Increasing the Efficiency of Combinatorial Logical Data Analysis in Recognition and Classification Problems

E. V. Djukova[a], A. S. Inyakin[a], N. V. Peskov[a], and A. A. Sakharov[b]

[a] *Dorodnitsyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia*
e-mail: djukova@ccas.ru, andre_w@mail.ru, nick@motor.ru
[b] *Moscow State Pedagogical University, ul. Malaya Pirogovskaya 1, Moscow, 119992 Russia*
e-mail: caxap@mail.ru

Abstract—Problems of increasing the efficiency of combinatorial logical data analysis in recognition problems are examined. A technique for correct conversion of initial information for reduction of its dimensionality is proposed. Results of testing this technique for problems of real medical prognoses are given.

DOI: 10.1134/S1054661806040201

1. INTRODUCTION

In contrast to the statistical approach, procedures for combinatorial logical recognition and classification have a number of significant advantages including the possibility of obtaining an acceptable result in the case where a few objects are described with numerous features [1]. Moreover, a decision rule derived with the use of these procedures is usually easily interpreted. This circumstance is particularly important in medical diagnostics, ecological monitoring, sociological interview processing, etc.

Combinatorial logical data analysis in recognition problems is based on construction of the fragments of descriptions of objects having properties that are extreme in some sense [1, 2, 6–8, 10, 12–15]. The constructed fragments must reflect certain regularities in descriptions of teaching objects. Such fragments serve as elementary classifiers and allow one to classify new objects. The technique is particularly efficient for the case where the data is integer and the number of the allowed values of every feature is not large. To process high-dimensionality data, it is required either to adapt logical recognition procedures (e.g., to use algorithms of voting by representative sets with partial conversion [9]) or to convert the initial data during their preprocessing.

Preliminary conversion is usually based on partition of the range of each feature into intervals. As examples, we can cite the following partition methods: partition of
1 each range of features into equilength intervals, parti-
2 tion of each feature range into equipotential intervals,
partition with the use of the class-attribute independence maximization (CAIM) algorithm [16], and division with the use of Senko's algorithm [11]. In general, the cited methods do not necessarily preserve a given partition of objects into classes, which makes impossible the use of a number of combinatorial logical recognition procedures (e.g., test algorithms [3, 10]). Therefore, the concept of correct conversion appears to be important. Correct conversion means learning data conversion such that the descriptions of objects from different classes remain distinguishable. The problem can be reduced to finding the covering of a Boolean matrix that is constructed in a special way using the learning sample. The idea of this reduction was put forward by Yu. I. Zhuravlev. The number of correct conversions increases exponentially with the problem dimensions. Therefore, an intricate problem of choosing the "best" conversion appears.

In this paper, a technique is proposed for rapidly constructing correct conversions that are best in a certain sense and guarantee a high quality of solution of the recognition problem. This technique is tested on a real medical prognosis problem.

2. CODING COVERINGS

Let M be the analyzed set of objects, and every object from M be describable in the feature system $\{x_1, \dots, x_n\}$, $\{S_1, \dots, S_m\}$ be the learning sample, Si = $(a_{i1}, \dots, a_{in})$, $i = 1, 2, \dots, m$ (here, a_{ij} is the value of the feature x_j, $a_{ij} \in \Re$, and $\Re$ is the set of real numbers). For simplicity, let us consider the case where objects $S_1, \dots, S_m$ are divided into two disjoint classes K_1 and K_2, $|K_1| = m_1$, $|K_2| = m_2$.

We refer to the value $(a_{i_1 j} + a_{i_2 j})/2$ as the threshold for the feature x_j if $a_{ij} \notin (a_{i_1 j}, a_{i_2 j})$ for $i = 1, 2, \dots, m$

Received March 31, 2005

and objects S_{i_1} and S_{i_2} belong to different classes. We designate the set of all thresholds for the feature x_j as $D^{(j)}$. Let $D_j^* \subseteq D^{(j)}$, $D_j^* = \{d_1^{(j)}, \ldots, d_{u_j}^{(j)}\}$; we designate the sequence of numbers $d_1^{(1)}, \ldots, d_{u_1}^{(1)}, d_1^{(2)}, \ldots, d_{u_2}^{(2)}, \ldots, d_1^{(n)}, \ldots, d_{u_n}^{(n)}$ as $\pi(D_1^*, \ldots, D_n^*)$. The sum of the objects S_{i_1} and S_{i_2} over the set of thresholds $\pi(D_1^*, \ldots, D_n^*)$ is defined as the string

$$\Big(a_{i_1 1} \oplus a_{i_2 1}\big|_{d_1^{(1)}}, a_{i_1 1} \oplus a_{i_2 1}\big|_{d_2^{(1)}}, \ldots, a_{i_1 1} \oplus a_{i_2 1}\big|_{d_{u_1}^{(1)}},$$
$$a_{i_1 2} \oplus a_{i_2 2}\big|_{d_1^{(2)}}, a_{i_1 2} \oplus a_{i_2 2}\big|_{d_2^{(2)}}, \ldots, a_{i_1 2} \oplus a_{i_2 2}\big|_{d_{u_2}^{(2)}},$$
$$\ldots, a_{i_1 n} \oplus a_{i_2 n}\big|_{d_1^{(n)}}, a_{i_1 n} \oplus a_{i_2 n}\big|_{d_2^{(n)}}, \ldots, a_{i_1 n} \oplus a_{i_2 n}\big|_{d_{u_n}^{(n)}}\Big),$$

where $a_{i_1 j} \oplus a_{i_2 j}\big|_d$ is 1 if $a_{i_1 j}$ and $a_{i_2 j}$ lie on different sides of the threshold d and equals 0 otherwise.

Let us construct a Boolean matrix L. The matrix L is an $h \times N$ matrix, where $h = m_1 m_2$ and $N = |D^{(1)}| + \ldots + |D^{(n)}|$. Each row of this matrix is obtained as a result of the pairwise summation of objects that belong to different classes over the set of thresholds $\pi(D_1, \ldots, D_n)$. According to the construction, the set of thresholds $D^{(j)}$ corresponds to the group of the $|D^{(j)}|$ columns of the matrix L, which is denoted as G_j.

The value a_{ij} of the feature x_j for S_i generally defines two adjacent thresholds $d_1^{(ij)}$ and $d_2^{(ij)}$ from $D^{(j)}$ such that $d_1^{(ij)} < a_{ij} < d_2^{(ij)}$. If the value a_{ij} is the least or the largest value of the feature x_j, then it defines only one threshold (the minimum or the maximum element in $D^{(j)}$, respectively). We designate the set of thresholds defined by the value a_{ij} as $D_i^{(j)}$. A case where $D_i^{(j)} = \varnothing$ is generally possible. According to the construction, the set of the thresholds $D_i^{(j)}$, $j \in \{1, 2, \ldots, n\}$, corresponds to a group of columns of the matrix L, which is denoted as G_j.

Let $S_i \in K_l$, $i \in \{1, 2, \ldots, m\}$, $l \in \{1, 2\}$, $\overline{m}_l = m - m_l$. We assign object S_i to a submatrix L^i of the matrix L. The matrix L^i is an $\overline{m}_l \times N_i$ matrix, where $N_i = |D_i^{(1)}| + \ldots + |D_i^{(n)}|$. Each row of this matrix is obtained by summation of the object S_i with every learning object that does not belong to the same class with S_i over the set of thresholds $\pi(D_i^{(1)}, \ldots, D_i^{(n)})$.

The set H of columns of the matrix L is called a complete coding covering if the following two conditions are fulfilled: (i) H is a covering of L—i.e., for every row of the matrix L, at least one column in H has 1 at the intersection with this row; and (ii) $H \cap G_j \neq \varnothing$ for $j = 1, 2, \ldots, n$.

Let $i \in \{1, 2, \ldots, m\}$. A set H of the columns of the matrix L is called an i-partial coding covering if the following three conditions are fulfilled: (i) every column entering into H belongs to $\bigcup_{j=1}^{n} G_{ij}$; (ii) for every row of the matrix L^i, there is a column in H having 1 at the intersection with this row; and (iii) $H \cap G_{ij} \neq \varnothing$ for $G_{ij} \neq \varnothing$, $j = 1, 2, \ldots, n$.

Let H be a complete coding covering for L. The number $\max\limits_{j \in \{1, 2, \ldots, n\}} |H \cap G_j| + 1$ is called the dimensionality of H. When H is an i-partial coding covering for L, the number $\max\limits_{j \in \{1, 2, \ldots, n\}} |H \cap G_{ij}| + 1$ is called the dimensionality of H. Thus, the dimensionality of the i-partial coding covering is at most three.

For every $S \in M$, coding covering H apparently defines a mapping conv(H, S): $S \longrightarrow S^H$ based on the replacement of feature values for object S by numbers from $\{0, 1, \ldots, k - 1\}$, where k is the dimensionality of H. Indeed, let a_{pj}, $p \in \{1, 2, \ldots, m\}$ be some value of the feature x_j for object S_p, and $D_H^{(j)} = \{d_1, \ldots, d_\nu\}$, $d_1 < \ldots < d_\nu$, be the set of thresholds from $D^{(j)}$ corresponding to the columns from H. Three variants are possible: (i) $a_{pj} < d_1$; (ii) $d_t < a_{pj} < d_{t+1}$, $t \in \{1, 2, \ldots, \nu - 1\}$; and (iii) $d_\nu < a_{pj}$. In cases (i), (ii), and (iii) the element a_{pj} is coded by 0, t, and ν, respectively. If $D_H^{(j)} = \varnothing$, then the value of the element a_{pj} is replaced by 0. Similarly, we define conv(H, a_{pj}) as the mapping of the element a_{pj} with respect to the coding covering H.

3. MODEL OF RECOGNIZING ALGORITHMS WITH PARTIAL CONVERSIONS

Let features $x_1, \ldots, x_n$ have integer values and $(w_1, \ldots, w_n)$ be a description of an object W from M (w_j is the value of the feature x_j, $j = 1, 2, \ldots, n$). Let Q be a set of r different features, $Q = \{x_{j_1}, \ldots, x_{j_r}\}$. The set Q separates a fragment $(W, Q) = (w_{j_1}, \ldots, w_{j_r})$ in the description of the object W. Let two objects W' and W'' from M be given, $W' = (w_1', \ldots, w_n')$ and $W'' = (w_1'', \ldots, w_n'')$. The similarity of the objects W' and W'' with respect to the set of features $Q = \{x_{j_1}, \ldots, x_{j_r}\}$ is estimated as

$$B(W', W', Q) = \begin{cases} 1, \text{ if } w'_{j_t} = w''_{j_t} \text{ for } t = 1, 2, \ldots, r, \\ 0, \quad \text{otherwise.} \end{cases}$$

Thus, the objects W' and W'' are similar with respect to Q if and only if the fragments (W', Q) and (W'', Q) coincide with each other.

Let $K \in \{K_1, K_2\}$, W be a teaching object from the class K. The fragment (W, Q) is called a representative set for class K if the equality $B(W, W', Q) = 0$ is valid for every teaching object $W' \notin K$.

Let $Q^{(t)} = Q\backslash\{x_{j_t}\}$, $t \in \{1, 2, \ldots, r\}$. The representative set (W, Q) for class K is called a blind set if the fragment $(W, Q^{(t)})$ is not a representative set for K for any $t \in \{1, 2, \ldots, r\}$.

Consider a model of recognition algorithms that use partial conversion. Let H be an i-partial conversion $S_i \in K$. The aggregate of all blind representative sets for K that is generated by the object $\mathrm{conv}(H, S_i)$ is denoted as $P_i(H)$. The set of all irreducible coverings of the matrix L^i is denoted as $\Im(L^i)$. It is easy to see that the set of feature values L^i of the object $(\mathrm{conv}(H, a_{ij_1}), \ldots, \mathrm{conv}(H, a_{ij_r}))$ belongs to $\mathrm{conv}(H, S_i)$ if and only if the set of columns of the matrix $P_i(H)$ corresponding to the set of thresholds $\pi(D_H^{(j_1)}, \ldots, D_H^{(j_r)})$ belongs to $\Im(L^i)$. Let $\{H_1^i, \ldots, H_{\nu_i}^i\}$ be the set of all i-partial conversions for L. Let us designate

$$P_i = \bigcup_{t=1}^{\nu_i} P_i(H_t^i).$$

The proposed model of recognizing algorithms is based on constructing the sets $P_1, \ldots, P_m$. In view of the above consideration, the problem of constructing a set P_i, $i \in \{1, 2, \ldots, m\}$ is reduced to constructing the set $\Im(L^i)$.

We set $P(K_j) = \bigcup_{\mathrm{conv}(H, S_i) \in K_j} P_i$, $j = 1, 2$. The classification of the object S among one of the classes is determined by calculating the estimates $\Gamma(S, K_1)$ and $\Gamma(S, K_2)$, where

$$\Gamma(S, K_j) = \frac{1}{|P(K_j)|}$$
$$\times \sum_{(\mathrm{conv}(H, S'), Q) \in P(K_j)} B(\mathrm{conv}(H, S), \mathrm{conv}(H, S'), Q),$$
$$j \in \{1, 2\}.$$

Object S belongs to the class for which this estimate is maximal. If the estimates $\Gamma(S, K_1)$ and $\Gamma(S, K_2)$ are equal to each other, then the algorithm fails to recognize the object S.

4. ALGORITHM FOR CHOOSING THE BEST CORRECT CONVERSION

Below, an initial data conversion algorithm is proposed on the basis of analysis of the Boolean matrix L, which allows one to rapidly obtain a good correct conversion. To construct this algorithm, the well-known idea of steepest descent is used.

We designate a submatrix of the matrix L that is composed of the columns from G_j, $j \in \{1, 2, \ldots, n\}$ as L_{G_j}.

A column of the matrix covers a row of this matrix if 1 stands on their intersection. Here, we give the description of the conversion algorithm.

Let us specify an integer q, $1 \leq q \leq \min_{j \in \{1, \ldots, n\}} |D^{(j)}|$.

Step 1. We set $p = 1$. In every submatrix L_{G_j}, $j = 1, 2, \ldots, n$, we choose a column with the maximum number of ones (this column corresponds to the threshold that recognizes the objects from different classes in the best way). We mark the chosen column of the submatrix L_{G_j} and the rows covered by it in this matrix. If $q = 1$, then set $p = 0$ and go to Step 3.

Step 2. In every matrix L_{G_j}, $j = 1, 2, \ldots, n$, we choose (if possible) a column with the maximum number of ones in unmarked rows. The chosen column of the submatrix L_{G_j} and the rows covered by it in this matrix are marked. If $p = 0$, then Step 2 is repeated q times. Otherwise, Step 2 is repeated $q - 1$ times.

Step 3. We check if the aggregate of the marked columns is a covering of the matrix L. If it is, then the algorithm terminates. Otherwise, we erase all marks in the rows of the matrices L_{G_j}, $j = 1, 2, \ldots, n$. In every matrix L_{G_j}, $j = 1, 2, \ldots, n$, we mark the rows with the ordinal numbers that equal the ordinal numbers of the rows in the matrix L that are covered by the mentioned aggregate of the marked columns. We set $p = 0$ and go to Step 2.

It is obvious that the derived aggregate of the marked columns is a complete coding covering of the matrix L. According to this coding covering, a correct conversion of the learning sample is constructed. The described procedure is iterated for every admissible value of q. Among all the obtained correct conversions, we choose the one that gives the best results in sliding control.

Testing results

Conversion algorithm	Recognition result in sliding control
Correct conversion	83%
CAIM	77%
SENKO	62%
Partial conversion	66%

The described conversion method was compared with other conversion methods in sliding control. Here, the algorithm of voting by representative sets was chosen as the recognizing algorithm. This algorithm was tested for a real medical prediction problem. In the 3 experiment, the survival rate of patients with osteogenic sarcoma during one year after therapy was investigated. The problem of prediction of the survival rate of patients using the set of the cytological parameters of a tumorous tissue sample, which were obtained using biopsy during the course of treatment, was studied. The previous investigations revealed that the problem is very complicated, because not only the state of tumor cells, but also many other objective factors such as the 4 immunity and psychophysical state of the patient, the environment, etc., affect the treatment outcome.

The available sample included information on 78 patients; 25 and 53 of them lived less and more than a year after a course of treatment, respectively. The dimension of the feature space is 7. In the initial data, from six to 38 thresholds correspond to each feature.

The testing results are presented in the table.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project no. 04-01-00795) and by the Council of the President of the Russian Federation for Support of Young Scientists and Leading Scientific Schools (project no. NSh-1721.2003.1).

REFERENCES

1. E. V. Djukova and Yu. I. Zhuravlev, “Discrete Analysis of Feature Descriptions in Recognition Problems of High Dimensionality,” Zh. Vychisl. Mat. Mat. Fiz. **40** (8), 1264–1278 (2000) [Comput. Math. Math. Phys. **40**, 1214–1227 (2000)].
2. E. V. Djukova and N. V. Peskov, “Search for Informative Fragments in Description of Objects in Discrete Recognition Procedures,” Zh. Vychisl. Mat. Mat. Fiz. **42** (5), 741–753 (2002) [Comput. Math. Math. Phys. **42**, 711–723 (2002)].
3. E. V. Djukova, “On the Asymptotically Optimal Algorithm of Blind Test Generation,” Dokl. Akad. Nauk SSSR **233** (4), 527–530 (1977).
4. E. V. Djukova, Yu. I. Zhuravlev, N. V. Peskov, and A. A. Sakharov, “Processing of Real-Valued Data by Logical Recognition Procedures,” Iskusst. Intellekt, No. 2, 80–85 (2004).
5. E. V. Djukova, “On the Complexity of Implementation of Some Recognition Procedures,” Zh. Vychisl. Mat. Mat. Fiz. **27** (1), 114–127 (1987).
6. E. V. Djukova, “Recognition Algorithms of Core Type: Realization Complexity and Metric Properties,” in *Recognition, Classification, and Prediction (Mathematical Methods and Their Application)* (Nauka, Moscow, 1989), Vol. 2, pp. 99–125.
7. E. V. Djukova and A. S. Inyakin, “Classification Procedures Based on the Construction of Class Coverings,” Zh. Vychisl. Mat. Mat. Fiz. **43** (12), 1884–1895 (2003) [Comput. Math. Math. Phys. **43**, 1812–1822 (2003)].
8. E. V. Djukova, “On the Implementation Complexity of Discrete (Logical) Recognition Procedures,” Zh. Vychisl. Mat. Mat. Fiz. **44** (3), 562–572 (2004) [Comput. Math. Math. Phys. **44**, 532–541 (2004)].
9. E. V. Djukova and I. L. Korneeva, “Models of Recognition Algorithms Based on Various Methods of Conversion of Input Data,” in *Mathematical Methods in Pattern Recognition and Discrete Optimization* (Vychisl. Tsentr Akad. Nauk SSSR, Moscow, 1990).
10. Yu. I. Zhuravlev, “On the Algebraic Approach to the Solution of Recognition and Classification Problems,” Probl. Kibern. **33**, 5–68 (1978).
11. O. V. Sen’ko, “Conversion of Continuous Features into the Discrete Form Based on Optimizing the Approximation Accuracy Functional,” in *Proc. of Int. Conf. on Pattern Recognition and Information Processing, Minsk–Szczecin, Russia–Poland, 1997.*
12. E. V. Djukova, “Discrete (Logical) Recognition Procedures: Principles of Construction, Complexity of Realization, and Basic Models,” J. Pat. Rec. Image Anal. **13** (3), 417–425 (2003).
13. E. V. Djukova, A. S. Inyakin, and N. V.Peskov, “Methods of Combinatorial Analysis in Synthesis of Efficient Recognition Algorithms,” J. Pat. Rec. Image Anal. **13** (3), 426–432 (2003).
14. E. V. Djukova and N. V. Peskov, “Selection of Typical Objects in Classes for Recognition Problems,” J. Pat. Rec. Image Anal. **12** (3), 243–249 (2002).
15. E. V. Djukova and Yu. I. Zhuravlev, “Discrete Methods of Information Analysis and Algorithm Synthesis,” J. Pat. Rec. Image Anal. **7** (2), 192–205 (1997).
16. L. A. Kurgan and K. J. Cios, “CAIM Discretization Algorithm,” IEEE Trans. Knowl. Data Eng. **16** (2), 145–153 (2004).

Djukova Elena V. Born 1945. Graduated from Moscow State University in 1967. Candidate's degree in Physics and Mathematics in 1979. Doctoral degree in Physics and Mathematics in 1997. Dorodnitsyn Computing Center, Russian Academy of Sciences, leading researcher. Moscow State University, lecturer. Moscow Pedagogical University, lecturer. Scientific interests: discrete mathematics and mathematical method of pattern recognition. Author of 70 papers.

Peskov Nikolai V. Born 1978. Graduated from Moscow State University in 2000. Candidate's degree in 2004. Dorodnicyn Computing Center, Russian Academy of Sciences, junior researcher. Scientific interests: discrete mathematics and mathematical methods of pattern recognition. Author of ten papers.

Inyakin Andrey S. Born 1978. Graduated from Moscow State University in 2000. Dorodnitsyn Computing Center, Russian Academy of Sciences, junior researcher. Scientific interests: discrete mathematics and mathematical methods of pattern recognition. Author of ten papers.

Sakharov Aleksei A. Born 1980. Graduated from Moscow State University in 2003. Moscow Pedagogical University, graduate student. Scientific interests: discrete mathematics and mathematical method of pattern recognition. Author of three papers.

SPELL: 1. equilength, 2. equipotential, 3. osteogenic, 4. psychophysical