Classification and Clustering

OPTIMAL DECISION RULES IN LOGICAL RECOGNITION MODELS

Anatol Gupal, Vladimir Ryazanov

Abstract: The task of smooth and stable decision rules construction in logical recognition models is considered. Logical regularities of classes are defined as conjunctions of one-place predicates that determine the membership of features values in an intervals of the real axis. The conjunctions are true on a special no extending subsets of reference objects of some class and are optimal. The standard approach of linear decision rules construction for given sets of logical regularities consists in realization of voting schemes. The weighting coefficients of voting procedures are done as heuristic ones or are as solutions of complex optimization task. The modifications of linear decision rules are proposed that are based on the search of maximal estimations of standard objects for their classes and use approximations of logical regularities by smooth sigmoid functions.

Keywords: precedent-recognition recognition, logical regularities of classes, estimate calculation algorithms, integer programming, decision rules, sigmoid formatting rules

Conference: The paper is selected from International Conference "Classification, Forecasting, Data Mining" CFDM 2009, Varna, Bulgaria, June-July 2009

Introduction

The paper is dedicated to development of recognition algorithms that are based on partial-precedence principle (logical-combinatorial methods, estimate calculation algorithms). The first studies in this field were made by Yu.I.Zhuravlev (a test algorithm [Dmitriev, 1966] , estimate calculation algorithms [Zhuravlev, 1971]), recognition model based on voting over representative sets [Baskakova, 1981]. The well-known practical recognition algorithm Kora has been presented in [Vaintsvaig, 1973]. The basic principle of these algorithms is the search of irredundant fragments of objects descriptions in terms of features that are the incident ones to the classes. Such important fragments are used later for recognition of new objects. These models were elaborated for *k*-valued features. To work with real-valued features, the data quantization is made in advance that preserves the separability of classes on training sample [Zhuravlev, 1978], [Zhuravlev, 1998], [Zhuravlev, 2002], [Dyukova, 2000], [Dyukova, 1989]. Later, the term logical regularity (LR) will be used. The predicate $P(S) = A_1(S) \& A_2(S) \& \ldots \& A_k(S)$ will be understood as logical regularity, where $A_1, A_2, \ldots, A_k$ are one-placed predicates that depend on one of the features and determine the membership of the value of this feature in a certain interval of the real axis. The LR is true for all reference objects of some "no extending" subsets $\tilde{S}^*$ of training sample $\tilde{S}$ belonging to class K_i , moreover $P(S) = 0, \forall S \in CK_i \cap \tilde{S}$.

In [Kochetkov, 1989], recognition algorithms have been proposed that are invariant under some transformations of features, and some practical method for LR search was described [Bushmanov, 1988]. In [Ryazanov, 2007], [Kovshov, 2008], the parametrical approach was considered. The LR is described by vector of binary parameters and LR search is reduced to solution of special integer-valued mathematical programming task. It was proposed relaxation, combinatorial and genetic algorithms for LR search.

This paper is an extension of investigation [Ryazanov, 2007]. Let the sets of LR of all classes have been found by training sample. The LR of minimal complexity and equivalent to some one LR is calculated. To recognize any object S, the weighted sum of values of one-parametric sigmoid approximations of LR for S is calculated. Some restrictions for weight coefficients in terms of equations are used. Finally, the task of construction of stable and smooth decision rule is reduced to linear programming problem. Coefficients of matrix of restrictions are the functions of smooth parameter. The algorithm for construction of stable smooth decision rules have been approved successfully by model and real data.

Main Definitions

We consider the standard recognition task by precedents with n numerical features $x_1, x_2, ..., x_n$, l nonintersecting classes $K_1, K_2, ..., K_l$ and training sample $\tilde{S} = \{S_1, S_2, ..., S_m\}$. We use notation $\tilde{K}_i = \tilde{S} \cap K_i, i = 1, 2, ..., l,$ and suppose that $\tilde{K}_i \neq \varnothing,\ i = 1, 2, ..., l$.

Let $S = (x_1(S), x_2(S), ..., x_n(S))$, $S \in \bigcup_{i=1}^{l} K_i$, $S_t = (a_{t1}, a_{t2}, ..., a_{tn}), a_{tj} = x_j(S_t)$, $x_i \in R$.

The next parametric set of elementary predicates is considered

$$P^{1,c_j}(x) = \begin{cases} 1, & c_j \le x, \\ 0, & \text{otherwise}, \end{cases} \quad P^{2,c_j}(x) = \begin{cases} 1, & x \le d_j, \\ 0, & \text{otherwise} \end{cases}, \text{ where } c_j, d_j \in R, j = 1,2,...,n.$$

Let $\Omega \subseteq \{1,2,...,n\}$.

Definition. The predicate $$P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x}) = \underset{j\in\Omega_1}{\&} P^{1,c_j}(x_j) \underset{j\in\Omega_2}{\&} P^{2,d_j}(x_j) \quad (1)$$

is called a logical regularity of the class K_λ, $\lambda = 1, 2, ..., l$, if it holds that

$\exists\, S_t \in \tilde{K}_\lambda : P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(S_t) = 1$,

$\forall\, S_t \notin \tilde{K}_\lambda$, $P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(S_t) = 1$,

$$\Phi(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x})) = \underset{P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'}(\mathbf{x})}{extr} \Phi(P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'}(\mathbf{x}))$$

Later, we consider the predicate objective function $\Phi(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x})) = \left|\{S_i : S_i \in \tilde{K}_\lambda, P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(S_i) = 1\}\right|$ to be maximized. The set $N(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}) = \{\mathbf{x} \in R^n : c_j \le x_j, j \in \Omega_1, x_j \le d_j, j \in \Omega_2\}$ is called the interval of the predicate $P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x})$. The predicates $P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x})$, $P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'}(\mathbf{x})$ are said to be equivalent if $P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(S_t)$ = $P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'}(S_t), t = 1, 2, ..., m$. Two intervals $N(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}})$, $N(P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'})$ are said to be equivalent if their predicates are equivalent ones.

The feasible predicate $P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x})$ is local-optimal with respect to the criterion $\Phi(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x}))$ if there are no feasible predicates $P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'}(\mathbf{x})$ such that $N(P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'}) \supseteq N(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}})$, $\Phi(P^{\Omega'_1,\Omega'_2,\mathbf{c}',\mathbf{d}'}) > \Phi(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}})$.

Optimization of Logical Decision Rules

Assume that we have the set of LRs $\mathrm{P}_\lambda = \{P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x})\}$ for each class K_λ, and the set of intervals $\{N(P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}) : P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}} \in \mathrm{P}_\lambda\}$ coves $\tilde{K}_\lambda$. The algorithms for finding P_λ have been proposed in [13]. We say that the LR from P_λ has the minimal complicity if there is not any equivalent one that has smaller value of $\Omega_1 + \Omega_2$. Let some LR $P^{\Omega_1,\Omega_2,\mathbf{c},\mathbf{d}}(\mathbf{x}) \in \mathrm{P}_\lambda$ is known. The equivalent LR of minimal complicity is founded as some solution of the following integer linear programming task:

$$\sum_{i\in\Omega_1} y_{1i} + \sum_{i\in\Omega_2} y_{2i} \to \min,$$

$$\sum_{i\in\Omega_1} (1 - P^{1,c_i}(a_{ti}))y_{1i} + \sum_{i\in\Omega_2} (1 - P^{2,d_i}(a_{ti}))y_{2i} \geq 1, \forall S_t \in \tilde{S} \setminus \tilde{K}_\lambda,$$

$$y_{1i} \in \{0,1\}, i \in \Omega_1, y_{2i} \in \{0,1\}, i \in \Omega_2.$$

The unities in y_{1i}, y_{2i} define corresponding subsets Ω_1, Ω_2 for LR to be find. Later, we assume that the sets P_λ consist of LRs of minimal complicity.

The standard approach to recognizing of any object S by estimate calculation algorithms is the following one.

1. The estimation $$\Gamma_j(S) = \sum_{P_t \in \mathrm{P}_j} P_t(S) \qquad (2)$$

is calculated for any object S and class K_j.

2. The standard decision rule $\alpha_j^A(S) = \begin{cases} 1, & \sum_{i=1}^{l} \delta_i^j \Gamma_i(S) \geq \delta_i^0, \\ 0, & \text{otherwise.} \end{cases}$ is used (or the simpler $\alpha_j^A(S) = \begin{cases} 1, & \Gamma_j(S) > \Gamma_i(S), i \neq j, \\ 0, & \text{otherwise.} \end{cases}$).

The notation $\alpha^A_j(S)=1$ ($\alpha^A_j(S)=0$) denotes the solution $S \in K_j$ ($S \notin K_j$) of algorithm A. Parameters δ_i^j are founded in optimization process of recognition model with the use of control sample. The given scheme of recognition has obvious lacks.

1. An arbitrariness in calculation of estimations (2) as result of absence of weight factors of LRs.
2. Graduated character of estimations as functions of signs does not allow estimating stability of a solving rule.
3. Now there are no effective methods of optimization of standard criterion of quality of models of calculation of estimations with use of control sample.

Let's notice that as $\Gamma_j(S_t) > 0, S_t \in \tilde{K}_j$ and $\Gamma_j(S_t) = 0, S_t \in \tilde{S} \setminus \tilde{K}_j$, the algorithm is faultless on objects of the table of training at use of the elementary solving rule. Its extrapolating abilities thus to the user are not known. The following updating resulted above the general scheme of algorithms of calculation of estimations is offered.

Estimations for classes are calculated according to (3)

$$\Gamma_j(S) = \sum_{P_t \in \mathrm{P}_j} \gamma_t f_t\ (S), \tag{3}$$

where $\gamma_t = \gamma_t(P_t^{\Omega_1^t, \Omega_2^t, \mathbf{c}_t, \mathbf{d}_t})$ - the non-negative parameters characterizing "weight" corresponding LT $P_t^{\Omega_1^t, \Omega_2^t, \mathbf{c}_t, \mathbf{d}_t} \in \mathrm{P}_j$, $f_t\ (S)$ - approximating LR $P_t^{\Omega_1^t, \Omega_2^t, \mathbf{c}_t, \mathbf{d}_t}$ sigmoid kind function $f_t\ (S) = \prod_{i \in \Omega_1^t} \frac{1}{1 + \exp(-\delta(x_i(S) - c_{ti}))} \prod_{i \in \Omega_2^t} \frac{1}{1 + \exp(\delta(x_i(S) - d_{ti}))}$.

Classification of S is spent on a maximum of estimations (3). The parameter δ sets «smoothness degree» of LR approximations. Parameters $\gamma_t, t = 1,2,...,N$ (N - total number of logical regularities of all classes) are the solution of the following problem of linear programming:

$$\sigma \to \max, \tag{4}$$

$$\sum_{P_t \in \mathbf{P}_j} \gamma_t f_t\ (S_t) \geq \sigma,\ S_t \in K_j, t = 1,2,...,m, j = 1,2,...,l \tag{5}$$

$$\sum_{i=1}^{N} \gamma_i = N,\ \gamma_i \geq 0,\ i = 1,2,...,N, \tag{6}$$

In a problem (4) - (6) there are such weights factors for LRs of classes at which estimations of standards for classes will be maximum one. Thus, for the set degree of smoothness δ there are weight parameters $\gamma_i,\ i = 1,2,...,N,$ providing steadiest solutions on the training data. The given approach is direct analogue search of the maximum gap in a support vector machine [Burges, 1998]. The algorithm of construction of steady smooth solving rules is successfully approved on the model and real data.

Acknowledgements

The authors are glad to acknowledge support of the following organizations for execution of the described research: RFBR (projects 08-01-90427 ukr, 08-01-00636). The work has been also supported by the Presidium's program N2 of RAS.

Bibliography

[Dmitriev, 1966] A.N.Dmitriev, Yu.I.Zhuravlev, and F.P.Krendelev, "On Mathematical Principles of Classification of Objects and Phenomena", in Discrete Analysis (Institute of Mathematics, Siberian Division, USSR Academy of Sciences, Novosibirsk, 1966), No. 7, pp.3-11 [in Russian].

[Zhuravlev, 1971] Yu.I.Zhuravlev and V.V.Nikiforov, "Recognition Algoithms based on Estimates Calculation", Kibernetika, No.3, pp. 1-11 (1971).

[Zhuravlev, 1978] Yu.I.Zhuravlev, "An Algebraic Approach to Recognition or Classification Problems ", in Problems of Cybernetics , Issue 33 (Nauka, Moscow, 1978; Hafner, 1986), pp.5-68.

[Vaintsvaig, 1973] M.N.Vaintsvaig, "Kora: A Learning Algorithm for Pattern Recognition", in Learning Algorithms for Pattern Recognition (Sovetskoe Radio, Moscow, 1973), pp. 8-12 [in Russian].

[Baskakova, 1981] L.V.Baskakova and Yu.I.Zhuravlev, "A Model of Recognition Algorithms with Representative Sets and Systems of Support Sets", Zh. Vychisl. Mat. Mat. Fiz. 21, pp.1264-1275 (1981).

[Zhuravlev, 1998] Yu.I.Zhuravlev, Selected Works (Magistr, Moscow, 1998) [in Russian].

[Zhuravlev, 2002] Yu.I.Zhuravlev. "Recognition Algorithms with Representative Sets (Logic Algorithms) Algorithms)" Zh. Vychisl. Mat. Mat. Fiz. 42, 1425–1435 (2002) [Comput. Math. Math. Phys. 42, 1372–1382 (2002)].

[Dyukova, 2000] E. V. Dyukova and Yu. I. Zhuravlev, "Discrete Analysis of Feature Descriptions in Recognition Problems of High Dimensionality," Zh. Vychisl. Mat. Mat. Fiz. 40, 1264–1278 (2000) [Comput. Math. Math. Phys. 40, 1214–1227 (2000)].

[Dyukova, 1989] E. V. Dyukova, "The Recognition Algorithms Kora: Complexity of Implementation and Metric Properties," in *Recognition, Classification, and Prediction: Mathematical Methods and Applications* (Nauka, Moscow, 1989), No. 2, pp. 99–125 [in Russian].

[Kochetkov, 1989] D. V. Kochetkov, "Recognition Algorithms that are Invariant under Transformations of the Space of Features," in *Recognition, Classification, and Prediction: Mathematical Methods and Applications* (Nauka, Moscow, 1989), No. 1, pp. 82–113; No. 2, pp. 178–206; No. 3, pp. 64–88 [in Russian].

[Bushmanov, 1988] O.N.Bushmanov, E. V. Dyukova, Yu. I. Zhuravlev, Дюкова Е.В., Yu.I.Zhuravlev, D. V. Kochetkov, V.V.Ryazanov "Program system for data analysis and pattern recognition" in *Recognition, Classification, and Prediction: Mathematical Methods and Applications* (Nauka, Moscow, 1988), No. 2, pp. 250-273 [in Russian].

[Ryazanov, 2007] V.V.Ryazanov, "Logical Regularities in Pattern Recognition (Parametric Approach)" , Zh. Vychisl. Mat. Mat. Fiz. 2007. T.47, № 2.

[Kovshov, 2008] N. V. Kovshov, V. L. Moiseev, and V. V. Ryazanov, "Algorithms for Finding Logical Regularities in Pattern Recognition", Zh. Vychisl. Mat. Mat. Fiz.*, 2008, Vol. 48, No. 2, pp. 314–328.*

[Burges, 1998] Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition, Appeared in: Data Mining and Knowledge Discovery 2, 121-167, 1998.

Authors' Information

Gupal A.M. – Head of Department, Glushkov Institute of Cybernetics NAS Ukraine, Akademision Glushkov st., 40, Kiev, 03680 MCP, Ukraina, e-mail: gupal_anatol@mail.ru

Ryazanov V.V. – Head of Department, Computing Centre of the Russian Academy of Sciences, 40 Vavilova St., Moscow GSP-1, 119991, RUSSIAN FEDERATION, e-mail: rvvccas@mail.ru