

Arkady D. Zakrevskij

## FINITE PREDICATES WITH APPLICATIONS IN PATTERN RECOGNITION PROBLEMS

We extend the theory of Boolean functions, especially in respect to representing these functions in the disjunctive or conjunctive normal forms, onto the case of finite predicates. So, we show that it is useful to apply the language of Boolean vectors and matrices, developing efficient methods for calculation over finite predicates. This means that finite predicates should be decomposed into some binary units, which will correspond to components of Boolean vectors and matrices and should be represented as combinations of these units. Further, we define probabilities in data bases using Boolean matrices representing finite predicates. We also show that it is natural to try and present knowledge in the most compact form, which allows reducing the time of inference, by which the recognition problems are solved.

*Corresponding Member of the National Academy of Sciences of Belarus,  
United Institute of Informatics Problems, National Academy of Sciences of Belarus,  
Minsk, Belarus  
e-mail: zakrevskij@tut.by*

### 1. Finite predicates and their matrix forms

One of the most important problems of artificial intelligence is the problem of pattern recognition [2, 4]. To solve it, various formal methods were applied, usually based on the theory of Boolean functions [7, 9]. However, they become insufficient when dealing with objects described in terms of multi-valued attributes, so other means should be involved in this case, finite predicates for example [10].

The finite predicates are two-valued functions, which arguments are variables with restricted number of values. Denote these variables by  $x_1, x_2, \dots, x_n$ . Let them receive values accordingly from finite sets  $X_1, \dots, X_n$ , which direct product  $X_1 \times X_2 \times \dots \times X_n$  generates a space  $M$ . The mapping  $M \rightarrow \{0, 1\}$  of the set  $M$  onto the two-element set  $\{0, 1\}$  (it is equivalent to {false, true}) is called a *finite* predicate.

If solving practical problems is related to the usage of finite predica-

tes, it is useful to represent the latter whenever possible in a more compact form. Here it is possible to use experience of the theory of Boolean functions, developed chiefly for the case when the considered functions are represented in the disjunctive normal form (DNF). The most effective methods of minimization of Boolean functions and solution of logical equations are designed just for this form. It is reasonable to extend these methods onto finite predicates.

According to tradition, let us assume that an *elementary conjunction*  $k$  represents the characteristic function of some interval  $I$  of space  $M$ , and this interval is defined as the direct product of nonempty subsets  $\alpha_i$ , taken by one from every  $X_i$ :

$$I = \alpha_1 \times \alpha_2 \times \dots \times \alpha_n, \quad \alpha_i \subseteq X_i, \quad \alpha_i \neq \emptyset, \quad i = 1, 2, \dots, n.$$

This means that an elementary conjunction  $k$  is defined as a conjunction of several one-argument predicates  $x_i \in \alpha_i$  ( $x_i$  receives a value from subset  $\alpha_i$ ) and is represented by the expression

$$k = (x_1 \in \alpha_1) \wedge (x_2 \in \alpha_2) \wedge \dots \wedge (x_n \in \alpha_n).$$

The multiplicands, for which  $\alpha_i = X_i$  (in this case predicate  $x_i \in \alpha_i$  becomes identical to true), can be dropped.

Note that in the simplest case, when all arguments become two-valued, this definition coincides with the definition of elementary conjunction in Boolean algebra.

Similarly, we shall define an *elementary disjunction*  $d$  as a disjunction of one-argument predicates distinct from true:

$$d = (x_1 \in \alpha_1) \vee (x_2 \in \alpha_2) \vee \dots \vee (x_n \in \alpha_n), \quad \alpha_i \subset X_i, \quad i = 1, 2, \dots, n.$$

If  $\alpha_i = \emptyset$ , the term  $x_i \in \alpha_i$  can be deleted from any elementary disjunction, as representing the identically false expression.

The disjunctive and conjunctive normal forms are defined by the standard way: *DNF* is a disjunction of elementary conjunctions, and *CNF* is a conjunction of elementary disjunctions.

The characteristic functions of elements of space  $M$  are naturally represented as *complete* elementary conjunctions, i.e. elementary conjunctions, in which all sets are one-element:  $|\alpha_i| = 1$  for all  $i = 1, 2, \dots, n$ . Any DNF, composed of complete elementary conjunctions, is called *perfect (PDNF)*. The number of its terms is equal to the power of characteristic set  $M_\varphi$  of predicate  $\varphi$ , represented by the given PDNF.

Developing efficient methods for calculation over finite predicates, it is useful to apply the language of Boolean vectors and matrices, immediately

representable in computer. And it means that all considered objects should be decomposed into some binary units, which will correspond to components of Boolean vectors and matrices and should be represented as combinations of these units.

For representation of such combinations we shall use *sectional Boolean vectors*. They are divided into sections set in one-to-one correspondence with arguments, and the components of these sections are put in correspondence with values of the arguments. Value 1 in component  $j$  of section  $i$  is interpreted as the expression “variable  $x_i$  has value  $j$ ”. The sectional Boolean vectors shall be used for representation of elements and some areas of space  $M$ , and collections of such vectors for representation of finite predicates.

Elements of space  $M$ , i.e. some concrete sets of values of all arguments, shall be represented by sectional vectors having exactly one 1 in each section, defining in such a way uniquely values accepted by the arguments. The sectional vectors of more general type, which could contain several 1s in each section, have double interpretation. Firstly, they can be understood as elementary conjunctions (conjunctions of one-argument predicates corresponding to intervals of space  $M$ , i.e. direct products of nonempty subsets taken by one from  $X_1, X_2, \dots, X_n$ ). Secondly, they can be interpreted as similarly defined elementary disjunctions, which can be regarded as the complements of appropriate elementary conjunctions. Let us call such vectors *conjuncts* and *disjuncts*, respectively. Each section of a conjunct should contain no less than one 1, each section of a disjunct no less than one 0 (otherwise the conjunct degenerates to 0, the disjunct to 1).

The correspondence between elements of sectional vectors, on the one hand, and arguments and their values, on the other hand, is set by a *cliché* – the linear enumeration of arguments and their values. Let us assume that in the considered below examples all vectors are interpreted on a uniform cliché, for example, as follows:

$$\begin{array}{cccc} & a & & b & & c \\ 1 & 2 & 3 & . & 1 & 2 & 3 & 4 & . & 1 & 2 \end{array}$$

Thus, if it is known that vector

$$1 \ 1 \ 0 \ . \ 0 \ 1 \ 0 \ 1 \ . \ 0 \ 1$$

represents a conjunct, it is interpreted as a predicate receiving value 1 when

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

and if this vector is regarded as a disjunct, it is interpreted as a predicate accepting value 1 if and only if

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

Collections of sectional vector-rows can form *sectional Boolean matrices* of two types: conjunctive and disjunctive ones. *Conjunctive matrices* consist of row-conjuncts and are convenient for interpreting as disjunctive normal forms (DNFs) of finite predicates. *Disjunctive matrices* consist of row-disjuncts and are interpreted as conjunctive normal forms (CNFs).

## 2. Representation of data and knowledge

The main concepts used by solving pattern recognition problems are *world model*, *data* and *knowledge*.

The world model is defined as a set  $W$  (called *world* below) of some objects represented by combinations of values of their attributes, which compose the set  $X = \{x_1, x_2, \dots, x_n\}$ . The attributes could be multi-valued, for example, such as the colour, which can be red, dark blue, green, etc., but should receive only one of these values. The world  $W$  is regarded as a subset of space  $M$  and is presented by the corresponding predicate  $\varphi$ . Usually  $|W| \ll |M|$ .

It is natural to define the data as any information about individual objects, and the knowledge about world  $W$  as a whole [9, 12]. 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. These relationships prohibit some other combinations of properties ( $Q$ ) by equations  $k_i = 0$ , where  $k_i$  is a conjunction over the set of attributes  $X$ , or by equivalent to them equations  $d_i = 1$  called *disjuncts* below (with elementary disjunction  $d_i = \bar{k}_i$ ). 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. In case when these prohibitions are represented by disjuncts they are called *implicative regularities* [8].

Reflecting availability of the mentioned combinations by the predicates  $P$  and  $Q$ , one can present the data by affirmations  $\exists w \in W : P(w)$  with the existential quantifier  $\exists$  (there exists), and the knowledge by affirmations  $\neg \exists w \in W : Q(w)$  with its negation  $\neg \exists$  (there does not exist). The latter

ones could be easily transformed into affirmations  $\forall w \in W : \neg Q(w)$  with the generality quantifier  $\forall$  (for every).

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 objects from some world  $W$  could be described in such a way but only a relatively small part of them which forms a random selection  $F$  from  $W$ :  $|F| \ll |W|$ . Selection  $F$  can be represented by a set of selected points in space  $M$ .

The distribution of these points reflects the regularities inherent in the world: every implicative regularity generates some empty, i.e. free of selected points, interval in the space  $M$ . The reverse affirmation suggests itself: maybe any empty interval generates the corresponding regularity. But such an affirmation is a hypothesis which could be accepted if only it is plausible enough. The matter is that an empty interval 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. Thus the problem of plausibility evaluation arises which should be solved on the stage of inductive inference, where some regularities are extracted from the data.

### **3. Inductive inference**

A lot of papers are devoted to the problem of knowledge discovery in data bases [1, 3, 6, etc.]. Inductive inference is used for its solution.

In our case it consists in suggesting hypotheses about regularities represented by those disjuncts, which do not contradict the data. However, these hypotheses could be accepted if only they are reliable enough, and this means that at least these disjuncts should correspond to rather big intervals of space  $M$ .

Consider some disjunct. It does not contradict the database if the corresponding interval of space  $M$  is empty – if it does not intersect with the random selection  $F$  from  $W$ . Therefore, it is possible to put forward a hypothesis affirming that the whole world  $W$  as well does not contain elements of that interval. However, it is necessary to take into account the possibility that the considered interval has appeared empty quite accidentally. The less is the probability of such possibility, the more reasonable would be to accept the hypothesis.

The formula for evaluation of such a probability is rather complicated. But it can be approximated by the mathematical expectation  $w$  of the num-

ber of empty intervals of the given size, and the less is that value, the more precise is the approximation.

That expectation  $w$  was evaluated in [10] for the case of two-valued attributes, as a function of parameters:

$m$  is the number of elements in the random selection  $F$ ,

$n$  is the number of binary attributes,

$k$  is the rank of the regarded disjunct (the number of variables in it), determining the size of considered intervals.

The following formula was proposed to calculate it:

$$w(m, n, k) = C_n^k 2^k (1 - 2^{-k})^m,$$

where  $C_n^k$  is the number of different  $k$ -element subsets of an  $n$ -element set.

In order to evaluate the indicated probability for the case of many-valued attributes, we shall carry out the following imaginary experiment. Suppose that the selection  $F$  is formed during  $m$  steps, on each of which one element is selected from space  $M$  at random.

Considering a disjunct, we shall count up the probability  $p$  that it will be satisfied with an accidentally selected element of space  $M$  (this element will not enter the corresponding interval):

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

where  $n$  denotes the number of attributes,  $s_i$  – the number of values of attribute  $x_i$ ,  $r_i$  is the number of those of them, which do not enter the disjunct. For example, the probability  $p$  for disjunct 00.1000.101 is equal to  $1 - 2/2 \cdot 3/4 \cdot 1/3 = 3/4$ .

Let's divide all conceivable disjuncts into classes  $D_i$ , consisting of disjuncts with equal values of  $p$ , number these classes in ascending order  $p$  and introduce the following characteristics:

$q_i$  is the number of disjuncts in class  $D_i$ ,

$p_i$  is the value of parameter  $p$  for elements of class  $D_i$ .

The expectation  $w_i$  of the number of disjuncts from class  $D_i$ , which do not contradict the considered random selection, is

$$w_i = q_i p_i^m,$$

and the similar expectation for the union of classes  $D_1, D_2, \dots, D_k$  is

$$w_k^+ = \sum_{i=1}^k w_i.$$

Just this value can be used for the quantitative estimation of hypotheses plausibility. Any disjunct not contradicting to the data can be accepted as a regularity only when this value is small enough. In this case it is impossible to explain the emptiness of the corresponding interval by an accident; hence we have to admit that the disjunct represents some regularity reflected in the database.

#### 4. Knowledge base and its simplification

After extracting regularities from a database, a knowledge base is created playing the main role during recognition of new objects of the researched subject area. It is natural to try and present knowledge in the most compact form, which will allow reducing the time of inference, by which the recognition problems are solved.

The knowledge base is created as a disjunctive matrix  $D$ , representing CNF of some finite predicate. Therefore its compression is performed as minimization of this finite predicate. Minimizing a predicate we obtain its most compact description. Usually that means finding its *shortest* DNF, which contains a minimum number of terms. This task can be formulated as the task of finding a shortest minor cover of a Boolean matrix.

Let  $u$  and  $v$  be some rows of a disjunctive matrix  $D$ , and  $p$  and  $q$  – some of its columns. Let's assume, that vectors  $a$  and  $b$  are in ratio  $a \geq b$  if this ratio is fulfilled component-wise (for example,  $011.0010.101 \geq 010.0010.100$ ).

The following rules of reduction allow simplifying a disjunctive matrix  $D$  by deleting some rows or columns.

*Rule 1.* If  $u \geq v$ , row  $u$  is deleted.

*Rule 2.* If row  $u$  contains complete (without zeros) domain (section), it is deleted.

*Rule 3.* If column  $p$  is empty (without ones), it is deleted.

*Rule 4.* If a row exists containing ones only in one domain, all columns of that domain which contain zeros in the given row are deleted.

The enumerated rules form a set of basic equivalence transformations of the disjunctive matrix (not changing the represented predicate). Alongside with the given rules one more transformation can be applied for simplification of matrix  $D$ . Its use can change the set of solutions, but does not disturb the property of consistency: any consistent matrix remains consistent, any inconsistent – remains inconsistent.

*Rule 5.* If  $p \geq q$  and the columns  $p$  and  $q$  belong to the same domain, the column  $q$  is deleted.

## 5. Resolution rules

Let  $\mathbf{u}$  and  $\mathbf{v}$  be some disjuncts,  $\mathbf{D}$  and  $\mathbf{C}$  be disjunctive matrices, specifying some CNFs, and  $E(\mathbf{u})$ ,  $E(\mathbf{v})$ ,  $E(\mathbf{D})$ ,  $E(\mathbf{C})$  be their characteristic sets, i.e. collections of elements of space  $M$ , presenting the solutions for  $\mathbf{u}$ ,  $\mathbf{v}$ ,  $\mathbf{D}$  and  $\mathbf{C}$ , accordingly. Besides, let  $\bar{\mathbf{u}}$  be the vector obtained from  $\mathbf{u}$  by its component-wise negation, and  $\mathbf{D} \wedge \mathbf{u}$  be the matrix obtained from  $\mathbf{D}$  by component-wise conjunction of its each row with vector  $\mathbf{u}$ .

Let us say that disjunct  $\mathbf{v}$  follows from disjunct  $\mathbf{u}$  (it is its logical conclusion), denoting it as  $\mathbf{u} \rightarrow \mathbf{v}$ , if and only if  $E(\mathbf{u}) \subseteq E(\mathbf{v})$ . Similarly,  $\mathbf{D} \rightarrow \mathbf{u}$  if and only if  $E(\mathbf{D}) \subseteq E(\mathbf{u})$ ,  $\mathbf{D} \rightarrow \mathbf{C}$  if and only if  $E(\mathbf{D}) \subseteq E(\mathbf{C})$ , etc.

It is easy to show that  $\mathbf{u} \rightarrow \mathbf{v}$  if and only if vector  $\mathbf{v}$  covers vector  $\mathbf{u}$ .

The following problem is formulated in the mode typical for the logic inference theory. A disjunctive matrix  $\mathbf{D}$  and a disjunct  $\mathbf{u}$  are given. The question is to find out, whether  $\mathbf{u}$  follows from  $\mathbf{D}$ .

### Affirmation 1

Disjunct  $\mathbf{u}$  logically follows from disjunctive matrix  $\mathbf{D}$  if and only if matrix  $\mathbf{D} \wedge \bar{\mathbf{u}}$  is inconsistent.

The procedure of checking CNF for consistency is useful for conversion of a disjunctive matrix to an irredundant form, which could be sometimes a good approximation to the optimum solution.

A disjunctive matrix is called *irredundant* when at deleting of any row or at changing value 1 of some element for 0 it turns to a matrix not equivalent to the initial one. One can make any disjunctive matrix irredundant by applying operations of these two types while it is possible, i.e. while after their execution the matrix remains equivalent to the initial one.

It is obvious that a row can be deleted from matrix  $\mathbf{D}$  if it is a logical conclusion of the remaining set. And the check of this condition is circumscribed above.

Sometimes a row cannot be deleted, but it is possible to change value 1 of some of its component for 0, having reduced by that the number of 1s in the matrix.

### Affirmation 2

Element  $d_i^{jk}$  of disjunctive matrix  $\mathbf{D}$  can change its value 1 for 0 if and only if a disjunct follows from  $\mathbf{D}$ , which can be obtained from row  $\mathbf{d}_i$  by replacement of domain  $\mathbf{d}_i^j$  by other one, where  $d_i^{jk} = 0$  and the remaining components have value 1.

## 6. Deductive inference in pattern recognition

Consider now the disjunctive matrix  $D$  as a system of regularities, which are obligatory for all elements of a subject area (class), formally identified with some sets of values of attributes, i.e. with elements of the space  $M$ . Thus we shall consider every disjunct representing a particular tie between attributes bounding the set of “admittable” objects.

Let us assume that regarding an object from the researched class we receive the information about values of some attributes. It is convenient to define as a *quantum* of such an information the elementary prohibition  $x_j \neq k$ : the value of attribute  $x_j$  is distinct from  $k$ . Having received several such quanta, we can present the total information by a sectional Boolean vector  $r$ , in which the components corresponding to elementary prohibitions, take value 0, the remaining – value 1. This vector sets some elementary conjunction  $r$  and is interpreted as the conjunctive equation  $r = 1$ . Let us call it a *conjunct*.

For example, conjunct

$$r = 1\ 1\ 1\ .\ 0\ 0\ 1\ 1\ .\ 0\ 1$$

is interpreted as equation

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

This means that the considered object cannot have value 1 or 2 of attribute  $b$ , and also value 1 of attribute  $c$ . In other words, vector  $r$  sets an interval where the object is localized, it is known only that the element of space  $M$  representing this object is somewhere inside the indicated interval.

A problem of recognition arises in this situation, consisting in further localization of the object by the way of deductive inference [12, 13, 15]. The information contained in matrix  $D$  is used for that. It represents a system of disjunctive equations to which the objects of the given class should be submitted [11].

The best solution of this problem could be achieved via simplifying of this system by its “tuning” onto the interval represented by vector  $r$ . This operation is performed by deleting values 1 in columns of matrix  $D$ , corresponding to those components of conjunct  $r$ , which have value 0.

### Affirmation 3

A disjunctive matrix  $D$  in aggregate with a conjunct  $r$  is equivalent to the disjunctive matrix  $D^* = D \wedge r$ .

The operation of deleting 1s in some columns could be followed by further reducing the disjunctive matrix by means of standard conversions of equivalence.

The “interval” localization of the object is of interest at recognition, when some more components of vector  $\mathbf{r}$  can change value 1 for 0. Such a localization well corresponds to the traditional formulation of the problem of recognition, when the values of some selected (target) attributes are searched. The process of such localization could be reduced to search of separate elementary prohibitions, when questions of the following type are put forward: whether it follows from matrix  $\mathbf{D}^*$ , that the considered object cannot have value  $k$  of attribute  $x_j$ ?

Obviously, at the positive answer to this question a disjunct follows logically from matrix  $\mathbf{D}^*$ , represented by the sectional Boolean vector  $\mathbf{s}(j, k)$ , in which all components of domain  $j$ , except number  $k$ , have value 1, and all rest components have value 0.

#### Affirmation 4

The value of component  $r^{jk}$  of vector  $\mathbf{r}$  can be changed from 1 to 0 if and only if the disjunctive matrix  $\mathbf{D}^* \wedge \bar{\mathbf{s}}(j, k)$  is inconsistent.

Regard an example with variables  $a, b, c$ , receiving values accordingly from sets  $A = \{1, 2, 3\}$ ,  $B = \{1, 2, 3, 4\}$ ,  $C = \{1, 2\}$ . Let

$$\mathbf{D} = \begin{bmatrix} 001.0010.00 \\ 000.0011.01 \\ 010.1100.10 \\ 001.0000.01 \end{bmatrix}$$

and suppose it is known that some object of the considered class has value 1 of attribute  $c$ . Then

$$\mathbf{r} = [111.1111.10]$$

and

$$\mathbf{D}^* = \begin{bmatrix} 001.0010.00 \\ 000.0011.00 \\ 010.1100.10 \\ 001.0000.00 \end{bmatrix}.$$

If we are interested in attribute  $b$ , it is possible at once to initiate check of its values and to find out, for example, that  $b$  cannot have value 1, because

$$\mathbf{s}(b, 1) = 0\ 0\ 0\ .\ 0\ 1\ 1\ 1\ .\ 0\ 0,$$

and matrix  $D^* \wedge \bar{s}(b, 1)$  takes value

$$\begin{bmatrix} 001.0000.00 \\ 000.0000.00 \\ 010.1000.10 \\ 001.0000.00 \end{bmatrix}.$$

It is obvious that it is inconsistent, because there is a row containing only zeros.

## References

- [1] 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.
- [2] M. Bongard. *Pattern recognition*, Spartan Books, New York, 1970.
- [3] W. J. Frawley, G. Piatetsky-Shapiro and C. J. Matheus. “Knowledge discovery in data bases: an overview”, *Knowledge discovery in data bases* (ed. by Piatetsky-Shapiro and Frawley), Cambridge, Mass: AAAI/MIT Press, pp. 1–27, 1991.
- [4] E. B. Hunt. *Artificial intelligence*, Academic Press, New York – San Francisco – London, 1975.
- [5] N. J. Nilsson. *Problem-solving methods in artificial intelligence*, McGraw-Hill Book Company, New York, 1971.
- [6] G. Piatetsky-Shapiro. “Discovery, analysis, and presentation of strong rules”, *Knowledge discovery in databases* (eds. G. Piatetsky-Shapiro and W. Frawley), Mento Park, Calif.: AAA Press, pp. 229–248, 1991.
- [7] 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.
- [8] 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.
- [9] A. D. Zakrevskij. *Logic of recognition*, Minsk, Nauka i tekhnika (in Russian), 1988.
- [10] 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, v. B, pp. 197–200, 1993.

- [11] A. D. Zakrevskij. “Pattern recognition as solving logical equations”, *Special Issue 1999 – SSIT’99 (AMSE)*, pp. 125–136, 1999.
- [12] A. D. Zakrevskij. “A logical approach to the pattern recognition problem”, *Proceedings of the International Conference KDS-2001 “Knowledge – Dialog – Solution”*, St. Petersburg, v. 1, pp. 238–245, June 2001.