Optimization Models for Detection of Patterns in Data ⋆

Igor Masich, Lev Kazakovtsev, and Alena Stupina

Siberian State University of Science and Technology, Krasnoyarsk, Russia
is.masich@gmail.com

Abstract. The paper reviews the issues of detection of hidden rules (or patterns) in data sets and their use to support decision making in recognition. The problem of finding patterns is considered as the problem of conditional optimization of monotone pseudo-Boolean functions. For comparison of patterns, three criteria are used: simplicity, selectivity and evidence, as well as their possible overlap. We consider the types of patterns obtained in accordance with these criteria, which are of the greatest interest for supporting decision making in recognition. The problem of searching for informative patterns by means of formalizing this search in the form of a conditional pseudo-Boolean optimization problem is investigated. The analysis of properties of the optimization model is carried out, and a new alternative optimization model is proposed to search for strong spanned patterns.

Keywords: Pseudo-Boolean optimization · Recognition · Rule-based algorithms

1 Introduction

At present, when solving classification problems, in addition to the requirement of high accuracy, the necessity often arises for interpretability and validity of the obtained solutions. Especially interpretability and validity are key factors in solving those practical problems in which losses from making the wrong decision can be great. Therefore, the decision support system used for such problems should justify possible solutions and interpret the result.

To create such system, data classification algorithms are required that, in addition to the solution itself, provide an explicit rule, that is, they reveal knowledge from the available data. This is true for logical classification algorithms, the working principle of which is to identify patterns in data and formalize them in the form of rules set, i.e. a set of patterns described by a simple logical formula.

⋆ Results were obtained in the framework of the state task 2.5527.2017/8.9 of the Ministry of Education and Science of the Russian Federation

The process of forming logical rules is accompanied by solving the problems of choosing the best alternatives in accordance with some criteria. In the studied method the formalization of the formation process of logical rules is implemented in the form of a number of problems of combinatorial optimization, which forms a flexible and efficient algorithm for finding patterns in the primary data. By combining a number of patterns in the composition, the obtained classifier will solve the problem.

However, at present time, there are a number of problems associated with the application of the method of logical data analysis in solving practical classification problems. One of them is the construction of optimization models for the informative patterns formation. While considering this problem, first of all, it is necessary to determine the criteria and constraints that are the foundation of these optimization models.

The main objective is increase of the interpretability of the classifier and the quality of recognition of new observations, that is, to improve the classifier's generalizing abilities.

The main object of the research conducted by the authors is the method of logical analysis of data derived from the theory of combinatorial optimization [3]. This method belongs to rule based classification algorithms and is based on identification of logical patterns from the data sample. This paper analyses the existing optimization model for searching patterns and offers an alternative optimization model for constructing strong spanned patterns.

2 The Concept of Pattern

The creation and use of logical classification algorithms is based on the identification in the primary data of the patterns from which a decision function is formed. The search for patterns can be considered as a task of combinatorial optimization. To obtain a more efficient solution, the choice of the optimization algorithm should be made proceeding from the characteristic properties inherent in the optimization considered problem. In this paper some properties of optimization problems are considered that can be solved in the course of searching of logical patterns in the data.

Let us consider the problem of recognizing objects described by binary variables and divided into two classes

$$K = K^{+} \bigcup K^{-} \subset B_2^n,$$

where $B_2^n = B_2 \times B_2 \times \cdots \times B_2$, $B_2 = \{0, 1\}$.

In this case, the classes do not intersect:

$$K^{+} \bigcap K^{-} = \emptyset.$$

An observation $X \in K$ is described by a binary vector $X = (x_1, x_2, ..., x_n)$ and can be represented as a point in the hypercube of the space of binary variables B_2^n. Observations of the class K^{+} will be called positive points of the sample K, and observations of the class K^- will be called negative points of the sample.

Let us consider a subset of points from B_2^n, for which some variables are fixed and identical, and the rest take an arbitrary value:

$$T = \{x \in B_2^n | x_i = 1 \text{ for } \forall i \in A \\ \text{and } x_j = 0 \text{ for } \forall j \in B\}$$

for some subsets $A, B \subseteq \{1, 2, \ldots, n\}$, $A \bigcap B = \emptyset$. This set can also be defined as a Boolean function that takes a true value for the elements of the set:

$$t(x) = \left(\bigwedge_{i \in A} x_i \right) \wedge \left(\bigwedge_{j \in B} \overline{x}_j \right).$$

The set of points x, for which $t(x) = 1$, is defined as $S(t)$. $S(t)$ is a subcube in the Boolean hypercube B_2^n , the number of points of the subcube is $2^{(n-|A|-|B|)}$.

The binary variable x_i or its negation $\bar{x}_i$ in a term is called a literal. Record x_i^α indicates x_i if $\alpha = 1$ and $\bar{x}_i$ if $\alpha = 0$. Thus, a term is a conjunction of various literals that does not simultaneously contain a certain variable and its negation. The multiplicity of literals in the term t is defined as $Lit(t)$.

Let us consider that the term t covers the point $a \in B_2^n$, if $t(a) = 1$, that is, this point belongs to the corresponding subcube.

In this case under a *pattern* P (or *rule*) is meant a term that covers at least one observation of a certain class and does not cover any observations of another class [1]. That is, the pattern corresponds to a subcube having a non-empty intersection with one of the sets (K^+ or K^-) and an empty intersection with another set (K^- or K^+ relatively). A consistent pattern P, which does not intersect with K^- , will be called positive and a consistent pattern P', which does not intersect with K^+ will be called negative. Below for definiteness only positive patterns will be considered. A set of observations that are covered by the pattern P is defined as $Cov(P)$.

Patterns are elementary blocks for constructing logical recognition algorithms. The most useful for this purpose are patterns with the largest coverage (maximum patterns), that is, those for which $|Cov(P)|$ is maximum.

One of the ways in constructing a set of patterns for the recognition algorithm is to search for patterns that are based on the values of the attributes of specific objects.

3 Criteria for Selection and Optimality of Patterns

There are no single unambiguous criteria for comparing logical patterns with each other. While analyzing various data, different requirements can be imposed on the quality and features of the formed patterns. In accordance with [5], to evaluate the quality of pure patterns (homogeneous, non-covering observations of other classes), three criteria are used – simplicity, selectivity and evidence, as well as their possible overlap.

The criterion of simplicity (or compactness) is often used to compare patterns among themselves, including those obtained by different learning algorithms.

A pattern P_1 is more preferable to P_2 by the criterion of *simplicity* (define $P_1 \succeq_\sigma P_2$), if $Lit(P_1) \subseteq Lit(P_2)$.

In [2] a pattern P is prime if after removing any literal from $Lit(P)$ a term is formed and is not a (pure) pattern (that is, it covers the observations of another class).

It is obvious that the optimality of a pattern by the criterion of simplicity is identical with the assertion that this pattern is the prime.

The search for simpler patterns has well-founded premises. First, such patterns are better interpreted and understandable for humans when used in decision making. Secondly, it is often believed that simpler patterns have a better generalizing ability, and their use leads to better recognition accuracy. However, this assertion is debatable; moreover, it has been shown in [6] that a decrease in simplicity can lead to greater accuracy.

The usage of simple and, therefore, short patterns leads to the fact that the number of incorrectly recognized positive observations (false negative) decreases, but at the same time this can lead to an increase in the number of incorrectly recognized negative observations (false positive observations). The natural way to reduce the number of erroneous positive observations is the formation of more selective observations, which is achieved by reducing the size of the subcube that determines the pattern.

A pattern P_1 is more preferable to P_2 by criterion of *selectivity* (define $P_1 \succeq_\Sigma P_2$) if $S(P_1) \subseteq S(P_2)$.

It should be noted that the two criteria considered above are opposite to each other, that is, $Lit(P_1) \subseteq Lit(P_2) \Leftrightarrow S(P_1) \supseteq S(P_2)$.

A pattern that optimal by the selectivity criterion is a minterm, that is, a pattern covering the only positive observation. The use of this criterion by itself is, of course, not effective, since the resulting patterns (minterms) do not possess any generalizing ability. But the selectivity criterion is extremely useful when used in conjunction with other criteria, which will be discussed later.

Another useful criterion is a criterion based on coverage, that is, the number of positive observations of the training sample that satisfy the conditions of the pattern. Undoubtedly, the patterns with greater coverage have a greater generalizing ability. And the observations of the training sample, covered by the pattern, are a kind of proof of the applicability of this pattern to the decision.

A pattern P_1 is more preferable to P_2 by criterion of *evidence* (define $P_1 \succeq_\varepsilon P_2$), if $Cov(P_1) \supseteq Cov(P_2)$.

The patterns that are optimal by the criterion of evidence are called *strong.* That is, a pattern P is strong in such case that there is no such pattern *P'*, that $Cov(P') \supset Cov(P)$.

It is important to note that the considered criteria are not completely independent. So, the criteria of simplicity and selectivity are opposite to each other. Moreover, the following dependencies can be noted:

$$P_1 \succeq_\sigma P_2 \Rightarrow P_1 \succeq_\varepsilon P_2,$$

$$P_1 \succeq_\Sigma P_2 \Rightarrow P_2 \succeq_\varepsilon P_1.$$

To combine the considered criteria with each other two methods are used.

For given criteria π and ρ a pattern P_1 is more preferable with respect to P_2 by intersection of $\pi \wedge \rho$ (define $P_1 \succeq_{\pi\wedge\rho} P_2$), if and only if $P_1 \succeq_\pi P_2$ and $P_1 \succ_\rho P_2$.

For given criteria π and ρ a pattern P_1 is more preferable with respect to P_2 on lexicographical refinement $\pi|\rho$ (define $P_1 \succeq_{\pi|\rho} P_2$), if and only if $P_1 \succ_\pi P_2$ or ($P_1 \approx_\pi P_2$ and $P_1 \succeq_\rho P_2$).

In [2] it is shown that only three of all possible combinations of criteria deserve attention. Patterns that are optimal by criterion $\Sigma\wedge\varepsilon$ are called *spanned*. Patterns that are optimal by criterion $\varepsilon|\sigma$ are called *strong prime*. And patterns that are optimal by criterion $\varepsilon|\Sigma$ are called *strong spanned*.

Among all the types of patterns obtained in accordance with the above criteria and their combinations, patterns of prime, strong prime and strong spanned are most useful for identifying informative patterns and their use to support decision making in recognition.

4 Search for Optimal Patterns

Let us emphasize some $a \in K^+$ observation and define pattern P^a, covering the observation a. Those variables that are fixed in P^a , are equal to the corresponding values of the attributes of the object a.

To define the pattern P^a, the binary variables $Y = (y_1, y_2, \ldots, y_n)$ are introduced:

$$y_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ attribute is fixed in } P^a, \\ 0, & \text{otherwise.} \end{cases}$$

Some point $b \in K^+$ will be covered by the pattern P^a only if $y_i = 0$ for all i for which $b_i \neq a_i$. On the other hand, some point $c \in K^-$ will not be covered by the pattern P^a in case if $y_i = 1$ at least for one variable i, for which $c_i \neq a_i$.

A condition that says that a positive pattern must not contain any point K^- requires that for each observation $c \in K^-$ variable y_i take the value 1 for at least one i, for which $c_i \neq a_i$ [2]:

$$\sum_{\substack{i=1 \\ c_i \neq a_i}}^{n} y_i \geq 1.$$

Strengthening the constraint to increase resistance to errors is made by replacing the number 1 on the right-hand side of the inequality by a positive integer d.

On the other hand, a positive observation $b \in K^+$ will be a part of considerate subcube when the variable y_i takes the value 0 for all indices i for which $b_i \neq a_i$.

Thus, the number of positive observations covered by a-pattern can be calculated as:

$$\sum_{b \in K^+} \prod_{\substack{i=1 \\ b_i \neq a_i}}^{n} (1 - y_i).$$

Thus, the problem of finding the maximum pattern can be written in the form of a problem of finding such values $Y = (y_1, y_2, \ldots, y_n)$, in which the obtained pattern P^a covers as many possible points $b \in K^+$ and does not cover any points $c \in K^-$ [4]:

$$\sum_{b \in K^+} \prod_{\substack{i=1 \\ b_i \neq a_i}}^{n} (1 - y_i) \to \max_{Y}, \quad (1)$$

$$\sum_{\substack{i=1 \\ c_i \neq a_i}}^{n} y_i \geq 1 \text{ for all } c \in K^-. \quad (2)$$

This problem is the problem of conditional pseudo-Boolean optimization, that is, the optimization problem in which the objective functions and the functions on the left-hand side of the constraint are pseudo-Boolean functions (real functions of Boolean variables).

The problem of finding the maximum negative patterns is formulated similarly.

It should be noted that any point $Y = (y_1, y_2, \ldots, y_n)$ corresponds to a subcube in the space of binary variables $X = (x_1, x_2, \ldots, x_n)$, including the basic observation a. When $Y \in O_k(Y^1)$ (in other words Y differs from Y^1 by value of k coordinate), where $Y^1 = (1, 1, ..., 1)$, the number of points of this subcube is 2^k.

Each found pattern is characterized by coverage (the number of covered observations of the certain class) and the degree (the number of fixed variables that determine this pattern). According to the above optimization model (1)–(2), the obtained patterns do not cover any observations of another class (from the training sample).

The most valuable patterns are those that have the greatest coverage. The larger the coverage, the better the pattern reflects the image of the class.

5 Optimization Problem Properties

Let us consider optimization problem properties (1)–(2). For this, first of all, we mention the basic concepts [1].

- Points $X^1, X^2 \in B_2^n$ are called *k-nearly*, if they differ in value k coordinate, $k = 1, \ldots, n$.

- Set $O_k(X)$, $k = 0, \ldots, n$, of all points B_2^n, k-nearly to the point X, is called *k-th level* of point X.
- Point $X \in B_2^n$ is called k-nearly to the set $A \subset B_2^n$, if $A \bigcap O_k(X) \neq \emptyset$ and $\forall l = 0, \ldots, k-1 : A \bigcup O_l(X) = \emptyset$.
- Point $X^* \in B_2^n$, for which $f(X^*) < f(X), \forall X \in O_1(X^*)$, is called *local minimum* of a pseudo-Boolean function f.
- A pseudo-Boolean function having only one local minimum is called *unimodal function* on B_2^n.
- Unimodal function f is called *monotone* on B_2^n, if $\forall X^k \in O_k(X^*), k = 1, \ldots, n$: $f(X^{k-1}) \leq f(X^k)$, $\forall X^{k-1} \in O_{k-1}(X^*) \bigcap O_1(X^k)$, and *strictly monotone*, if this condition is satisfied with the strict inequality sign.
- A set of points $W(X^0, X^l) = \{X^0, X^1, \ldots, X^l\} \subset B_2^n$ is called *path* between points X^0 and X^l, if $\forall i = 1, \ldots, l$ point X^i is nearly to X^{i-1}.
- Path $W(X^0, X^l) \subset B_2^n$ between k-nearly points X^0 and X^l is called *shortest*, if $l = k$.
- $\forall X, Y \in B_2^n$ combination of all shortest paths $W(X, Y)$ is called *subcube* B_2^n and define as $K(X, Y)$.

Let us consider the basic properties of the set of feasible solutions of the conditional pseudo-Boolean optimization problem. There is a problem of the following form:

$$C(X) \to \max_{X \in S \subset B_2^n}, \tag{3}$$

where $C(X)$ is a monotone increasing from X^0 pseudo-Boolean function, $S \subset B_2^n$ is some feasible subspace of Boolean variables, defined by a given system of constraints, for example:

$$A_j(X) \leq H_j, j = 1, \ldots, m.$$

Let us introduce a number of concepts for a subset of points of the space of Boolean variables [1].

- Point $Y \in S$ is a *boundary point* of the set S, if there are $X \in O_1(Y)$, for which $X \notin S$.
- Point $Y \in O_i(X^0) \bigcap S$ is called *limiting point* of the set S with base point $X^0 \in S$, if $\forall X \in O_1(Y) \bigcap O_{i+1}(X^0)$ condition $X \notin S$ is fulfilled.
- The constraint defining the subarea of the space of Boolean variables are called *active* if the optimal solution of the conditional optimization problem (3) does not coincide with the optimal solution of the corresponding optimization problem without the considering constraint.

One of the properties of the set of feasible solutions is as follows:

Property 1. If the objective function is a strictly monotonic unimodal function and the constraint is active, then the optimal solution of problem (3) is a point belonging to the subset of limiting points of the set of feasible solutions S with the base point X^0 at which the objective function takes minimum value:

$$C(X^0) = \min_{X \in B_2^n} C(X).$$

Let us consider a separate restriction in the optimization problem (1)–(2):

$$A_j(Y) \geq 1, \text{ for some } c^j \in K^-, j = \{1, 2, ..., |K^-|\},$$

$$\text{where } A_j(Y) = \sum_{\substack{i=1 \\ c_i^j \neq a_i}}^{n} y_i.$$

Introduce the notation

$$\delta_i^j = \begin{cases} 1, & \text{if } c_i^j \neq a_i; \\ 0, & \text{if } c_i^j = a_i. \end{cases}$$

Then

$$A_j(Y) = \sum_{i=1}^{n} \delta_i^j y_i.$$

Function $A_j(Y)$ is monotone decreasing from the point $Y^1 = (1, 1, \ldots, 1)$.

The limiting points of the feasible domain are points $Y_k \in O_{n-1}(Y^1)$ (either, which is the same thing, $Y_k \in O_1(Y^0)$), and such that Y_k differ from $Y^0 = (0, 0, \ldots, 0)$ by the value of the k-th component, for which $\delta_k^j = 1$.

The set of feasible solutions is the combination of the subcubes formed by the limiting points of the feasible space and the point Y^1:

$$\bigcup_{k:\delta_k^j = 1} K(Y_k, Y^1).$$

Now let us move on to the entire system of constraints

$$A_j(Y) \geq 1, \text{ for all } j = 1, 2, \ldots, |K^-|.$$

This system will be satisfied by points belonging to the set

$$\bigcap_{j=1}^{|K^-|} \bigcup_{k:\delta_k^j = 1} K(Y_k^j, Y^1),$$

which, in the final analysis, is the union of a finite number of subcubes. The limiting points of the feasible space can be at completely different levels of the point Y^1. And their number, in the worst case, can reach the value $C_n^{[n/2]}$ that is the average power level.

Next, let us consider the objective function

$$C(Y) = \sum_{b \in K^+} \prod_{\substack{i=1 \\ b_i \neq a_i}}^{n} (1 - y_i)$$

for some "basic" observation $a \in K^+$. Or it can be written as

$$C(Y) = \sum_{j=1}^{|K^+|} \prod_{i=1}^{n} (1 - \Delta_i^j y_i),$$

$$\text{where } \Delta_i^j = \begin{cases} 1, & \text{if } b_i^j \neq a_i \\ 0, & \text{if } b_i^j = a_i. \end{cases}$$

The function $C(X)$ increases monotonically from the point $Y^1 = (1, 1, ..., 1)$, taking the value 1 in it, which corresponds to the covering of the "basic" observation a. The largest value, equal to $|K^+|$, the function $C(X)$ takes the value $Y^0 = (0, 0, ..., 0)$ at a point.

Let us suppose that the observation $b \in K^+$ closest to observation a differs in the value of the s component.

$$s = \min_{b \in K^+ \setminus \{a\}} \sum_{i=1}^{n} |a_i - b_i|.$$

At all points $Y \in O_k(Y^1)$, $k = 0, 1, ..., s-1$, the value of the objective function are the same and equal to 1. The presence of such a set of constancy complicates the work of optimization algorithms starting to search from a feasible point Y^1 and leading it along nearly points, since the calculation of the objective function in a nearly system consisting of nearly points does not give an information on the best direction of the search.

In solving practical problems of large dimensions this set of constancy can be such that most of the points of the feasible space belong to it. This complicates or makes the operation of such algorithms as a genetic algorithm, a local search with a multistart impossible.

Another consequence of the sets of constancy presence of the objective function is that the optimal solution is not only a point belonging to a subset of limiting points of an feasible space, but it can be an integer set of points representing a set of constancy of the objective function. The following statement is true.

Proposition 1. *The limiting points of the feasible space of problem (1)–(2) correspond to the prime patterns.*

The proof follows from the concepts of the limiting point and the prime pattern.

It should be noted that the found prime pattern will not necessarily be a strong pattern. This property is garanteed only for the optimal solution of the problem.

Proposition 2. *The optimal solution of problem (1)–(2) corresponds to a strong prime pattern.*

Proof. From the past statement it follows that the optimal solution corresponds to the prime pattern. The value of the objective function of the problem is a covering of pattern. If the solution is optimal, then there is no pattern P^a (covering the basic observation a), better by the criterion of evidence, and hence this solution corresponds to a strong pattern.

Thus, applying approximate optimization algorithms, it can be confirmed that the found pattern will be prime, but it will not necessarily be strong. If the exact optimization algorithm will be used, then found pattern will be a strong prime pattern.

6 Search for Strong Spanned Patterns

The optimization model is determined, first of all, by the way in which the variables that define the alternatives of the optimization problem are introduced. In the optimization model discussed above, the alternatives, that is, the patterns, are determined by including or not including in the pattern the conditions that are satisfied in some basic observation. Let us consider an alternative way of specifying the pattern.

Let us introduce the binary variable

$$z_b = \begin{cases} 1, \text{ if observation } b \in K^+ \\ \text{ is covered by } P^a, \\ 0, \text{ otherwise.} \end{cases}$$

To ensure that the resulting pattern is encompassing, let us introduce into it all the literals available for all positive observations that this pattern covers, that is, those literals for which the condition is true:

$$\prod_{\substack{b \in K^+ : \\ z_b = 1}} (1 - |b_i - a_i|) = 1.$$

Now let us formulate the problem of finding the maximum pattern as the maximization of the coverage of positive observations under the condition that the coverage of negative observations is in admissible:

$$\sum_{b \in K^+} z_b \to \max_Z, \tag{4}$$

$$\sum_{\substack{i = 1 \\ c_i \neq a_i}}^{n} \prod_{\substack{b \in K^+ : \\ z_b = 1}} (1 - |b_i - a_i|) \geq 1 \text{ for all } c \in K^-. \tag{5}$$

Having solved this optimization problem, we will define a set of positive observations covered by the desired pattern. The pattern itself can be determined using the characteristic variables $Y = (y_1, y_2, ..., y_n)$:

$$y_i = \prod_{\substack{b \in K^+ : \\ z_b = 1}} \left(1 - |b_i - a_i|\right), i = 1, ..., n.$$

In this formulation of the optimization problem, the objective function is strictly monotonic, and there are no sets of constancy of the objective function.

Proposition 3. *In the problem (4)–(5), the limiting points of the feasible space, and only they, correspond to strong spanned patters.*

Proof. Let us take a point $Z_k \in O_k(Z^0)$, where $Z^0 = (z_1, ..., z_{m^+})$. For any point $Z_{k+1} \in O_1(Z_k) \bigcap O_{k+1}(Z^0)$, $Cov(T_k) \subset Cov(T_{k+1})$ is fulfilled, where T_k and T_{k+1} are terms, corresponding to the points Z_k and Z_{k+1} respectively. If Z_k is an limiting point of an feasible set, then any point of $Z_{k+1} \in O_1(Z_k) \bigcap O_{k+1}(Z^0)$ is not permissible, and the term T_{k+1} is not a pattern, consequently, the term T_k is a strong pattern (according to the concept of a strong regularity).

In turn, if the point Z_k is not an limiting point then there is an admissible point of $Z_{k+1} \in O_1(Z_k) \bigcap O_{k+1}(Z^0)$ and the corresponding term T_{k+1}, which is a pattern with a large covering $Cov(T_{k+1}) \supset Cov(T_k)$, so the term T_k is not strong pattern.

7 Experimental Research

The results of experimental studies conducted on the problem of predicting complications of myocardial infarction are presented in [4]. The sample consists of 1700 observations (patients), each of which is described by 112 different types of symptoms. According to the input criteria, it is required to make a prediction about the onset of four complications of the disease: ventricular fibrillation (VF), auricular fibrillation (AF), pulmonary edema (PE), cardiac rupture (CR). 85% of sample observations were used for training, the rest – for control. The search for patterns was implemented with the help of two optimization models: the search for the maximum prime patterns (model 1) and the search for strong spanned patterns (model 2).

The optimization problems were solved using the algorithms of conditional pseudo-Boolean optimization described in [1].

Table 1 presents comparative results for the average degree of obtained patterns (the number of literals in the term) and the accuracy of recognition of control observations. As can be seen from the results, the use of strong spanned patterns makes it possible to obtain classifiers with a better generalizing ability.

Table 1. Results of experimental research

The problem of forecasting	Class	Model 1		Model 2	
		Degree	Precision	Degree	Precision
VF	neg.	5	0.90	6	0.90
	pos.	3	0.78	4	0.83
AF	neg.	7	0.80	9	0.80
	pos.	6	0.68	6	0.68
PE	neg.	7	0.68	9	0.68
	pos.	4	0.89	5	0.89
CR	neg.	5	1.00	6	1.00
	pos.	3	0.79	4	0.93

8 Conclusions

The paper examines the problem of searching for informative patterns by means of formalizing of this search in the form of a conditional pseudo-Boolean optimization problem. The analysis of the properties of the constructed optimization model is implemented and a new alternative model is proposed for optimization, designed to search for strong spanned patterns.

References

1. Antamoshkin, A.N., Masich, I.S.: Search algorithms for conditional pseudo-Boolean optimization. Control, communication and security systems 1, 103–145 (2016)
2. Bonates, T.O., Hammer, P.L., Kogan, A.: Maximum patterns in datasets. Discrete Applied Mathematics 156(6), 846–861 (2008)
3. Crama, Y., Hammer, P.L., Ibaraki, T.: Cause-effect Relationships and Partially Defined Boolean Functions. Annals of Operations Research 16, 299–325 (1988)
4. Golovenkin, S.E., Gorban, A.N., Shulman, V.A.: Complications of myocardial infarction: a database for approbation of recognition and prognosis systems: Preprint 6; Krasnoyarsk: Computing Center of the SB RAS (1997)
5. Hammer, P.L., Kogan, A., Simeone B,., Szedmak, S.: Pareto-optimal patterns in Logical Analysis of Data. Discrete Applied Mathematics 144(1–2), 79–102 (2004)
6. Holte, R.C., Acker, L., Porter, B.: Concept learning and the problem of small disjuncts. In: Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89). pp. 813–818. Morgan Kaufmann, Detroit, MI (1989)