THE KNOWLEDGE: ITS PRESENTATION AND ROLE IN RECOGNITION SYSTEMS

A. D. Zakrevskij

Abstract: The concept of knowledge is the central one used when solving the various problems of data mining and pattern recognition in finite spaces of Boolean or multi-valued attributes. A special form of knowledge representation, called implicative regularities, is proposed for applying in two powerful tools of modern logic: the inductive inference and the deductive inference. The first one is used for extracting the knowledge from the data. The second is applied when the knowledge is used for calculation of the goal attribute values. A set of efficient algorithms was developed for that, dealing with Boolean functions and finite predicates represented by logical vectors and matrices.

INTRODUCTION

Knowledge is the central concept in a wide variety of investigations dedicated to pattern recognition problems [1, 3, 6]. Solving them begins with the choice of a proper *world model* – an abstract artificial world reflecting some important qualities of real subject areas – important from the point of view of the problems to be solved.

In this paper, we use a very simple model which defines this abstract world as a finite set of objects possessing some combinations of qualities. More formally, this model is defined as a set W of logical n-vectors presenting values of attributes composing the set $X = \{x_1, x_2, ..., x_n\}$. The attributes could be binary (Boolean) or multi-valued. In the latter case each of the attributes x_i is characterized by a corresponding finite set V_i of alternative values. The Cartesian product of these sets $V_1 \times V_2 \times ... \times V_n$ (or $\{0, 1\}^n$ in the binary case) constitutes the space of attributes M. Elements from W can be regarded as abstract models of real objects of a natural subject area. The world as a whole is represented by a subset $W \subseteq M$ or by the corresponding finite predicate $\varphi (x_1, x_2, ..., x_n)$ which takes value 1 on the elements of the set W. In case of two-valued attributes this predicate is a Boolean function $f (x_1, x_2, ..., x_n)$. That approach is rather simple, inasmuch the world is considered only as the set of its elements (without regarding any relations between them), but it is sufficient for solving many practical problems.

Usually, only partial information about the world W is known represented in terms of data and knowledge. Suppose that the *data* present information concerning some separate elements from W, describing these elements by corresponding logical vectors. Taken together, these vectors represent a so called sampling population (a reliable selection from the subject area) and constitute the set F serving further as a *database*. As a rule, $|W|<<|M|$ and $|F|<<|W|$. The *knowledge*, on the contrary, presents information about qualities of the whole subject area, expressed by some inherent in the regarded subject area regularities which establish some relationships between attributes.

Both data and knowledge present information about the regarded world, but they are of different type. Comparing them, we could say that data consist of affirmations about the *existence* of some objects with definite combinations of qualities. Unlike that, the knowledge puts some restrictions on what is possible, affirming the *non-existence* of objects with some other combinations of qualities.

The pattern recognition process, taken as a whole, may be roughly divided into two main stages: obtaining some knowledge by data mining and predicting values of goal attributes by using this knowledge. The methods of inductive and deductive inference are applied at these stages, accordingly. Their efficiency depends greatly on the form in which the knowledge is presented. A special attention is paid below to this point.

THE KNOWLEDGE – CONCEPT AND FORMAT

Within the framework of our world model, the knowledge is defined as a set of regularities. The key question is to choose a proper model for them. Starting from general assumptions it is accepted that any regarded regularity defines a logical connection between some attributes: it means that some combinations of attribute values are declared impossible (prohibited).

Evidently, the less attributes are connected by some regularity, the stronger is the latter. That is confirmed, for instance, by the long history of investigations in physics and other nature sciences. On the other hand, if we choose several attributes and decide to connect them by a regularity, it will be the weakest when it forbids only one combination of those attributes values.

In the Boolean case such a regularity can be expressed by the logical equation $k_i = 0$ or by the equivalent to it equation $d_i = 1$, where k_i is a conjunct formed of some attributes (in direct or inverse mode) from the set *X*, and d_i is the corresponding disjunct, satisfying the relation $d_i = \neg k_i$. For instance, equations $ab'c = 0$ and $a' \vee b \vee c' = 1$ (' is the symbol of inversion, equivalent to $\neg$) represent the same regularity, which prohibits the combination of values $a = 1$, $b = 0$, $c = 1$.

A regularity of this kind is called *implicative* (more general than functional one) [5]. 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. In other words, the regularity affirms that this interval is empty, not containing any elements of the world *W*. The size of that interval (the number of its elements) equals 2^{n-r}, where *n* is the number of all attributes and *r* is the rank of the implicative regularity – the number of attributes coming into it. It becomes clear now how the strength of the regularity is defined by its rank.

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. All these elements (α) are "between" two elements 100010 (minimal) and 101111 (maximal), satisfying the inequality $100010 \leq \alpha \leq 101111$, and hence justifying the term *interval*. 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 a *sequent* (its left part is always a conjunction, and the right part – 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 we could obtain the following set of the equivalent equations

$ae \rightarrow b$ (if $a = 1$ and $e = 1$, then $b = 1$), $ab' \rightarrow e'$, $a \rightarrow b \vee e'$, …, $1 \rightarrow a' \vee b \vee e'$.

The last one could be changed for the disjunctive equation $a' \vee b \vee e' = 1$, which may be coded by the ternary vector 01 - - 0 - .

A set of regularities given in such a form can be presented by a ternary disjunctive matrix *D*, called below a *knowledge matrix*, which columns are marked with attribute symbols. For example, the knowledge matrix

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
	1	-	-	0	-	-	0	-
D =	-	-	-	1	-	1	-	-
	0	1	-	-	-	-	-	-

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 values of some attributes:

$$(a = 0, d = 1, g = 1),\ (d = 0, f = 0) \text{ and } (a = 1, b = 0).$$

The set of these equations can be reduced to one equation $D = 1$ where *D* is a CNF (conjunctive normal form) represented by the matrix *D*.

$$D = (a \vee d' \vee g')\,(d \vee f)\,(a' \vee b) = 1.$$

By inverting both left and right parts of the equation $D = 1$ we get the equivalent equation $K = 0$ with the left part $K = \neg D$ – a DNF (disjunctive normal form) presenting a so called *veto function V*, which defines the prohibition area.

For the regarded example

$$K = a'dg \vee d'f' \vee ab' = 0.$$

The suggested form of implicative regularities turned out to be extremely convenient at the stage of deductive inference, where the methods developed for theorem proving automation are successfully applied [2]. As it is shown below, regularities of the considered type could be rather easily discovered in the database, and it is not difficult to evaluate their strength and plausibility, which is very important for their further application.

In the case of finite predicates generalized conjuncts and disjuncts could be used to present the knowledge [7, 8]. 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.

Suppose $X = \{x, y, z\}$, 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). Let $\alpha_1 = \{a\}$, $\alpha_2 = \{a, e, g\}$, $\alpha_3 = \{h, i\}$. The interval $I = \alpha_1 \times \alpha_2 \times \alpha_3$ 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

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

inasmuch as $(z = h) \vee (z = i) = 1$. If this product enters the equation $k = 0$ which reflects a regular connection between x and y, then $I \cap W = \varnothing$, i. e. the interval I turns out to be empty.

As it can be seen from the given 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 xi $\in V_i \backslash \alpha_i$. This is possible since all sets Vi are finite.

For the considered example

$$d = \neg k = (x \neq a) \vee ((y \neq a) \wedge (y \neq e) \wedge (y \neq g)) = (x = b) \vee (x = c) \vee (y = f).$$

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

In the case of multi-valued attributes, it is more convenient to use sectional Boolean vectors and matrices introduced for representation of finite predicates [7]. 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. For the example given above, the 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*. Obviously, 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 some fuzziness. The 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 quanta* and is marked by a zero in the corresponding component of the vector.

Giving an example of presenting the knowledge, suppose 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

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

0 0 1 . 0 1 0 0 . 0 1

may be interpreted as a set of disjunctive equations

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

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

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

DATA MINING

A very important part of the pattern recognition problem is obtaining knowledge from data [3]. 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 the set *F* in the space *M* (suppose it is Boolean) 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 *K* and looking for such combinations of attribute values which do not occur there.

The number of attributes coming into an implicative regularity is called its *rank*. It coincides with the rank of the corresponding interval. Remind that the less attributes are tied with a regularity, the stronger is the tie. So, it is worthwhile to look for regularities of smaller rank.

Consider, for example, the following data matrix *K*:

a	*b*	*c*	*d*	*e*	*f*
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

There are no empty intervals of the rank 1, because each column contains 1s and 0s. So we look further for empty intervals of the 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 the 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 make 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, while 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 express 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 if only this

probability is small enough. Its calculation is rather difficult, so it was proposed in [5] to approximate it by the mathematical expectation $E(n, m, r)$ of the number of empty intervals of the rank r.
That value can be calculated by the formula

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

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 the rank r in the space M, and $(1-2^{-r})^m$ is the probability of some concrete interval of the rank r to be 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 the Table 1 of the dependence E on r under fixed values of other parameters: $n = 100$ and $m = 200$.

Table 1. The dependence 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 the search for empty intervals and putting forward corresponding hypotheses can be restricted in this case by the relation $r < 4$. If some empty interval of the rank $r < 4$ is found, we can formulate the corresponding regularity with good reason, 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 the ranks more then 3. The search for regularities could be strongly restricted in that case by checking for emptiness only intervals of the ranks 1, 2 and 3, which number is

$$C_{100}^1\times2^1 + C_{100}^2\times2^2 + C_{100}^3\times2^3 = 1{,}333{,}400.$$

Not much, compared with 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 the 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 dependence of the maximum rank r_{max} on parameters n and m

$n \setminus m$	20	50	100	200	500	1000
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 – 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 enough 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. It analyzes the elements of the selection F consecutively, one by one, and fix such two-element combinations which have occur, using a symmetrical square Boolean $2n\times2n$ matrix S for that, with rows and columns corresponding to the values $x_1 = 0$, $x_1 = 1$, $x_2 = 0$, $x_2 = 1$, etc. Its elements corresponding to occurring combinations are marked with 1. The rest combinations (not occurring) are presented by zero (empty) elements and accepted as forbidden. The regularities presented by them connect some attributes in pairs and are called syllogistic [6]. For example, regarding the following selection F (only to illustrate the algorithm, despite the fact that the selection is too small for $r_{max} = 2$):

a	b	c	d	e	
0	1	0	0	1	1
1	1	0	1	1	2
1	0	0	1	1	3
0	1	1	0	0	4
1	0	0	1	1	5
0	1	1	0	0	6

we shall find in the end ten two-element combinations which do not occur in *F*, and consider them as syllogistic regularities. They can be presented by the following ternary knowledge matrix *D*:

D =

a	b	c	d	e
0	0	-	-	-
0	-	-	1	-
1	-	1	-	-
1	-	-	-	0
-	0	1	-	-
-	0	-	0	-
-	0	-	-	0
-	-	1	1	-
-	-	1	-	1
-	-	-	1	0

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.

It is a little more difficult to extract knowledge from the space of *n* multi-valued attributes $x_1, x_2, \ldots, x_n$ [9, 11]. To begin with, define the probability *p* that some concrete disjunct will be satisfied by an accidentally chosen element of the space. It could be calculated by the formula

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

where s_i is the number of all values of the attribute x_i, and r_i – the number of those of them which do not enter this disjunct. For instance, $p = 1 - 2/2\times3/4\times1/3 = 3/4$ for the disjunct 00.1000.101. Let us divide all disjuncts into classes D_j, forming them from disjuncts with the same value of *p*. And let us number these classes in order of increasing *p* and introduce the following conventional signs: q_j – the number of disjuncts in the class D_j, p_j – 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:

$$E_j = q_j\,(p_j)^m ,$$

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

$$E_k^+ = \sum_{l=1}^{k} E_j .$$

Inductive inference is performed by consecutive regarding classes D_j in order of their numbers and summarizing corresponding values E_j until the sum surpasses a threshold ω, which is introduced with taking into account the specific of the problems to be solved. 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*.

The 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*. 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 little number of logical conclusions, but more reliable ones, while extending the used part of the knowledge base extends also the set of obtained logical conclusions, at the expense of their plausibility.

We do not regard here the important problem (touched in [12]) of extracting knowledge from partial data – when values of some attributes of the elements from *F* remain unknown.

SOLVING EQUATIONS OF DEDUCTIVE INFERENCE

The recognition problem can be regarded as the problem of a closer definition of qualities of some observed object not belonging to the experimental selection from the subject area [5, 14]. It is formulated in terms of logical equations, Boolean or predicate, and the tree searching technique of deductive inference is applied for their solution [4, 10, 13].

Suppose, we know the values of *s* from *n* attributes of this object. That is equivalent to location of the object in a certain interval of the Boolean space *M* presented by the corresponding elementary conjunction *k* of the rank *s*. The problem is to define by logical reasoning, as sure as possible, the values of the remaining $n - s$ attributes, using for that the information contained in the knowledge ternary matrix *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 the 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 3 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 3. 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}	∅

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 last 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 of attribute *e*, but there are no restrictions on other attributes (*c* and *d*). If by the same function *V* the object α is characterized by $k = c'e'f$, then

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

and that means that the object contradicts the knowledge.

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

Suppose, we regard the same (already minimized) knowledge matrix 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 the matrix D. 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, using the following rule: $x\,(x' \vee H) = x\,H$, where x is a Boolean variable and H – an arbitrary Boolean formula.

	a b c d e f			b d e f		b d e f
	0 - 0 - - 1	1		- - - 1		- - - 1
D^* =	- 0 - - 1 0	2	→	0 - 1 0	→	0 - 1 -
	1 - - 1 0 -	3		1 0 - 0		1 0 - -
	- 1 - 0 - 0	4				
	- 1 1 1 - -	5				

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).

Solving the recognition problem in a special case, when the values of all attributes are known except the goal one, could be facilitated by preliminary partitioning the Boolean space of attributes into four regions. After that it would be sufficient only to find out 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 3. The operations $f : x_i$ and $f : x_i'$ changing the argument x_i of the function f for constant 1 or 0, accordingly, are used by that. 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 $V^*(x_i) = (V : x_i) \wedge (V : x_i')$.

Using the same example we obtain:

$$V = acf' \vee be'f \vee a'd'e \vee b'df \vee b'c'd',$$
$$V : f = be' \vee a'd'e \vee b'd \vee b'c'd',$$
$$V : f' = ac \vee a'd'e \vee b'c'd',$$
$$V^-(f) = (be' \vee a'd'e \vee b'd \vee b'c'd')' \wedge (ac \vee a'd'e \vee b'c'd')',$$
$$V^1(f) = (be' \vee a'd'e \vee b'd \vee b'c'd')' \wedge (ac \vee a'd'e \vee b'c'd'),$$
$$V^0(f) = (be' \vee a'd'e \vee b'd \vee b'c'd') \wedge (ac \vee a'd'e \vee b'c'd')',$$
$$V^*(f) = (be' \vee a'd'e \vee b'd \vee b'c'd') \wedge (ac \vee a'd'e \vee b'c'd').$$

DEDUCTIVE INFERENCE IN FINITE PREDICATES

In the case of multi-valued attributes the disjunctive knowledge matrix D turns out to be a sectional Boolean matrix presenting some finite predicate. There is some specific in dealing with it [7, 10].

Let us state the central problem of deductive inference: a disjunctive matrix D and a disjunct d mated with D (that means defined on the same pattern) are given, the problem is to find out whether d is a logical consequence of D. In other words, the question is if the conjunctive term d is derived from CNF D, that means it becomes 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 d is obtained or the set expansion is exhausted without finding d. At the last case it is proved that d does not follow from D.

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

But far not all disjuncts obtained in such a way deserve 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 u and v but does not follow from u or v taken separately. Let us call it a *resolvent* of disjuncts u and v, and determine the rules for its obtaining.

Disjuncts u and v are called *adjacent in regard to 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 rest domains a component with the value 0 in both vectors. Note that at violating the first condition a disjunct is obtained which follows either from u or from v, whereas at violating the second condition a trivial (identical to 1) disjunct is found, which follows from any other disjunct.

Affirmation 1. If disjuncts u and v are adjacent in regard to the variable x_i and $w = u<x_i>v$, then the disjunct w is a resolvent of the disjuncts u and v.

For example,

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

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

$$\begin{array}{rl} u<a>v = & 0\,0\,0\,.\,1\,0\,.\,0\,1\,1\,1 \\ u<c>v = & 1\,1\,0\,.\,1\,0\,.\,0\,0\,1\,0 \end{array}$$

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 d does not follow from D. So, the problem is reduced to the regarded above problem of checking some disjunctive matrix for consistency.

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

Affirmation 2. A disjunct d follows from a disjunctive matrix D if and only if the disjunctive matrix $D \wedge \neg d$ is not consistent.

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

For instance, if

$$D = \begin{array}{l} 0\,0\,1\,.\,0\,0\,1\,0\,.\,0\,0 \\ 1\,1\,0\,.\,0\,0\,1\,1\,.\,0\,1 \\ 0\,1\,0\,.\,1\,1\,0\,0\,.\,1\,0 \\ 0\,0\,1\,.\,0\,1\,0\,0\,.\,0\,0 \end{array},$$

$$d = 011.1000.00\ ,$$

then the following disjunctive matrix should be checked for consistency

$$D \wedge \neg d = \begin{matrix} 000.0010.00 \\ 100.0011.01 \\ 000.0100.10 \\ 000.0100.01 \end{matrix}$$

This matrix is not consistent, hence the disjunct d follows from D.

CONCLUSION

Implicative regularities were proposed to fix the knowledge extracted from data on the stage of inductive inference and to play the role of axioms on the stage of deductive inference, when the values of goal attributes should be forecasted. A set of algorithms was developed to implement those operations. The suggested means were used when constructing several expert systems of various purposes where the pattern recognition problem was the central one. The computer experiments testified the high efficiency of the proposed approach.

Bibliography

1. Bongard M. Pattern recognition. – Spartan Books, New York, 1970.
2. Chang C. L., Lee R. C. T. Symbolic logic and mechanical theorem proving. – Academic Press, New York – San Francisco – London, 1973.
3. Frawley W. J., Piatetsky-Shapiro G., and Matheus C. J. Knowledge discovery in data bases: an overview. – In: Knowledge discovery in data bases (ed. by Piatetsky-Shapiro and Frawley), Cambridge, Mass: AAAI/MIT Press, 1991, pp. 1-27.
4. Nilsson N. J. Problem-solving methods in artificial intelligence. – McGraw-Hill Book Company, New York, 1971.
5. Zakrevskij A. D. Revealing of implicative regularities in the Boolean space of attributes and pattern recognition. – Kibernetika, No 1, 1982, pp. 1-6 (in Russian).
6. Zakrevskij A. D. Logic of recognition. – Minsk: Nauka i tekhnika, 1988 (in Russian).
7. Zakrevskij A. D. Matrix formalism for logical inference in finite predicates. – Philosophical bases of non-classical logics. Institute of Philosophy AN SSSR, Moscow, 1990, pp. 70-80 (in Russian).
8. Zakrevskij A. D., Levchenko V. I., Pechersky Yu. N. Instrumental expert system EDIP, based on finite predicates. – Applied systems of artificial intelligence. Mathematical researches, issue 123, Inst. Math. AN Mold. SSR, Kishinev, 1991, pp. 24-40 (in Russian).
9. Zakrevskij A. D. EXSYLOR – expert system for logical recognition. – Upravlyayushchie sistemy i mashiny, 1992, No 5/6, pp. 118-124 (in Russian).
10. Zakrevsky A. D. 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, 1993, v. B, pp. 197-200.
11. Zakrevsky A. Logical recognition in the space of multivalued attributes. – Computer Science Journal of Moldova, 1994, v. 2, No 2, pp. 169-184.
12. Zakrevskij A. D., Vasylkova I. V. 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, May 1997, v. 1, pp. 322-326.
13. Zakrevskij A. D. Pattern recognition as solving logical equations. – Special Issue 1999 – SSIT'99 (AMSE), pp. 125-136.
14. Zakrevskij A. D. A logical approach to the pattern recognition problem. – Proceedings of the International Conference KDS-2001 "Knowledge – Dialog – Solution", S.- Peterburg, June 2001, v. 1, pp. 238-245.

Author information

Arkadij Zakrevskij - The United institute of informatics problems of the NAS of Belarus. Surganov Str. 6, 220012 Minsk, Belarus. e-mail: zakr@newman.bas-net.by