ALGORITHM OF CALCULATION OF ESTIMATES IN CONDITION OF FEATURES' CORRELATIONS

Mirzoyan Kamilov[1], Shavkat Fazilov[2], Nomaz Mirzaev[3], and Sobirjon Radjabov[4]

Institute of Mathematics and Information Technologies, Tashkent, Uzbekistan
[1]*kamilov@yandex.ru,* [2]*sh.fazilov@mail.ru,* [3]*mnm2005@rambler.ru,* [4]*s_radjabov@yahoo.com*

Introduction. Research works on pattern recognition have actively been carried out for the past several years. Up to now several types of models have been worked out and investigated, among them [1, 2] could be shown: models based on the principle of division; statistical models; models on the base of the principle of potencials; models on the base of calculation of estimates and etc.

Analysis of these models shows that currently model of algorithms for solving such problems, where objects are described in the uncorrelated (or weakly correlated) features space, are mostly working out. That is why actuality of the problems related to investigating of the models of algorithms for solving object classification problems, where objects are defined in the correlated features space, is increasing.

The goal of this work is to build recognition algorithms considering features' correlations. As an initial model algorithm of calculation of estimates is used [1].

Statement of the problem. **Let** $\{S\}$ - be set of possible objects, which consist of l disjoint subsets $K_1, K_2, ..., K_l$. Knowledge about the division of $\{S\}$ is not complete, only some initial information I_0 about classes $K_1, K_2, ..., K_l$ is given:

$$I_0 = \{S_1, ..., S_i, ..., S_m;\ \ \tilde{\alpha}(S_1), ..., \tilde{\alpha}(S_i), ..., \tilde{\alpha}(S_m)\},$$

where $\tilde{\alpha}(S_i)$ ($\tilde{\alpha}(S_i) = (\alpha_{i1}, ..., \alpha_{ij}, ..., \alpha_{il})$) – informational vector of the object S_i, α_{ij} - value of the predicate $P_j(S_i) = "S_i \in K_j"$.

It is assumed that for any object S ($S \in \{S\}$) n-dimensional vector $\bar{a} = (a_1, ..., a_i, ..., a_n)$ could be matched in the feature space $X = (x_1, ... x_i, ..., x_n)$. The problem is to build algorithm A, which transfers the set (I_0, $\tilde{S}^q$) to informational matrix $\|\beta_{ij}\|_{qxl}$, ($\beta_{ij} = P_j(S_i^{'})$): $A(I_0, \tilde{S}^q) = \|\beta_{ij}\|_{qxl}$, $\beta_{ij} \in \{0, 1, \Delta\}$. Here $\tilde{S}^q$ - the set of control objects ($\tilde{S}^q = \{S_1^{'}, ..., S_q^{'}\}$, $\tilde{S}^q \in \{S\}$). Value β_{ij} is also interpreted as in [1].

Solving method*.* One of the approaches to solving the problem of building of recognition algorithm in condition of features' correlations is considered in this work. Model of modified recognition algorithms based on the estimating features' correlations and extracting preferred model of correlations for each subset of strongly correlated features is worked out on the base of this approach.

We consider the main stages of building of calculation of estimates type recognition algorithms based on the estimating of features' correlations:

1. *Extracting subsets of strongly correlated features*. The first stage in defining recognition operator B in the model of recognition algorithms based on calculation of estimates is determining the system W_A of "uncorrelated" subsets.

2. *Forming of the set of representative features*. In this stage the search for representative features is carried out. In the process of representative features' formation it is required that each extracted feature be typical representative of the extracted subset of the strongly correlated features. After completing this stage we get reduced features space, dimension of which is much smaller than initial space ($n^{'} < n$). From now on we define formed features space by Y ($Y = (y_1, ..., y_{n^{'}})$).

3. *Determining the models of correlations in each subset of features for the class* $K_j\ (j=\overline{1,l})$. Let x_i be any feature belonging to the subset Ω_q. It is assumed that elements of Ω_q are linearly ordered by the features' indexes ($x_i < x_j$, if $i<j$). Initial element (x_0) of the subset Ω_q is y_q, remaining elements are defined by x_i ($N_q = card(\Omega_q)$; $i=1,...,N_q-1$). Then model of correlations in Ω_q could be defined as

$$x_i = F(\bar{c},\ y_q),\quad x_i \in \Omega_q \setminus y_q,$$

where $\bar{c}$ is vector of unknown parameters, F- function from some given class $\{F\}$.

Calculating the values of the vector of unknown parameters $\bar{c}$ defines the model of correlations in the features subset Ω_q for the class $K_j\ (j=\overline{1,l})$. Depending on the parametrical type of $F(\bar{c},\ x)$ and method of determining of $\bar{c}$ we obtain different models of correlations in the features' set $\Omega_q\quad (q=\overline{1,n'})$.

4. Extracting preferred models of correlations. Let N_q be the power of the subset Ω_q of strongly correlated features. It is assumed that (N_q-1) models of correlations is defined in Ω_q for the class K_1:

$$x_i = F(\bar{c}, y_q),\quad x_i \in \Omega_q \setminus y_q,\ i=\overline{1,(N_q-1)},$$

where y_q is representative feature ($y_q \in \Omega_q$).

We introduce following definitions: $E_1 = I_0 \cap K_1$, $E_2 = I_0 \cap K_2$. The search for the preferred models of correlations in Ω_q is carried out on the base of estimating dominancy of the considering models for the objects belonging to the set I_0 [3]:

$$T_i = \frac{L_2 \sum_{s\in E_2} (x_i - F(\bar{c}, y_q))^2}{L_1 \sum_{s\in E_1} (x_i - F(\bar{c}, y_q))^2},\quad L_1 = card(E_1),\quad L_2 = card(E_2).$$

How bigger T_i is, more preference is given to i- model of correlations. If several models have same preference, any of them could be selected.

After completing this stage preferred model of correlations for the features' subset Ω_q is determined, and it is defined as $x_i = F_q(\bar{c}, y_q)$. Hereinafter only these models of correlations are considered.

Next stages are closely related to the algorithms of calculation of estimates [1, 4].

5. Defining the system of supporting sets. Let $H_{\tilde{\omega}}$ be all the possible subsets of the set $\{y_1, y_2, ..., y_{n'}\}$. We define collection of such subsets by Ω. Fifth stage of the modified recognition algorithms based on the calculation of estimates is defining the system of supporting sets Ω_A ($\Omega_A \subseteq \Omega$).

6. Defining the distance function between objects. Let's consider possible objects S and S_u. In the sixth stage of the modified recognition algorithms based on the calculation of estimates distance function $\mu_{\tilde{\omega}}(S, S_u)$ between objects S and S_u in the $\tilde{\omega}$-part of the features' space is defined.

7. Calculating of estimates on the objects of fixed supporting set. In the seventh stage of the modified recognition algorithms based on the calculation of estimates numerical characteristic called estimate $\Gamma_{\tilde{\omega}}(S, S_u)$ is calculated:

$$\Gamma_{\tilde{\omega}}(S,S_u) = \lambda_u \mu_{\tilde{\omega}}(S,S_u),$$

where λ_u is given parameter of the algorithm.

8. Calculating of estimates for the class on the fixed supporting set. Lets assume that values $\Gamma_{\tilde{\omega}}(S,S_u)$ ($S_u \in \tilde{K}_j$) were calculated. Estimate for the class is determined as:

$$\Gamma_{\tilde{\omega}}(S,K_j) = \sum_{S_u \in \tilde{K}_j} \gamma_u \Gamma_{\tilde{\omega}}(S,S_u).$$

Here γ_u – given parameter of the algorithm.

9. Estimate for the class K_j on the system of supporting sets. Let numerical parameter $\tau(\tilde{\omega})$ corresponds for each vector $\tilde{\omega}$. Estimate on the system of supporting sets Ω_A is defined as

$$\Gamma(S,K_j) = \sum_{\tilde{\omega} \in \Omega_A} \tau(\tilde{\omega}) \Gamma_{\tilde{\omega}}(S,K_j).$$

10. *Decision rule.* In the last stage of the modified recognition algorithms based on the calculation of estimates decision rule is defined as [1]:

$$\beta_{ij} = C(\Gamma(S_i,K_j)) = \begin{cases} 0, \text{ if } \ \Gamma(S_i,K_j) < c_1; \\ 1, \text{ if } \ \Gamma(S_i,K_j) > c_2; \\ \Delta, \text{ if } \ c_1 \le \Gamma(S_i,K_j) \le c_2, \end{cases}$$

where c_1, c_2 are parameters of the algorithm.

We defined class of the modified recognition algorithms based on the calculation of estimates. Any algorithm *A* from this model is fully defined by the set of parameters $\pi = (n', k, \{\varepsilon_i\}, \{\lambda_u\}, \{\gamma_u\}, \{\tau(\tilde{\omega})\}, c_1, c_2)$. We define the collection of all recognition algorithms form the proposed model by $A(\pi, S)$. The search for the best algorithm is carried out in the parameters' space π.

The difference between proposed algorithms and traditional recognition algorithms based on the calculation of estimates is that proposed algorithms are based on the estimating of features correlations. That is why these algorithms are used when there are some correlations among features. Obviously, this correlation should be different for each class. This makes it possible to describe objects of each class by different models.

If features are weakly correlated, classical model of the recognition algorithms (for example, model discussed in [1, 4]) are used. Therefore proposed model of recognition algorithms is not alternative to the calculation of estimates based models, but it is only complementary to them.

Experimental results. The proposed algorithms were tested by experimental investigations on solving model problems.

There are 2 classes in the experiments. There are 200 samples (100 for each class) in the training set. There are 200 samples (100 for each class) in the test set. The number of features is 110. The number of strongly correlated features is 10.

The average effectivity estimates of recognition algorithms for model problems are calculated for all experiments. During the experimental investigations 10 experiments were carried out.

These problems were solved by using both the classical calculation of estimates algorithm [5] and modified calculation of estimates algorithm.

All correlated features were found and the effective recognition algorithm based on these features was built. The following recognition results were achieved during the experiments: average recognition errors during training and testing were 0% and 1,5% respectively. These results for the classical calculation of estimates algorithm were 12,5% and 24% respectively.

Comparative analysis of these algorithms shows that by using modified calculation of estimates algorithm 197 objects form 200 were correctly recognized and recognition rate was 98,5%. This is 22% higher than that of using the classical calculation of estimates algorithm. This could be explained that features' correlations are not considered in the classical calculation of estimates algorithm.

Conclusion. Model of recognition algorithms based on features correlations, which improve recognition rate and application areas in solving practical problems was introduced. As an initial model for pattern recognition model of algorithms of calculation of estimates was considered. This model of algorithms reduces considerably the number of calculations in recognizing unknown object and it could be used in different software packages for solving problems of forecasting and recognition of objects in correlated features' space.

References

1. Yu.I. Zhuravlev. Selected papers. (in Russian) "Magistr" Publishing House, Moscow (1998) 420 p.
2. M.M. Kamilov, N.M. Mirzaev, S.S. Radjabov. Urgent problems of building of models of recognition algorithms. (in Russian) Chemical technology. Control and management, Tashkent (2009) pp. 67-72.
3. N.M. Mirzaev. Mean distance type recognition algorithms based on the features correlations. (in Russian) Theoretical and practical problems of modern information technologies: Proceedings of 10th All-Russian scientific-technical conference, Ulan-Ude (2009) pp. 418 - 423.
4. Yu.I. Zhuravlev, V.V. Ryazanov, O.V. Senko. Recognition. Mathematical methods. Software system. Practical application. (in Russian) "Fasiz" Publishing House, Moscow (2006) 159 p.
5. Yu.I. Zhuravlev, M.M. Kamilov, Sh.E. Tulyaganov. Calculation of estimates algorithms and their application. (in Russian) "Fan" Publishing House, Tashkent (1974) 119 p.