

HEURISTICS AND CRITERIA FOR CONSTRUCTING LOGICAL PATTERNS IN DATA *

A N Antamoshkin^{1,2}, I S Masich¹ and R I Kuzmich²

¹Department of System Analysis and Operation Research, Siberian State Aerospace University named after academician M. F. Reshetnev, Krasnoyarsk, Russia

²Department of Economics and Information Technology of Management, Siberian Federal University, Krasnoyarsk, Russia

E-mail: i-masich@yandex.ru

Abstract. The article considers various optimization models for constructing patterns in the method of logical analysis of data. Application techniques of the proposed models are specified and comparison of their classification against the accuracy on the task of predicting complications of myocardial infarction is provided

Introduction

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 [1]. This method was successfully used to solve a number of problems in various areas [2-4]. It belongs to rule based classification algorithms and is based on identification of logical patterns from the original data sample. The original sample consists of two disjoint sets Ω^+ and Ω^- of observations which are n-dimensional vectors belonging to positive or negative class respectively. Vector elements, also called attributes, can be numerical, nominal or binary. The goal is to determine the class for some new observation, which is also a vector of n variables.

To achieve this goal, a classification model is to be developed, which consists of a set of patterns identified from the original data sample. Each pattern should be informative, i.e. covers a number of objects of the same class and does not cover (or covers a small amount of) the objects of another class [5].

This paper offers several optimization models for constructing a pattern in order to build an adequate classifier able to classify a new incoming object, i.e. an object, previously not participating in its building.

Model for constructing patterns with maximum coverage

The considered method is used for working with data samples, where attributes take binary values. Since the original sample can consists of different types of attributes, it needs binarizing. In this paper the “unitary” binarization method is used, discussed in detail in [6].

The notion of a pattern [7] is used as the basis for the considered method. A positive pattern can be defined as a subcube of the space of binary variables B_2^t , and this subcube intersects with the set Ω^+ and has no common elements with the set Ω^- . A negative pattern is defined similarly.

* Results were obtained in the framework of the state task № 346 of the Ministry of Education and Science of the Russian Federation



The positive ω -pattern for $\omega \in \{0,1\}^t$ is a pattern, which contains ω points. The problem is to find the maximum ω -pattern for each point $\omega \in \Omega^+$, i.e. covering the largest number of points in Ω^+ .

Let us define the corresponding subcube using variables y_j :

$$y_j = \begin{cases} 1, & \text{if } i\text{-th criterion fixed in the subcube,} \\ 0, & \text{in another case.} \end{cases}$$

That is, by fixing l variables of the original cube with dimension t , we get a subcube with dimension $(t-l)$ and with the number of points 2^{t-l} .

Requirement, which says that a positive pattern should not contain any points in Ω^- , demands that for each observation $\rho \in \Omega^-$ variable y_j takes on a value 1 at least for one j , where $\rho_j \neq \omega_j$:

$$\sum_{\substack{j=1 \\ \rho_j \neq \omega_j}}^t y_j \geq 1 \text{ for any } \rho \in \Omega^-.$$

For increasing stability to faults, the constraint may be strengthened by replacing the number 1 on the right side of the constraint on the positive integer d .

On the other hand, a positive observation of $\sigma \in \Omega^+$ is included in the considered subcube when the variable y_j takes on a value 0 for all indices j , where $\sigma_j \neq \omega_j$. Thus, the number of positive observations covered by a ω -pattern can be calculated as:

$$\sum_{\substack{\sigma \in \Omega^+ \\ \sigma_j \neq \omega_j}} \prod_{j=1}^t (1 - y_j).$$

Thus, we have the problem of constrained pseudo-Boolean optimization with algorithmically defined functions:

$$\sum_{\substack{\sigma \in \Omega^+ \\ \sigma_j \neq \omega_j}} \prod_{j=1}^t (1 - y_j) \rightarrow \max, \quad (1)$$

$$\sum_{\substack{j=1 \\ \rho_j \neq \omega_j}}^t y_j \geq d \text{ for any } \rho \in \Omega^-, y \in \{0,1\}^t. \quad (2)$$

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

Each pattern is characterized by two parameters: coverage, i.e. the number of objects of a particular class, which are covered by this pattern; and degree, i.e. the number of variables involved in its formation. Specificity of recognition problems encountered in practice is in the fact that a database includes a large number of unmeasured values (missing data), and performed measurements may be inaccurate or incorrect. Noises and outliers can lead to the situation where objects of different classes overlap each other, thus entering to the "field" of

the opposite class. As a result, obtained patterns have a greater degree and substantially less coverage, than in case of no outliers and inaccuracies. At the same time the classifier consists of a large number of small patterns (with a small coverage). It does not allow constructing an effective classifier with «well-interpreted» conventions (in which a small number of attributes are involved) and with high accuracy recognition.

To increase the stability of the method to outliers, the constraint (2) should be relaxed. It means creating possibilities for a pattern to cover some small number of objects of another class. Then, the degree of obtained patterns reduces and coverage is increased.

Thus, the constraint of the optimization model may be wrighten as follows:

$$\sum_{\rho \in \Omega^-} z_{\rho} \leq D, \quad (3)$$

$$\text{where } z_{\rho} = \begin{cases} 0, & \text{if } \sum_{\substack{j=1 \\ \rho_j \neq \omega_j}}^t y_j \geq d, \\ 1, & \text{in another case;} \end{cases}$$

D is the number (non-negative integer) of objects of another class which are allowed being covered by the pattern [8].

The procedure of growing patterns to increase their informativeness

In accordance to models 1 and 3, the most preferred patterns are those with the greatest coverage. As a result the formed patterns have small degree, i.e. consist of a small number of terms and use only some attributes [8].

Patterns with small degree correspond to large areas in the attribute space. This leads to the possible covering of objects of another class (missing in the training sample) and increasing the number of incorrectly recognized objects. This feature can influence on pattern informativeness reducing it. Therefore, to increase informativeness, we propose the growing procedure for patterns. The growing procedure is applied to each built pattern and means maximizing degree of patterns under fixing coverage:

$$\sum_{j=1}^t y_j \rightarrow \max ,$$

$$fc(Y) = fc'(Y),$$

where $fc(Y)$ is value of the objective function (1) (coverage) for the pattern before the growing procedure, $fc'(Y)$ is value of the objective function for the pattern after the growing procedure.

As a result of the growing procedure, patterns with maximum coverage and with higher degree are formed, thereby increasing reliability of decisions made by the classifier obtained accordance to this heuristic. Reliability of decisions is growing due to increased informativeness of patterns, as the number of objects of "own" class (the value of the objective function for the pattern) remains unchanged, and the number of objects of another

class covered by pattern decreases. Thus, the growing procedure contributes to increasing informativeness of each pattern and, consequently, the classifier in general.

Model for formation of patterns with various covering

The approach proposed consists in modification of the objective function (1) in order to increase various of patterns in the classifier [9].

The objective function (1) maximizes coverage for each formed pattern, capturing objects that are typical representative of the class, while atypical objects of the class remain uncovered and the classifier has no patterns comprising this objects. Therefore, we obtain a set of similar patterns for the class, thereby reducing the quality of classification. To obtain a classifier with higher various in patterns that allows to select substantially different subsets of objects, it is proposed to modify the objective function (1) for finding the positive pattern as follows:

$$\sum_{\sigma \in \Omega_s^+} K_\sigma \cdot \prod_{\substack{j=1 \\ \sigma_j \neq \omega_j}}^t (1 - y_j) \rightarrow \max, \quad (4)$$

where K_σ is weight of positive observation $\sigma \in \Omega_s^+$; this weight should be decreased if this observation is covered by the pattern, thereby this reduce participation priority of the observation in the formation of the next pattern for the sake of uncovered observations.

Similarly, the objective function in the optimization model is formed for finding negative patterns.

In order to use the optimization model with the objective function (4) for the constructing patterns, it is necessary to assign initial weights for all objects and to define rule for weight changes for objects that taken part in formation of the current pattern. The initial weights are proposed to equate to 1 for each object in the training sample. The rule for weight changes for object which took part in the formation of the current pattern:

$$K_{i+1} = \max \left[0, K_i - \frac{1}{N_{\max}} \right],$$

where K_i , K_{i+1} – weights of the covered object in the formation of the current and next patterns; N_{\max} – parameter defined by the researcher, it means the maximum number of patterns covering the object of the training sample in the classifier.

Thus, using the optimization model with the objective function (4) and the constraint (3) for constructing patterns leads to various patterns obtained.

Results

To compare the proposed models of patterns formation, a series of experiments is conducted on the problem of predicting complications of myocardial infarction (MI): ventricular fibrillation (VF), atrial fibrillation (AF), pulmonary edema (PE) [10].

Previously, the problem of predicting complications of MI was solved with use of artificial neural networks [10]. It was observed that the classifier gives decisions which are not good enough in case of a significant difference in the number of objects of each class in the original sample, therefore, the following approach was proposed in that paper to solve this problem. The number of patients with certain complication (positive objects) is approximately ten times

less than the number of patients without the complication (negative objects). The original sample (1700 objects) was divided into the test sample and 10 training samples for each complication. Furthermore, the positive objects in training samples remain the same and negative objects change. The classifier was trained at each training sample separately, but it was tested on total test sample. The final decision on each object of the test sample is made by the majority of votes received from all classifiers obtained on the basis of 10 training samples.

For constructing patterns the following four optimization models was used:

- 1) rigid model that does not allow covering objects of another class by the pattern;
- 2) relaxed model, which allows the pattern to cover some limited number of objects of another class;
- 3) modified model including the procedure of growing patterns;
- 4) model for formation of patterns which covering substantially different subsets of the sample objects.

For finding suboptimal solutions in accordance with optimization models used, heuristic algorithms of constrained pseudo-Boolean optimization described in [11-13] and implemented in the software application [14] was applied.

Classifier forms when combining all built positive and negative patterns.

The following decision rules were used to classify a new (or test) observation [5]:

- 1) If the observation satisfies the requirements of one or more positive patterns and does not satisfy the requirements of any negative pattern then it is classified as positive.
- 2) If the observation satisfies the requirements of one or more negative patterns and does not satisfy the requirements of any positive pattern then it is classified as negative.
- 3) If the observation satisfies the requirements p' from p positive patterns and q' from q negative patterns then sign of observation is defined as $p' / p - q' / q$.
- 4) If observation does not satisfy the requirements of any patterns, positive or negative, it belongs to class having the lowest price error.

The number of patients with complications (positive objects) and without complications (negative objects) in the training sample and volume of the test sample for all considered complications of MI are presented in table 1.

Table 1. Volume of each sample for all complications of MI.

Problem	Number of positive objects	Number of negative objects	Number of objects of testing sample
VF	70	181	30
AF	170	180	50
PE	159	173	39

Results of the tests are presented in tables 2-4.

Table 2. Classification results for the problem of predicting VF.

Optimization problem	Set of conditions	Number of conditions	Covering of negative objects	Covering of positive objects	Degree of conditions	Classification accuracy, %
The objective function (1),	neg.	161	14	0	5	90
	pos.	60	0	14	3	78

delimitation (2)						
The objective function (1), delimitation (3)	neg.	161	26	5	5	90
	pos.	60	5	22	3	78
The objective function (1), delimitation (3) with building procedure	neg.	161	26	3	6	90
	pos.	60	3	22	4	83
The objective function (4), delimitation (3)	neg.	54	16	5	6	80
	pos.	51	5	18	5	83

Table 3. Classification results for the problem of predicting AF.

Optimization problem	Set of conditions	Number of conditions	Covering of negative objects	Covering of positive objects	Degree of conditions	Classification accuracy, %
The objective function (1), delimitation (2)	neg.	150	13	0	6	73
	pos.	150	0	13	5	58
The objective function (1), delimitation (3)	neg.	150	27	5	7	80
	pos.	150	5	27	6	68
The objective function (1), delimitation (3) with building procedure	neg.	150	27	4	9	80
	pos.	150	4	27	6	68
The objective function (4), delimitation (3)	neg.	82	16	5	6	67
	pos.	87	5	16	5	58

Table 4. Classification results for the problem of predicting PE.

Optimization problem	Set of conditions	Number of conditions	Covering of negative objects	Covering of positive objects	Degree of conditions	Classification accuracy, %
The objective function (1), delimitation (2)	neg.	152	18	0	7	62
	pos.	141	0	14	4	82
The objective function (1), delimitation (3)	neg.	152	36	5	7	68
	pos.	141	5	29	4	89
The objective function (1), delimitation (3) with building procedure	neg.	152	36	4	9	68
	pos.	141	4	29	5	89

The objective function (4), delimitation (3)	neg.	67	16	5	6	62
	pos.	73	5	16	4	93

Conclusion

Tables 2-4 show that applying constraint (3) in the optimization model when searching for patterns allows finding patterns with higher coverage and a more accurate classifier is constructed on the basis of these patterns. Application of the considered approach is proper for solving problems with the existence of outliers and noises and with a lot of omissions in the data samples.

As a result of application procedure of growing patterns, the patterns with maximum covering and with higher degree are obtained, thus increasing reliability of the decisions made by the classifier constructed on the basis of these patterns. Increase in reliability of decisions occurs due to growth of informativeness of patterns. This is due to the fact that the number of the covered objects of own class remains unchanged but the number of the covered objects of another class decreases.

Application of the objective function (4) in the optimization model allows simplifying the classifier, thus reducing the number of patterns in it with respect to the complete set of patterns.

References.

- [1] Hammer P L 1986 The Logic of Cause-effect Relationships *Lecture at the Int. Conf. on Multi-Attribute Decision via Operations Research-based Expert Systems (Passau, Germany)*
- [2] Alexe S, Blackstone E. and Hammer P L 2003 Coronary Risk Prediction by Logical Analysis of Data *Annals of Operations Research* **119** 15-42
- [3] Boros E, Hammer P L and Ibaraki T 2000 An Implementation of Logical Analysis of Data *IEEE Transactions on Knowledge and Data Engineering* **12**(2) pp 292-306
- [4] Hammer P L, Kogan A and Lejeune M 2004 Modeling Country Risk Ratings Using Partial Orders *RUTCOR Research Report* 24-2004
- [5] Kuzmich R I and Masich I S 2012 Building a model for classification as decomposition of information patterns *Control Systems and Information Technology* **2**(48) 18-22
- [6] Rastrigin L A and Freymanis E E 1988 Solving problems of different scaled optimization by methods of random search *Problems of random search* **11** 9-25
- [7] Hammer P L and Bonates T 2005 Logical Analysis of Data: From Combinatorial Optimization to Medical Applications *RUTCOR Research Report* 10-2005
- [8] Golovenkin S E, Gulakova T K, Kuzmich R I, Masich I S and Shulman V A 2010 Model of logical analysis for solving the problem of predicting myocardial infarction *Vestnik SibSAU* **4**(30) 68-73
- [9] Kuzmich R I and Masich I S 2014 Modification of the objective function in the construction of patterns to increase the different conditions in the classification model *Control Systems and Information Technology* **2**(56) 14-18
- [10] Golovenkin S E, Gorban A N, Shulman V A and Rossiev D A 1997 Complications of myocardial infarction: a database for testing recognition and prediction *Preprint* 1997.6 (Krasnoyarsk: Inst. of Computational Modeling SD RAS)

[11] Masich I S 2006 Approximate algorithms for finding the end-points for the problem of constrained pseudo-Boolean optimization *Vestnik SibSAU* **1(8)** 39-43

[12] Antamoshkin A N and Masich I S 2006 Heuristic search algorithms for monotone pseudo-boolean function conditional optimization *Engineering and automation problems* **1(5)** 55-61

[13] Antamoshkin A N and Masich I S 2007 Pseudo-Boolean optimization in case of unconnected feasible sets *Models and Algorithms for Global Optimization. Series: Springer Optimization and Its Applications* vol 4, ed A Törn and J Žilinskas (New York: Springer) pp 111-122

[14] Masich I S, Kuzmich R I and Kraeva E M 2011 *Logical analysis of data in classification problems* 2011612265 (Moscow: RosPatent)