tree presents the following reasoning. Suppose that a vector $\boldsymbol{w}$ exists orthogonal to every row of matrix $\boldsymbol{C}$. Consider the variable a corresponding to the most defined column (i.e., the one with the minimum number of *don't care* elements). It could have the value either 1 or 0. If a = 1, then the rows from 6 to 13 are satisfied. Now in order to satisfy row 1 it is necessary for $\boldsymbol{w}$ to have f = 1, otherwise $\boldsymbol{w}$ cannot be orthogonal to that row. For the same reason it is necessary to accept e = 1 (look at row 4), then c = 0 (row 2), then b = 1 (row 3). After that it becomes impossible to satisfy row 5, so we have to return to the beginning of the produced chain and try the other value of the variable a.

If a = 0, then the rows from 1 to 5 are satisfied and we have to satisfy the remaining rows, from 6 to 13. Consider now the variable e. If e = 1, then from necessity b = 1 (row 8), f = 0 (row 6) and d = 1 (row 9), after which it is impossible to satisfy row 10. Else if e = 0, then follow f = 0 (row 7), b = 0 (row 11), c = 0 (row 12), after that it is impossible to satisfy row 13.

So, we have to admit that the value 0 of the variable a is also invalid, and it is impossible to construct a vector $\boldsymbol{w}$ orthogonal to every row of matrix $\boldsymbol{C}$. Hence, that matrix is degenerate.

In the case of multi-valued attributes the knowledge is presented by a sectional disjunctive Boolean matrix $\boldsymbol{D}$. Any solution of the equation D = 1 corresponds to a column minor, which includes exactly one column from each domain and has at least one 1 in every row. Checking $\boldsymbol{D}$ for consistency is more difficult now. The tree searching technique is necessary here. To facilitate the regarded task some rules formulated below can be used for reducing the size of matrix $\boldsymbol{D}$.

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be some rows of matrix $\boldsymbol{D}$, while $\boldsymbol{p}$ and $\boldsymbol{q}$ are some of its columns. Consider vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ be satisfying the relation $\boldsymbol{a} \geq \boldsymbol{b}$ if the latter is satisfied component-wise. Let us say in this case that $\boldsymbol{a}$ covers $\boldsymbol{b}$.

The following reduction rules enable us to simplify matrix $\boldsymbol{D}$ by expelling some rows or columns.

Rule 1: If $\boldsymbol{u}$ covers $\boldsymbol{v}$, then row $\boldsymbol{u}$ is expelled.

Rule 2: If a column $\boldsymbol{p}$ is empty (not having 1s), then it is expelled.

Rule 3: If there exists a row $\boldsymbol{u}$ having 1s only in one domain, then all columns of that domain which have 0 in row $\boldsymbol{u}$ are expelled.

These rules define the equivalence transformations of $\boldsymbol{D}$ which do not change the set of solutions of equation $D = 1$. When checking $\boldsymbol{D}$ for consistency two more rules may be added. Their application can change the set of roots but does not violate the property of consistency: any consistent matrix remains consistent, and any inconsistent matrix remains inconsistent.

Rule 4: If $\boldsymbol{p}$ covers $\boldsymbol{q}$ and these columns belong to the same domain, then column $\boldsymbol{q}$ is expelled.

Rule 5: If a row $\boldsymbol{u}$ has a section without 0s, then it can be expelled.

Rules 1 and 4 are the most useful. For example, regarding the knowledge matrix

$$\boldsymbol{D} = \begin{array}{cc} \begin{array}{c} a_1a_2a_3 \quad b_1b_2b_3 \quad c_1c_2c_3c_4 \end{array} & \\ \begin{bmatrix} 0\,1\,1 \,.\, 0\,0\,1 \,.\, 0\,0\,0\,0 \\ 1\,0\,0 \,.\, 1\,0\,1 \,.\, 0\,1\,0\,0 \\ 1\,0\,1 \,.\, 0\,1\,0 \,.\, 1\,0\,0\,1 \\ 1\,0\,0 \,.\, 0\,1\,0 \,.\, 1\,0\,1\,0 \\ 0\,1\,1 \,.\, 0\,1\,0 \,.\, 0\,1\,0\,1 \\ 0\,1\,0 \,.\, 1\,0\,1 \,.\, 1\,0\,0\,0 \end{bmatrix} & \begin{array}{c} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{array} \end{array}$$

we can use these rules for deleting one by one the following columns and rows: column $\boldsymbol{b}_1$ (covered by $\boldsymbol{b}_3$), column $\boldsymbol{c}_3$ (covered by $\boldsymbol{c}_1$), row $\boldsymbol{d}_3$ (covering now $\boldsymbol{d}_4$), column $\boldsymbol{c}_4$ (covered by $\boldsymbol{c}_2$), column $\boldsymbol{a}_3$ (covered by $\boldsymbol{a}_2$) and row $\boldsymbol{d}_6$ (covering $\boldsymbol{d}_1$ after deleting $\boldsymbol{a}_3$). Executing these operations, we get a more compact matrix shown below:

$$\begin{array}{cc} \begin{array}{c} a_1a_2 \quad b_2b_3 \quad c_1c_2 \end{array} & \\ \begin{bmatrix} 0\,1 \,.\, 0\,1 \,.\, 0\,0 \\ 1\,0 \,.\, 0\,1 \,.\, 0\,1 \\ 1\,0 \,.\, 1\,0 \,.\, 1\,0 \\ 0\,1 \,.\, 1\,0 \,.\, 0\,1 \end{bmatrix} & \begin{array}{c} d_1 \\ d_2 \\ d_4 \\ d_5 \end{array} \end{array}$$

It is not difficult now to be convinced of the consistency of this matrix. For instance, it is satisfied by vector 10.01.01, presenting one of the solutions of the regarded system of disjunctive equations.

4.2 Simplification

Using a disjunctive knowledge matrix $\boldsymbol{D}$ in some expert system of logical recognition, it is natural to simplify it beforehand in order to facilitate subsequent repeated computations on its base. The simplification can consist in reducing the number of disjuncts as well as the number of 1s in the matrix in such a way that does not change the set of the matrix roots. So, there arises the problem of practical importance to look for a minimum disjunctive matrix equivalent to the given one.

That is the well-known problem of minimization of Boolean functions (or their DNFs, to be more correct), and hundreds of publications were devoted to it [Quine, 1952], [Karnaugh, 1953], [Nelson, 1955], [McCluskey, 1956], [Urbano and Mueller, 1956], [Zakrevskij, 1965], etc. We have no place to discuss them here. However, we would like to note that this problem was expanded onto finite predicates, which is vital for solving combinatorial tasks in the space of multi-valued attributes.

The problem of Boolean function minimization could be regarded both in exact and approximate formulations. Of course, looking for minimum DNFs is rather time-consuming, but suboptimal solutions could be very often quite satisfactory for practical purposes. For example, much less run-time is necessary to find some irredundant DNF which could serve as a good approximation to a minimum DNF. One can obtain it from an arbitrary DNF deleting from it some terms. Besides, some terms could be simplified by deleting some literals from them. The following algorithm for checking a ternary matrix for degeneration is very useful for these operations.

Let us regard the ternary matrix $\boldsymbol{C}$ presenting the DNF

$$acf' \vee be'f \vee a'd'e \vee de'f \vee b'c'f \vee b'df \vee c'e'f \vee b'c'd'.$$

$$\begin{array}{cc} & \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \boldsymbol{C} = & \begin{bmatrix} 1 & - & 1 & - & - & 0 \\ - & 1 & - & - & 0 & 1 \\ 0 & - & - & 0 & 1 & - \\ - & - & - & 1 & 0 & 1 \\ - & 0 & 0 & - & - & 1 \\ - & 0 & - & 1 & - & 1 \\ - & - & 0 & - & 0 & 1 \\ - & 0 & 0 & 0 & - & - \end{bmatrix} \end{array}$$

Checking its terms one by one we can see that some of them are implicants of the remaining part of the DNF and so can be deleted. Suppose the term represented by a row c_i of matrix $\boldsymbol{C}$ should be checked. Then this operation is reduced to checking for degeneration of a certain minor of matrix $\boldsymbol{C}$. That minor (denoted as $\boldsymbol{C} : \boldsymbol{c}_i$) is obtained by deleting from $\boldsymbol{C}$ row $\boldsymbol{c}_i$ together with all rows orthogonal to $\boldsymbol{c}_i$ and also columns where $\boldsymbol{c}_i$ has definite values (1 or 0). If the minor $\boldsymbol{C} : \boldsymbol{c}_i$ turns out to be degenerate, then row $\boldsymbol{c}_i$ (and corresponding term) should be deleted.

For example, the row (- - - 1 0 1) can be deleted from $\boldsymbol{C}$ but the row (- 1 - - 0 1) cannot, because the minor $\boldsymbol{C}$: (- - - 1 0 1) is degenerate and the minor $\boldsymbol{C}$: (- 1 - - 0 1) is not.

$$\boldsymbol{C}:(- - - 1\ 0\ 1) = \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} - & 1 & - \\ - & 0 & 0 \\ - & 0 & - \\ - & - & 0 \end{bmatrix} \end{array}, \qquad \boldsymbol{C}:(- 1 - - 0\ 1) = \begin{array}{c} \begin{array}{ccc} a & c & d \end{array} \\ \begin{bmatrix} 0 & 1 & - \\ - & - & 1 \\ - & 0 & 1 \end{bmatrix} \end{array}.$$

Checking all rows one by one, we find that three rows can be deleted from $\boldsymbol{C}$. As a result, we get an irredundant matrix $\boldsymbol{C}^*$. The corresponding irredundant knowledge disjunctive matrix $\boldsymbol{D}^*$ is easily obtained from $\boldsymbol{C}^*$ by the component-wise inversion.

$$\boldsymbol{C}^* = \begin{array}{c} \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{bmatrix} 1 & - & 1 & - & - & 0 \\ - & 1 & - & - & 0 & 1 \\ 0 & - & - & 0 & 1 & - \\ - & 0 & - & 1 & - & 1 \\ - & 0 & 0 & 0 & - & - \end{bmatrix} \end{array}, \qquad \boldsymbol{D}^* = \begin{array}{c} \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{bmatrix} 0 & - & 0 & - & - & 1 \\ - & 0 & - & - & 1 & 0 \\ 1 & - & - & 1 & 0 & - \\ - & 1 & - & 0 & - & 0 \\ - & 1 & 1 & 1 & - & - \end{bmatrix} \end{array}.$$

5. PATTERN RECOGNITION – DEDUCTIVE INFERENCE

5.1 Recognition in the Boolean Space

The recognition problem can be formulated as the problem of a closer definition of qualities of some observed object not belonging to the experimental selection from the subject area [Zakrevskij, 1999]. Suppose that we know the values of s from n attributes of this object. That is equivalent to locating the object in a certain interval of the Boolean space M presented by the corresponding elementary conjunction k of rank s. The problem is to define by logical reasoning, as well as possible, the values of the remaining $n - s$ attributes, using for that the information contained in the knowledge ternary matrix $\boldsymbol{D}$ and in the corresponding veto function V.

Let us regard the set X_k of attributes with known values and the set of all forbidden combinations of values of the remaining attributes – for the considered object. The latter set can be described by a proper Boolean veto function $V(k)$ that could be easily obtained from V. Indeed, it is sufficient for that to transform the formula representing the function V by changing symbols of attributes presented in k for values (0 or 1) satisfying the equation $k = 1$. Denote this operation as $V(k) = V{:}k$.

Suppose that we want to know the value of an attribute x_i which does not come into X_k. The necessary and sufficient condition for the prohibition of the value 1 of that attribute is presented by the formal implication $kx_i \Rightarrow V$, i.e. belonging of the interval presented by conjunction kx_i to the prohibition region described by the function V. Analogously, the necessary and sufficient condition for the prohibition of the value 0 is presented by $kx_i' \Rightarrow V$.

It is not difficult to deduce from here forecasting rules to define the value of the goal attribute x_i of the object characterized by k. These rules are shown in a compressed form in Table 5 presenting the decision (a set of possible values of x_i, the bottom row) as a function of predicates $kx_i \Rightarrow V$ and $kx_i' \Rightarrow V$.

Table 5. Forecasting the Value of the Attribute x_i.

$kx_i \Rightarrow V$	0	0	1	1
$kx_i' \Rightarrow V$	0	1	0	1
x_i	$\{0, 1\}$	$\{1\}$	$\{0\}$	$\varnothing$

Note that four outcomes could appear at this approach. On a level with finding the only value (0 or 1) for the attribute x_i, such situations could be met when both values are acceptable or neither of them satisfies the veto function V. At the later case the existence of the object α characterized by k contradicts the knowledge base, and that could stimulate some correction of the latter. However, the probability of such an event is low enough, taking into account the way of forming the knowledge base.

For example, if

$$V = acf' \vee be'f \vee a'd'e \vee b'df \vee b'c'd'$$

and $k = abf$, then $V(k) = V{:}abf = e'$. It could be concluded from this that the regarded object α has value 1 in attribute e, but there are no restrictions on the other attributes (c and d). If by the same function V the object α is characterized by $k = c'e'f$, then

$$V(k) = b \vee b'd \vee b'd' = 1 \text{ (all are forbidden)},$$

and that means that the object contradicts the knowledge.

Predicates $kx_i \Rightarrow V$ and $kx_i' \Rightarrow V$ are accordingly equivalent to predicates $V{:}kx_i = 1$ and $V{:}kx_i' = 1$, and that allows reducing their calculation by checking the corresponding submatrices of the knowledge matrix $\boldsymbol{D}$ for consistency. Fixing the values of some attributes in function V is changed for selecting a corresponding minor of matrix $\boldsymbol{D}$ by deleting some rows and columns, which could be followed by further possible simplification.

Suppose that we regard the same (already minimized) knowledge matrix $\boldsymbol{D}$ corresponding to the veto function $V = acf' \vee be'f \vee a'd'e \vee b'df \vee b'c'd'$ and know that for the observed object $a = 1$ and $c = 1$. Taking into account this new information we transform matrix $\boldsymbol{D}$ as follows. First we delete from it the columns marked with a and c because these variables became constant, and delete also the rows 3 and 5 now satisfied by these constants. Further simplification is rather evident, by using the following rule: $x\,(x' \vee H) = x\,H$, where x is a Boolean variable and H is an arbitrary Boolean formula.

$$D^* = \begin{array}{c} a\ b\ c\ d\ e\ f \\ \begin{bmatrix} 0 & - & 0 & - & - & 1 \\ - & 0 & - & - & 1 & 0 \\ 1 & - & - & 1 & 0 & - \\ - & 1 & - & 0 & - & 0 \\ - & 1 & 1 & 1 & - & - \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \quad \begin{array}{c} b\ d\ e\ f \\ \begin{bmatrix} - & - & - & 1 \\ 0 & - & 1 & 0 \\ 1 & 0 & - & 0 \end{bmatrix} \end{array} \quad \begin{array}{c} b\ d\ e\ f \\ \begin{bmatrix} - & - & - & 1 \\ 0 & - & 1 & - \\ 1 & 0 & - & - \end{bmatrix} \end{array}$$

We can conclude now that $f = 1$, by necessity. As to the remaining attributes, their values cannot be forecasted uniquely. They obey the next two conditions: $b' \vee e = 1$ and $b \vee d' = 1$. This system of logical equations has two solutions. Either $b = d = 0$ (with an arbitrary value of e), or $b = e = 1$ (with an arbitrary value of d).

Suppose the values of all attributes are known except the goal one. In that case solving the recognition problem could be facilitated by preliminary partitioning the Boolean space of attributes into four regions. After that it would be sufficient only to conclude to which of them the regarded object belongs and make the corresponding conclusion.

The characteristic Boolean functions of these regions are obtained on the base of the rules shown in Table 5. The region where the value of the attribute x_i remains unknown is described by the function $V(x_i) =$

$(V{:}x_i)'\wedge(V{:}x_i')'$, the region where x_i receives the value 1 is presented by the function $V^1(x_i) = (V{:}x_i)'\wedge(V{:}x_i')$, the region where x_i receives the value 0 – by the function $V^0(x_i) = (V{:}x_i)\wedge(V{:}x_i')'$, and the region of contradiction – by the function $V^*(x_i) = (V{:}x_i)\wedge(V{:}x_i')$.

By using the same example we obtain:

$$V = acf' \vee be'f \vee a'd'e \vee b'df \vee b'c'd',$$

$$V{:}f = be' \vee a'd'e \vee b'd \vee b'c'd',$$

$$V{:}f' = ac \vee a'd'e \vee b'c'd',$$

$$V^{-}(f) = (be' \vee a'd'e \vee b'd \vee b'c'd')' \wedge (ac \vee a'd'e \vee b'c'd')',$$

$$V^{1}(f) = (be' \vee a'd'e \vee b'd \vee b'c'd')' \wedge (ac \vee a'd'e \vee b'c'd'),$$

$$V^{0}(f) = (be' \vee a'd'e \vee b'd \vee b'c'd') \wedge (ac \vee a'd'e \vee b'c'd')',$$

$$V^{*}(f) = (be' \vee a'd'e \vee b'd \vee b'c'd') \wedge (ac \vee a'd'e \vee b'c'd').$$

5.2 Appreciating the Asymmetry in Implicative Regularities

So far it was silently assumed that every implicative regularity is symmetrical for all attributes included in it. In other words, all these attributes have equal rights. It was assumed, for example, that the disjunct $a \vee b \vee c$ can be transformed into any of the sequents $a'b' \to c$, $a'c' \to b$ and $b'c' \to a$, which could be used under recognition. It was assumed, therefore, that all these expressions are equivalent.

However, sometimes the symmetry of implicative regularities could be subjected to doubt. More accurate means may be suggested for their representation as well as the appropriate rules for using them in deductive inference.

Denote by $w(k)$ the number of elements from selection F which have sets of attribute values satisfying the equation $k = 1$ with a conjunctive term k (for example, the vector 10011 of values of the attributes a, b, c, d, e satisfies the equation $c'e = 1$).

Let us consider some irredundant disjunct, $a \vee b \vee c$ for instance, representing a regularity obtained as a result of the selection analysis fulfilled according to the formulated above rules. Evidently, $w(a'b'c') = 0$, but it could be that $w(ab'c') \neq 0$, $w(a'bc') \neq 0$ and $w(a'b'c) \neq 0$. Admit that the three last quantities can greatly differ by taking either too large or too small values, which is represented by the corresponding linguistic constants L and S.

Suppose that some object is observed and it is established that it has neither attribute a nor b. If it is known also that $w(a'b'c) = \mathrm{L}$, there are rather weighty reasons for the logical conclusion that the subject possesses the attribute c. Indeed, there are many subjects in the selection that also have neither a nor b and all of them have c. In other words, the hypothesis following from the observation has many confirmations and no refutation, hence it can be accepted as a regularity. But if $w(a'b'c) = \mathrm{S}$, the reasons for accepting this hypothesis seems to be flimsy and unconvincing, because the number of confirming examples is too small in this case. It can be said that the regarded situation itself has a low probability.

Hence, the transformation of the disjunct $a \vee b \vee c$ into the sequent $a'b' \rightarrow c$ seems to be reasonable when $w(a'b'c) = \mathrm{L}$ and not reasonable when $w(a'b'c) = \mathrm{S}$. Let us say that in the first case c is derived in the disjunct $a \vee b \vee c$ and the sequent $a'b' \rightarrow c$ is valid, and in the second case c is not derived and the sequent $a'b' \rightarrow c$ is not valid.

All that resembles the formalism of association rules [Agrawal, Imielinski, *et al.*, 1993]. However, unlike the latter, arbitrary (not only positive) conjunctive terms k are regarded. Besides, a set of formal transformations is suggested below for deriving new valid sequents from the given ones.

Consider a multiplace disjunct-regularity D, and choose in it a literal x^*, meaning that x^* could be either x or x'. Splitting D represent it as $x^* \vee D_1$. Let K_1 be the conjunctive term equal to the inversion of D_1. For instance, if $D_1 = a' \vee b$, then $K_1 = ab'$.

Affirmation 1. A literal x^* is derived in the disjunct $D = x^* \vee D_1$ if and only if $w(K_1) = \mathrm{L}$.

Characterize any disjunct by a list of derived literals marking them by an appropriate underlying in the formula. For example, $a \vee \underline{b} \vee \underline{c}$ is a disjunct that can be transformed into the sequents $a'b' \rightarrow c$ and $a'c' \rightarrow b$, but not into $b'c' \rightarrow a$. A disjunct is called *complete* if all its literals are derived.

According to the well-known resolution rule [Chang and Lee, 1973] it is possible under certain conditions to draw from some two disjuncts a third one that is their logical conclusion. This rule can be fully applied to complete disjuncts, but demands more precise definition in the general case.

Affirmation 2. The disjunct $D_3 = b \vee c$ follows logically from $D_1 = a \vee b$ and $D_2 = a' \vee c$. By that the literal c is derived in D_3 if and only if a is derived in D_1, and b is derived in D_3 if and only if a' is derived in D_2.

Proof: It follows from the existence of the disjuncts D_1 and D_2 that $w(a'b') = 0$ and $w(ac') = 0$. If a is derived in D_1, then $w(ab') = L$. This follows from the obvious equality $w(ab') = w(ab'c) + w(ab'c')$ and the above obtained $w(ac') = 0$, that $w(ab'c) = L$ and $w(b'c) = L$, i.e. c is derived in $D_3 = b \vee c$. On the other hand, if c is derived in D_3, then $w(b'c) = L$, and it follows from $w(a'b') = 0$ and $w(b'c) = w(ab'c) + w(a'b'c)$ that $w(ab'c) = L$ and $w(ab') = L$, i.e. a is derived in D_1. The proof of the second part of the theorem, concerning derivability of b in D_3, is quite similar. **End of proof.**

Corollary. The complete disjunct $\underline{b} \vee \underline{c}$ follows from the disjuncts $\underline{a} \vee b$ and $\underline{a}' \vee c$.

More complicated disjuncts can be characterized by lists of derived fragments − individual literals present only at a specific case. For the general case these fragments are represented by the right parts of valid sequents $K_1 \rightarrow D_2$ which could be generated by the corresponding disjuncts $D = D_1 \vee D_2$.

Affirmation 3. The sequent $K_1 \rightarrow D_2$ generated by the disjunct $D_1 \vee D_2$ is valid if and only if $w(K_1) = L$.

For example, $acd' \rightarrow b' \vee e$ generated by $a' \vee b' \vee c' \vee d \vee e$ is valid if and only if there are rather many subjects in F that have attributes a, c and have not d.

Affirmation 4. If the sequent K_1 is valid, then any other sequent obtained from it by transferring (with inversion) some literals from K_1 into D_2 is also valid.

For instance, if the sequent $a'b'c' \rightarrow d \vee e$ is valid, then $a'b' \rightarrow c \vee d \vee e$, $a'c' \rightarrow b \vee d \vee e$, $b' \rightarrow a \vee c \vee d \vee e$, etc. (called derivatives of $a'b'c' \rightarrow d \vee e$) are also valid.

It follows from Affirmation 4 that it is sufficient to mark in the characteristic of a disjunct only the minimum derived parts, because their arbitrary extensions will also be derived. For example, if $a \vee b$, $b \vee c \vee d$ and $c \vee d \vee e \vee f$ are marked in the disjunct $a \vee b \vee c \vee d \vee e \vee f$, then $a \vee b \vee e$, $b \vee c \vee d \vee f$, $a \vee c \vee d \vee e \vee f$, etc. are derived too. In this case only the sequents $c'd'e'f' \rightarrow a \vee b$, $a'e'f' \rightarrow b \vee c \vee d$ and $a'b' \rightarrow c \vee d \vee e \vee f$ as well as their derivatives are valid according to this marking.

Affirmation 5. The set of possible characteristics of a disjunct is isomorphic to the set of all monotonic Boolean functions of the disjunct literals.

Affirmation 6. For any disjuncts $D^* = x \vee D_1$ and $D^{**} = x' \vee D_2$ the disjunct $D = D_1 \vee D_2$ follows from them. By that D_1 is derived in D if and only if x' is derived in D^{**}, and D_2 is derived in D if and only if x is derived in D^*.

This affirmation can be regarded as a generalization of Affirmation 2 and can be proved in a similar way.

So, any system of implicative regularities constructed in accordance with the suggested definitions can be represented by boundary sequents, when it is not allowed to transfer symbols from the right part into the left one. In such a form it can be used in a system of deductive inference, facilitating the fulfillment of its procedures. Then the recognition of any observed subject is carried out by inserting into these sequents the values of some attributes obtained while observing the subject and by appropriate reducing the system.

5.3 Deductive Inference in Finite Predicates

In the case of multi-valued attributes the disjunctive knowledge matrix $\boldsymbol{D}$ turns out to be a sectional Boolean matrix presenting a finite predicate. There are some specifics in dealing with it as described in [Zakrevskij, 1990], [Zakrevskij, 1993].

Let us state the central problem of deductive inference: a disjunctive matrix $\boldsymbol{D}$ and a disjunct $\boldsymbol{d}$ mated with $\boldsymbol{D}$ (that means defined on the same pattern) are given, the problem is to find out whether $\boldsymbol{d}$ is a logical consequence of $\boldsymbol{D}$. In other words, the question is if the conjunctive term d is derived in CNF D? That is, does it become equal to 1 on all elements of the space M where CNF D takes value 1?

Two ways for solving such problems are known: the direct inference and the back inference.

When the direct inference is executed, the initial set of disjuncts is expanded consecutively by including new disjuncts following from some pairs of disjuncts existing already in the set. This procedure continues until the disjunct $\boldsymbol{d}$ is obtained or the set expansion is exhausted without obtaining $\boldsymbol{d}$ – in the last case it is proved that $\boldsymbol{d}$ does not follow from $\boldsymbol{D}$.

Any pair of disjuncts $\boldsymbol{u}$ and $\boldsymbol{v}$ can generate several disjuncts-consequents $\boldsymbol{w}_i$, obtained formally by the operation $\boldsymbol{w}_i = \boldsymbol{u}<x_i>\boldsymbol{v}$ which may be called the resolution in regard with the variable x_i and which can be considered as the generalization of the well-known in the theory of Boolean functions resolution operation onto finite predicates. It is defined as follows: the domain (section) of $\boldsymbol{w}_i$ corresponding to the variable x_i equals the component-wise conjunction of the corresponding domains from $\boldsymbol{u}$ and $\boldsymbol{v}$ (this can be considered as the unification by the variable x_i), and the rest domains equal the component-wise disjunction of the corresponding domains from $\boldsymbol{u}$ and $\boldsymbol{v}$.

However, not every disjunct obtained in such a way deserves subsequent consideration. There is no sense in including into the regarded set a disjunct which follows from some other disjunct belonging to the set, because it represents only some expansion of the latter one. For example, disjunct 110.0111.00 follows from disjunct 010.0110.00. It is reasonable to look only for non-trivial consequents. Such is a disjunct which follows from some pair of disjuncts $\boldsymbol{u}$ and $\boldsymbol{v}$ but does not follow from $\boldsymbol{u}$ or $\boldsymbol{v}$ taken separately. Let us call it a *resolvent* of disjuncts $\boldsymbol{u}$ and $\boldsymbol{v}$, and determine the rules for its obtaining.

Disjuncts $\boldsymbol{u}$ and $\boldsymbol{v}$ are called *adjacent* by the variable x_i if and only if the corresponding domains are incomparable (their component-wise disjunction differs from each of these domains) and there exists in each of the remaining domains a component with the value 0 in both vectors. Note that by violating the first condition a disjunct is obtained which follows either from $\boldsymbol{u}$ or from $\boldsymbol{v}$, whereas by violating the second condition a trivial (identical to 1) disjunct is found, which follows from any other disjunct.

Affirmation 7. If disjuncts $\boldsymbol{u}$ and $\boldsymbol{v}$ are adjacent by the variable x_i and $\boldsymbol{w} = \boldsymbol{u}\langle x_i\rangle\boldsymbol{v}$, then the disjunct $\boldsymbol{w}$ is a resolvent of the disjuncts $\boldsymbol{u}$ and $\boldsymbol{v}$.

For example:

$$
\begin{array}{rl}
 & \quad a \qquad b \qquad c \\
\boldsymbol{u} = & 1\,0\,0\,.\,1\,0\,.\,0\,0\,1\,1 \\
\boldsymbol{v} = & 0\,1\,0\,.\,0\,0\,.\,0\,1\,1\,0
\end{array}
$$

It is easy to see that these disjuncts are adjacent by a and also by c, but not by b. Hence, they give rise to the following two resolvents

$$
\begin{array}{rl}
\boldsymbol{u}\langle a\rangle\boldsymbol{v} = & 0\,0\,0\,.\,1\,0\,.\,0\,1\,1\,1 \\
\boldsymbol{u}\langle c\rangle\boldsymbol{v} = & 1\,1\,0\,.\,1\,0\,.\,0\,0\,1\,0
\end{array}
$$

The direct inference is simple but time-consuming because the number of obtained consequents could be very large. The back inference is more efficient. It solves the problem by transforming the initial system of disjuncts into such a system which is consistent if and only if $\boldsymbol{d}$ does not follow from $\boldsymbol{D}$. So, the problem is reduced to the regarded above problem of checking some disjunctive matrix for consistency.

Denoting by $\neg\,\boldsymbol{d}$ the vector obtained from $\boldsymbol{d}$ by its component-wise negation, and by $\boldsymbol{D} \wedge \neg\,\boldsymbol{d}$ the matrix obtained from $\boldsymbol{D}$ by the component-wise conjunction of each of its rows with vector $\neg\,\boldsymbol{d}$, the following rule may be formulated.

Affirmation 8. A disjunct $\boldsymbol{d}$ follows from a disjunctive matrix $\boldsymbol{D}$ if and only if the disjunctive matrix $\boldsymbol{D} \wedge \neg \boldsymbol{d}$ is not consistent.

Checking this condition is rather easy: 1s are expelled from all columns of $\boldsymbol{D}$ which correspond to components of the vector $\boldsymbol{d}$ having value 1, then the obtained disjunctive matrix is checked for consistency. For instance, if

$$\boldsymbol{D} = \begin{bmatrix} 0\,0\,1\,.\,0\,0\,1\,0\,.\,0\,0 \\ 1\,1\,0\,.\,0\,0\,1\,1\,.\,0\,1 \\ 0\,1\,0\,.\,1\,1\,0\,0\,.\,1\,0 \\ 0\,0\,1\,.\,0\,1\,0\,0\,.\,0\,1 \end{bmatrix}$$

and

$$d = \;0\,1\,1\,.\,1\,0\,0\,0\,.\,0\,0$$

then the following disjunctive matrix should be checked for consistency

$$\boldsymbol{D} \wedge \neg \boldsymbol{d} = \begin{bmatrix} 0\,0\,0\,.\,0\,0\,1\,0\,.\,0\,0 \\ 1\,0\,0\,.\,0\,0\,1\,1\,.\,0\,1 \\ 0\,0\,0\,.\,0\,1\,0\,0\,.\,1\,0 \\ 0\,0\,0\,.\,0\,1\,0\,0\,.\,0\,1 \end{bmatrix}$$

This matrix is not consistent. Hence, the disjunct $\boldsymbol{d}$ follows from $\boldsymbol{D}$.

5.4 Pattern Recognition in the Space of Multi-Valued Attributes

Let us describe a typical situation, where the problem of recognition arises, and some ways to solve it [Zakrevskij, 1992], [Zakrevskij, 1994], [Zakrevskij, 2001].

Suppose that an object from the world W is considered and some partial information about it is known, which can be represented by a set S of elementary prohibitions that are simplest information quanta of the type $x_i \neq v_j$: the value of attribute x_i in the object differs from v_j. In particular, this value is determined uniquely if all values except one are prohibited by such quanta. Suppose also that some definite attribute called a goal attribute is indicated, and its value should be found, without immediate measurement. This means that it must be calculated on the basis of the given information about the object and the known regularities inherent in the world W, to which the object belongs. Remember that these regularities are presented in the matrix $\boldsymbol{D}$.

Let us represent the partial information about the object by a vector-conjunct $\boldsymbol{r}$, where all elementary prohibitions are mapped as 0s in the corresponding components. This vector can be regarded as the interval of an initial localization of the object in the space M. Taking into account the regularities of the world W, one can make it more precise by reducing the area of the possible location of the object in the space of attributes.

It is easy to see that this area is defined as the intersection of the interval $\boldsymbol{r}$ with the set of solutions of the disjunctive matrix $\boldsymbol{D}$.

Affirmation 9. Let a vector-conjunct $\boldsymbol{r}$ present an object from the world W the regularities of which are represented by a disjunctive matrix $\boldsymbol{D}$. Then the area of possible location of the object is equal to the set of solutions of the disjunctive matrix $\boldsymbol{D}^* = \boldsymbol{D} \wedge \boldsymbol{r}$.

The disjunctive matrix $\boldsymbol{D}^*$ is found rather easily by deleting from matrix $\boldsymbol{D}$ all 1s in the columns which correspond to the elementary prohibitions. Getting matrix $\boldsymbol{D}^*$, we in a sense completely solve the recognition problem, by converting the knowledge about the world as a whole to the knowledge about a definite object. The latter knowledge can be reduced to a more compact form by the reduction rules described above. The values of some attributes, including the goal one, could be determined by that uniquely.

Rather general seems to be the formulation of the recognition problem as the problem of maximal expansion of the set of elementary prohibitions S. This problem can be interpreted as the interval localization of the object under recognition, the essence of which is that the area of its possible location should remain an interval after the reduction. It is convenient to decompose the process of such localization into the search for separate elementary prohibitions $x_i \neq v_j$, following from matrix $\boldsymbol{D}^*$.

Let us put in correspondence to the elementary prohibition $x_i \neq v_j$ the sectional Boolean vector $\boldsymbol{s}(i, v_j)$, in which all components of the domain i, except those which correspond to the value v_j, have value 1 and all components of the other domains have value 0.

Affirmation 10. The prohibition $x_i \neq v_j$ follows from matrix $\boldsymbol{D}^*$ if and only if the disjunctive matrix $\boldsymbol{D}^* \wedge \neg\, \boldsymbol{s}(i, v_j)$ is not consistent.

It can happen that matrix $\boldsymbol{D}^*$ is not consistent itself. That would testify the existence of some contradictions between regularities inherent in the set W and a partial information about the object. Otherwise solving the problem of the set S expansion leads to obtaining the minimum interval containing the area of possible location of the observed object, and this solution is always unique.

For example, if it is known that some object possesses value 1 of the attribute c, then by using vector $\boldsymbol{r}$ = 111.1111.10 we transform $\boldsymbol{D}$ into $\boldsymbol{D}^*$ and by simplifying the latter we get an equivalent matrix:

$$D = \begin{array}{c} \begin{array}{ccccccccccc} & a & & & & b & & & & c & \end{array} \\ \begin{bmatrix} 0\,0\,1 & . & 0\,0\,1\,0 & . & 0\,0 \\ 0\,0\,0 & . & 0\,0\,1\,1 & . & 0\,1 \\ 0\,1\,0 & . & 1\,1\,0\,0 & . & 1\,0 \\ 0\,0\,1 & . & 0\,1\,0\,0 & . & 0\,1 \end{bmatrix} \end{array}, \quad D^* = \begin{bmatrix} 0\,0\,0 & . & 0\,0\,1\,1 & . & 0\,0 \\ 0\,0\,0 & . & 0\,0\,0\,0 & . & 1\,0 \\ 0\,0\,1 & . & 0\,0\,0\,0 & . & 0\,0 \end{bmatrix}.$$

It follows from here that the initial vector-conjunct ***r*** transforms into 001.0011.10, which means that the regarded object is located in the space M to an accuracy of only two elements.

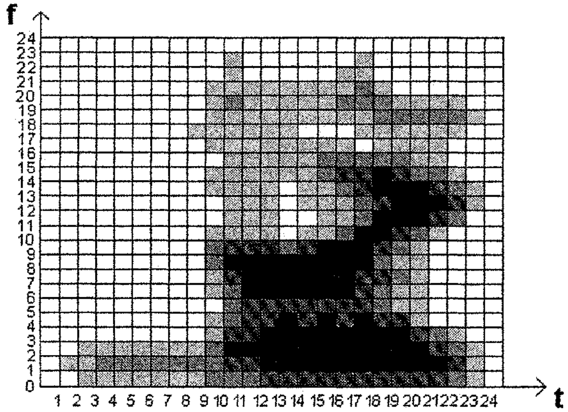
Sometimes expert systems of logical recognition must not only give correct logical conclusions but also provide them with clear explanations of some form. One such form can be a chain of consecutively executed resolutions leading to a sought-for disjunct-consequent. Of course, it can be found via direct inference of the regarded consequent, but in this case the chain as a rule turns out to be too long and inconvenient for visual perception. The problem of looking for minimized chains of logical inference arises in connection with that. A proper method for solving it was suggested in [Zakrevskij and Vasilkova, 1998]. It uses information contained in the search tree constructed when checking for consistency corresponding minors of the knowledge matrix D.

6. SOME APPLICATIONS

Several expert systems were constructed based on the suggested approach. For example, the system DIES was developed for running diagnostic tests on various engineering objects [Zakrevskij, Pechersky, *et al.*, 1988]. It uses the technique of logical inference in finite predicates.

The instrumental system EDIP was developed, based on the previous results and intended for running diagnostic tests on electric devices used in everyday life, for instance an iron [Zakrevskij, Levchenko, *et al.*, 1991]. It works in the dialogue mode: it first asks for some measurements and then it uses the results (together with knowledge provided by experts and presented by sectional Boolean matrices) to decide what is wrong with the device.

An experimental system for recognition of spoken words was developed in [Zakrevsky and Romanov, 1996]. The data for it are given in digital form and present the distribution of energy in the space "time × frequency" (see also Figure 4).



a) The Spectral Image of the Word

```
000000000000000000000000
000000000100000001000000
000000000100000110000000
000000001111111121000000
000000001211111222111100
000000001111111112222210
000000011111110001111100
000000001111111101111000
000000000111112222211000
000000001111112334322100
000000001111011234443210
000000000111011123444310
000000000111001124443200
000000001111011244321000
000000002333234443210000
000000002444444442110000
000000001344444433210000
000000001244444432110000
000000000233333332110000
000000001233434343310000
000000001334444444432100
001111111244444444443210
011222222233444444444310
001111112223333333333210
```

b) The Matrix of the Energy Levels

Figure 4. The Energy Distribution of the Pronunciation of the Russian Word "*nool*" (meaning "*zero*").

The information given by that distribution is reflected preliminary into a space of binary attributes, which present some specially selected geometrical characteristics of the image. That enables to apply the procedures of inductive and deductive inference described above.

The experimental system EXSYLOR for logical recognition in different areas was constructed [Zakrevskij, 1992]. That system is modeling subject areas in a finite space of multi-valued attributes. It can be tuned to any subject area (*World*) by enumerating names of essential attributes and their values and by forming in such a way a World model. Information about separate objects composing a representative selection from the World can be introduced into the system and fill a database. Description of regular connections between attributes represents information about the World as a whole; it is regarded as knowledge and constitutes the contents of a knowledge base. The knowledge is given in the form of sequents (if…, then…), prohibitions or disjuncts introduced into the system by an expert or deduced from the data. It is used in deductive procedures of computing values of the goal attributes of the objects under recognition.

The system EXSYLOR analyses data, produces and optimizes knowledge, extrapolates partial information about observed objects, and helps in planning measurements. If asked, it prepares information explaining its decisions. It is also provided with expert and recognizer interfaces.

The complete cycle of solving the problem of logical recognition by that system includes the following procedures:

- constructing the attribute space (defining all attributes and all their values);

- obtaining a representative selection of data, describing some set of objects from the investigated "World";
- executing inductive inference which enables to find some regularities inherent in the "World", revealing them from that selection;
- fulfilling deductive inference, that is, finding possible values of goal attributes from the revealed regularities and partial description of the regarded object;
- demonstrating the chain of logical inference conducted by the system during computation of a logical conclusion.

7. CONCLUSIONS

The proposed common logical approach to solving problems of data mining and pattern recognition combines two powerful tools of modern logic: the inductive inference and the deductive inference. The first one is used for extracting knowledge from data. The second is applied when that knowledge is used for calculation of the goal attribute values.

In a simple case the data are given as a set of points in the space M of Boolean attributes. This set is considered as a small part of some larger set defining a subject area. In the suggested method, the knowledge is presented by implicative regularities, as are called equations $k = 0$ with elementary conjunctions k. The search for knowledge is reduced to looking for empty (not containing data points) intervals of M and putting forward corresponding hypotheses which suggest that the revealed intervals do not contain any other elements of the subject area. Such hypotheses can be accepted as implicative regularities if they are plausible enough. There have been proposed some formulas for plausibility evaluation. The accepted regularities are presented by ternary vectors constituting a knowledge matrix.

That form of knowledge representation is convenient for use in the second stage of the problem, when the values of goal attributes should be found from the partial information about the observed object. This task is reduced to solving a system of logical equations.

In the more complicated case of multi-valued attributes finite predicates are used instead of Boolean functions, and sectional Boolean vectors and matrices are suggested for their representation.

The proposed means were used when constructing several expert systems of various purposes where the pattern recognition problem was the central one. The computer experiments testified a high efficiency of the proposed approach.

REFERENCES

R. Agrawal, T. Imielinski and A. Swami. "Mining association rules between sets of items in large databases", *Proceedings of the 1993 ACM SIGMOD Conference*, Washington, DC, May 1993.

M. Bongard. *Pattern recognition*, Spartan Books, New York, 1970.

G. S. Boolos and R. C. Jeffrey. *Computability and logic*, Cambridge University Press, 1989.

C.-L. Chang and R. C. T. Lee. *Symbolic logic and mechanical theorem proving*, Academic Press, New York - San Francisco - London, 1973.

S. A. Cook. "The complexity of theorem proving procedures", *Proc. 3rd ACM symposium on the theory of computing*, ACM, pp. 151-158, 1971.

M. Davis, G. Longemann and D. Loveland. "A machine program for theorem proving", *Communications of the ACM*, v. 5, pp. 394-397, 1962.

W. J. Frawley, G. Piatetsky-Shapiro and C. J. Matheus. "Knowledge discovery in databases: an overview", *Knowledge discovery in data bases* (G. Piatetsky-Shapiro and W.J. Frawley, Editors), Cambridge, Mass: AAAI/MIT Press, pp. 1-27, 1991.

R. L. Graham and J. H. Spencer. "Ramsey theory", *Scientific American*, Vol. 263, No. 1, July 1990.

E. B. Hunt. *Artificial intelligence*, Academic Press, New York - San Francisco - London, 1975.

M. Karnaugh. "The map method for synthesis of combinatorial logic circuits", *Transactions AIEE, Communications and Electronics*, v. 72, pp. 593-599, 1953.

W. Klösgen. "Efficient discovery of interesting statements in databases", *The Journal of Intelligent Information Systems*, 4(1), pp. 53-69, 1995.

C. J. Matheus, P. Chan and G. Piatetsky-Shapiro. "Systems for knowledge discovery in databases", *IEEE Transactions on knowledge and data engineering*, Vol. 5, No. 6, pp. 903-913, 1993.

E. J. McCluskey, Jr. "Minimization of Boolean functions", *Bell System Technical Journal*, v. 35, No 6, pp. 1417-1444, 1956.

R. L. Nelson. "Simplest normal truth functions", *Journal of Symbolic Logic*, Vol. 20, No. 2, pp. 105-108, 1955.

N. J. Nilsson. *Problem-solving methods in artificial intelligence*, McGraw-Hill Book Company, New York, 1971.

G. Piatetsky-Shapiro. "Discovery, analysis, and presentation of strong rules", *Knowledge discovery in databases* (G. Piatetsky-Shapiro and W.J. Frawley, Editors), Menlo Park, Calif.: AAA Press, pp. 229-248, 1991.

D. A. Pospelov, ed. *Artificial intelligence. V. 2. Models and methods*, Moscow, Radio i swyaz (in Russian), 1990.

W. V. Quine. "The problem of simplifying of truth functions", *American Mathematical Monthly*, Vol. 59, No. 8, pp. 521-531, 1952.

A. Thayse, P. Gribomont, P. Gochet, E. Grégoire, E. Sanchez and P. Delsarte. *Approche logique de l'intelligence artificielle*, Dunod informatique, Bordas, Paris, 1988.

E. Triantaphyllou. "Inference of a minimum size Boolean function from examples by using a new efficient branch-and bound approach", *Journal of Global Optimization*, Vol. 5, No. 1. pp. 64-94, 1994.

R. Urbano and R. K. Mueller. "A topological method for the determination of the minimal forms of a Boolean function", *IRE Trans., EC-5*, No. 3, pp. 126-132, 1956.

D. Waterman. *A guide to expert systems*, Addison-Wesley Publishing Company, 1986.

A. D. Zakrevskij. "Visual-matrix method of minimization of Boolean functions", *Avtomatika i telemekhanika*, Vol. 21, No. 3, pp. 368-373 (in Russian), 1960.

A. D. Zakrevskij. "Algorithms for minimization of weakly specified Boolean functions", *Kibernetika*, No. 2, p. 53-60 (in Russian), 1965.

A. D. Zakrevskij. "Revealing of implicative regularities in the Boolean space of attributes and pattern recognition", *Kibernetika*, No. 1, pp. 1-6 (in Russian), 1982.

A. D. Zakrevsky "Implicative regularities in formal cognition models", *Abstracts of the 8-th International Congress of Logic, Methodology and Philosophy of Science*, Moscow, Vol. 1, pp. 373-375, 17-22 August 1987.

A. D. Zakrevskij. *Logic of recognition*, Minsk, Nauka i tekhnika (in Russian), 1988.

A. D. Zakrevskij, Yu. N. Pechersky and F. V. Frolov. *DIES - expert system for diagnostics of engineering objects*, Preprint of Inst. Math. AN Mold. SSR, Kishinev (in Russian), 1988.

A. D. Zakrevskij. "Matrix formalism for logical inference in finite predicates", *Philosophical bases of non-classical logics*, Institute of Philosophy AN SSSR, Moscow, pp. 70-80 (in Russian), 1990.

A. D. Zakrevskij, V. I. Levchenko and Yu. N. Pechersky. "Instrumental expert system EDIP, based on finite predicates", *Applied systems of artificial intelligence. Mathematical researches*, issue 123, Inst. Math. AN Mold. SSR, Kishinev, pp. 24-40 (in Russian), 1991.

A. D. Zakrevskij. "EXSYLOR - expert system for logical recognition", *Upravlyayushchie sistemy i mashiny*, No. 5/6, pp. 118-124 (in Russian), 1992.

A. D. Zakrevskij. "Logical recognition by deductive inference based on finite predicates", *Proceedings of the Second Electro-technical and Computer Science Conference ERK'93*, Slovenia Section IEEE, Ljubljana, Vol. B, pp. 197-200, 1993.

A. D. Zakrevskij. "Method of logical recognition in the space of multi-valued attributes", *Proceeding of the Third Electro-technical and Computer Science Conference ERK'94*, Slovenia Section IEEE, Ljubljana, Vol. B, pp. 3-5, 1994.

A. D. Zakrevskij. "Logical recognition in the space of multivalued attributes", *Computer Science Journal of Moldova*, Vol. 2, No. 2, pp. 169-184, 1994.

A. D. Zakrevskij and V. I. Romanov. "Implementation of the logical approach to the recognition of spoken words", *Upravljajushchie sistemy i mashiny*, No. 6, pp. 16-19 (in Russian), 1996.

A. D. Zakrevskij and I. V. Vasilkova. "Inductive inference in systems of logical recognition in case of partial data", *Proceedings of the Fourth International Conference on Pattern Recognition and Information Processing*, Minsk-Szczecin, Vol. 1, pp. 322-326, May 1997.

A. D. Zakrevskij and I. V. Vasilkova. "Minimization of logical inference chains in finite predicates", *Proceedings of the International Conference on Computer Data Analysis and Modeling*, Minsk, Vol. 2, pp. 161-166, 1998.

A. D. Zakrevskij. "Pattern recognition as solving logical equations", *Special Issue 1999 - SSIT'99 (AMSE)*, pp. 125-136, 1999.

A. D. Zakrevskij. "A logical approach to the pattern recognition problem", *Proceedings of the International Conference KDS-2001 "Knowledge - Dialog - Solution"*, Saint Petersburg, Russia, Vol. 1, pp. 238-245, June 2001.

H. Zhang. SATO: "An efficient propositional prover", *Proceedings of the International Conference on Automated Deduction*, pp. 272-275, July 1997.

AUTHOR'S BIOGRAPHICAL STATEMENT

Dr. Arkadij D. Zakrevskij graduated in 1956 from the Tomsk State University (Russia) with the BS degree in Radio-physics and Electronics. In 1960 he earned the Ph.D. degree also from at Tomsk State University. His dissertation title was "A matrix method for synthesis of digital systems". In 1967 earned the Dr.Sc. degree from the Institute of Automatics and Remote Control (Moscow). The title of that dissertation was "Programming language LYaPAS and automation of discrete automata synthesis".

During the years 1956-1971 he was a postgraduate student, Assistant, Associate and Full Professor at Tomsk State University - Siberian Physical-Technical Institute. He also served as the Head of the Chair of mathematical logic and programming. He taught courses on the theory of digital systems, programming, logical foundation of computers, mathematical logic, theory of probabilities, theory of Boolean functions, automata theory, and other related subjects.

Since 1971 he is the Head of the laboratory of logical design and a principal researcher at the Institute of Engineering Cybernetics (now the United Institute of Informatics Problems) of the National Academy of Sciences of Belarus. He is also a Professor at the Belarussian State University and at Belarussian State University of Informatics and Radioelectronics, Minsk, Belarus. He has taught courses on Boolean functions, graph theory, logical design of digital systems, discrete mathematics with application to CAD systems, and on the theory of logical control devices.

His research areas are on the logical theory of digital devices, theory of Boolean functions, programming automation, hard logical-combinatorial problems, logical recognition, solving logical equations, and parallel algorithms for logical control. He has published 11 books (see list below) in the above areas and also over 400 research papers. In 1972 Dr. Zakrevskij was elected a member of the National Academy of Science (NAS) of Belarus.