Selection of Significant Features of Objects in the Classification Data Processing

Fazilov Shavkat, Mamatov Narzillo, Samijonov Abdurashid

Abstract: The formation of a characteristic space in classification problems can be divided into two stages: the choice of the initial description of objects and the formation of an informative description of objects on the basis of a reduction in the dimension of the space of the original description

Keywords: Object Classification, data Processing.

I. INTRODUCTION

One of the key issues of computer processing of information is the formation of an informative description of those objects that need to be identified, classified or recognized. The formation of informative descriptions, the detection of significant features of objects predetermines, ultimately, the success of the entire study, allows to achieve two usually mutually exclusive goals: improving the quality of recognition or classification (by eliminating "noisy" parameters) and reducing processing time (by reducing the amount of data) [1].

Forming an informative description of objects in data analysis tasks is essentially a transition from the original feature system $x = \left(x^1, x^2, \ldots, x^N \right)$ to the new $z = \left(z^1, z^2, \ldots, z^{\ell} \right)$ system, which includes a smaller number of features $\left(\ell < N \right)$ than the original system [4-6]. Usually, new features are formed as functions of the original features, i.e. $z = F(x)$, by solving an optimization problem. The latter consists in finding such a system of signs z, for which

$$I(\tilde{z}) = \max_{F \in \Omega} \{I(z)\}. \qquad (1)$$

Here $I(z)$ is a given measure of informatively ℓ - a dimensional system of features z, and F is the class of permissible transformations of the original features $x^1, x^2, \ldots, x^N$, which in general can be represented by one of the following types of transformations: linear, nonlinear discrete, continuous, logical.

Thus, the formation of an informative description of objects can be interpreted as a mapping of an N-dimensional vector x into ℓ —the dimensional vector z, which in the general case can be represented as $z = F(x)$, where F is a valid transformation, and $\ell \le N$.

The report uses the system of functions as permissible transformations.

$$z^i = f_i \left(x^i \right); \ i = \overline{1, N}, \qquad (2)$$

$$f \left(x^i \right) = \begin{cases} 0, \text{ if the feature } x^i \text{ is excluded;} \\ 1, \text{ if the feature } x^i \text{ is left.} \end{cases}$$

In this case, the new feature system is formed as a subset of the set of source features.

As a measure of the informativeness of features, heuristic criteria are used based on the assessment of the measure of separability of objects of a given training sample using the Euclidean metric [3, 7].

Suppose a training set is specified by objects $x_{11}, x_{12}, \ldots, x_{1m_1}, x_{21}, x_{22}, \ldots, x_{2m_2}, \ldots, x_{r1}, x_{r2}, \ldots, x_{rm_r}$, for which it is known that each group of objects $x_{p1}, x_{p2}, \ldots, x_{pm_p}$ belongs to a certain class $X_p, p = \overline{1, r}$*.

Each object x_{pi} is an N-dimensional vector of numerical features, i.e. $x_{pi} = \left(x_{pi}^1, x_{pi}^2, \ldots, x_{pi}^N \right)$.

For a given training sample of objects $x_{p1}, x_{p2}, \ldots, x_{pm_p} \in X_p, \ p = \overline{1, r}$, where x_{pi} is a vector in the N-dimensional feature space, we introduce the vector $\lambda = \left(\lambda^1, \lambda^2, \ldots, \lambda^N \right)$, $\lambda^k \in \{0;1\}, \ k = \overline{1, N}$, which uniquely characterizes a certain subsystem of signs. The components of the vector λ, equal to one, indicate the presence of corresponding signs in this subsystem, and zero components indicate the absence of corresponding signs.

The space of attributes $\left\{ x = \left(x^1, x^2, \ldots, x^N \right) \right\}$ will be considered Euclidean and denoted by R^N.

Definition. Let's call the vector λ ℓ - informative if the sum of its components is equal to ℓ, i.e.

$$\sum_{i=1}^{N} \lambda^i = \ell.$$

For each subsystem defined by ℓ - informative vector λ, its ℓ - dimensional characteristic subspace is defined. In each of these subspaces, we introduce the Euclidean norm with respect to truncation by λ

Fazilov Shavkat[1], Mamatov Narzillo[2], Samijonov Abdurashid[3]
[1,2]Scientific and Innovation Center of Information and Communication Technologies at TUIT named after Al-Kharezmi, Tashkent, Uzbekistan

$$\|x\|_{\lambda} = \sqrt{\sum_{j=1}^{N} \lambda^j \left(x^j\right)^2} .$$

Denote

$$\bar{x}_p = \frac{1}{m_p} \sum_{i=1}^{m_p} x_{pi}, \; p = \overline{1, r},$$

where $\bar{x}_p$ is the average class object X_p.

We introduce the function

$$S_p(\lambda) = \sqrt{\frac{1}{m_p} \sum_{i=1}^{m_p} \left\|x_{pi} - \bar{x}_p\right\|_{\lambda}^2} .$$

The $S_p(\lambda)$ function characterizes the average scatter of class X_p objects in a subset of features specified by the λ vector. Set the criterion informative subsystems in the form of a functional

$$I(\lambda) = \frac{\sum_{p,q=1}^{r} \left\|\bar{x}_p - \bar{x}_q\right\|_{\lambda}^2}{\sum_{p=1}^{r} S_p^2(\lambda)} . \qquad (3)$$

Denote

$$a = \left(a^1, a^2, ..., a^N\right); \; b = \left(b^1, b^2, ..., b^N\right),$$

$$a^j = \sum_{p,q=1}^{r} \left(\bar{x}_p^j - \bar{x}_q^j\right)^2, \; j = \overline{1, N};$$

$$b^j = \sum_{p=1}^{r} \left(\frac{1}{m_p} \sum_{i=1}^{m_p} \left(\bar{x}_{pi}^j - \bar{x}_p^j\right)^2 \right), \; j = \overline{1, N}.$$

Then the functional (3) is reduced to the form

$$I(\lambda) = \frac{(a, \lambda)}{(b, \lambda)}, \qquad (4)$$

where (*, *) is the scalar product of vectors.

a^j, b^j coefficients do not depend on λ and are calculated in advance. To calculate the functional $I(\lambda)$ for each λ required about N operations.

II. STATEMENT OF A PROBLEM AND THE CONCEPT OF THE PROBLEM DECISION

Consider an optimization problem

$$\begin{cases} I(\lambda) = \dfrac{(a, \lambda)}{(b, \lambda)} \to \max, \\ \lambda \in \Lambda^l, \lambda_i = \{0,1\}, \; i = \overline{1, N}, \\ a, b \in R^N, \; a_i \geq 0, b_i > 0, \; i = \overline{1, N}, \end{cases}$$

$$(5)$$

where Λ^l is ℓ - dimensional informative space of signs:

$$\Lambda^l = \left\{ \lambda \middle| \lambda_i = \{0,1\}, i = \overline{1, N}, \sum_{i=1}^{N} \lambda_i = l \right\}.$$

To solve the problem (5), we introduce the vector function

$$\phi(\lambda) = a(b, \lambda) - b(a, \lambda), \qquad (6)$$

which indicates the direction of the fastest growth of the functional $I(\lambda)$ at the λ point.

Theorem 1. If λ and μ - are two ℓ - informative vectors and $b^j > 0$, $j = \overline{1, N}$, then $I(\lambda) < I(\mu)$ then and only if $(\phi(\lambda), \mu) > 0$.

We introduce the operator (following)

$$\mu : \Lambda^{\ell} \to \Lambda^{\ell}$$

such that

$$(\phi(\lambda), \mu(\lambda)) = \max_{\eta \in \Lambda^{\ell}} (\phi(\lambda), \eta).$$

The operator μ has an obvious constructive representation. If we order the components of the vector $\phi(\lambda)$, i.e. find a set of pairwise distinct indices $j_1, j_2, ..., j_N$ such for which $\phi^{j_1}(\lambda) \geq \phi^{j_2}(\lambda) \geq ... \phi^{j_N}(\lambda)$, the components of the vector $\mu(\lambda)$ will be defined as

$$\mu^{j_1}(\lambda) = 1, \mu^{j_2}(\lambda) = 1, ..., \mu^{j_{\ell}}(\lambda) = 1, \mu^{j_{\ell+1}}(\lambda) = 0, \mu^{j_{\ell+2}}(\lambda) = 0, ..., \mu^{j_N}(\lambda) = 0$$

.

In other words, the components of the vector $\mu(\lambda)$, corresponding to the first ℓ - maximum components of the vector $\phi(\lambda)$, are equal to one, the rest are zero.

Obviously, $\mu(\lambda)$ is also an ℓ -informative vector, and

$$(\phi(\lambda), \mu(\lambda)) = \max\left\{ (\phi(\lambda), \eta) \middle| \eta \in \Lambda^{\ell} \right\}.$$

$$(7)$$

Property 1. For an arbitrary $\lambda(\lambda \in \Lambda^{\ell})$ true $(\phi(\lambda), \mu(\lambda)) \geq 0$.

The main corollary follows from the above Theorem 1 and Property 1.

$$I(\lambda) \leq I(\mu(\lambda)), \lambda \in \Lambda^{\ell}.$$

Theorem 2. If $I(\lambda) = I(\mu(\lambda))$, then

$$I(\lambda) = \max\left\{ I(\eta) \middle| \eta \in \Lambda^{\ell} \right\}.$$

Note that Theorem 2 guarantees the optimality of the solution obtained, i.e. the value of the functional $I(\lambda)$ at the found solution λ reaches its maximum on the set Λ^{ℓ}.

The proposed method is based on Theorems 1 and 2 and is implemented as an iterative procedure. And at the first step, an arbitrary ℓ is selected — an informative vector λ, for example, $\lambda = \left(\overbrace{1,1,...,1}^{\ell}, 0, 0, ..., 0 \right)$.

Further, at each iteration, the new vector λ is determined from the previous one with the help of the following operator $\mu(\lambda)$, i.e. simply assignment $\lambda = \mu(\lambda)$. The iteration process continues until the functional $I(\lambda)$ grows. In case growth stops, i.e. $I(\lambda) = I(\mu(\lambda))$, λ is the optimal solution. Usually this solution is achieved, as experiments have shown, at step 3–4.

We formulate the method in question as an algorithm:

Step 1. Input parameters are the values: ℓ - the required number of signs; N is the number of features; a and b are N-dimensional vectors.

Step 2. Setting the vector $\lambda = \left(\underbrace{\overbrace{1,1,...,1}^{\ell},0,0,...,0}_{N} \right)$.

Step 3. Calculate the value of the functional $I(\lambda)$.

Step 4. Calculate $\lambda_0 = \mu(\lambda)$ (the components of the vector $\mu(\lambda)$ are defined as a separate block).

Step 5. Calculate the value of the functional $I_0 = I(\lambda_0)$.

Step 6. Comparing $I_0 > I(\lambda)$. If the inequality holds, then we put $\lambda = \lambda_0$, $I(\lambda) = I_0$ and go to step 4. Otherwise, the procedure ends and λ is the optimal value.

Step 7. Output parameters: λ, $I(\lambda)$.

The computation unit $\lambda_0 = \mu(\lambda)$ provides the following successive operations:

Step 1. Input parameters: λ, ℓ, N, a, b.

Step 2. Calculate the vector $\phi(\lambda) = a(b,\lambda) - b(a,\lambda)$.

Step 3. Setting $\mu = (0,0,...,0), k = 1$.

Step 4. Setting $\phi_{max} = 10^{-60}, i = 1$.

Step 5. Checking the i-component of the vector μ. If $\mu_i = 0$, then go to the next step, otherwise - to step 8.

Step 6. Checking the i-components of the vectors ϕ_{max}, ϕ.
With $\phi_{max} < \phi$, the transition to the next step, otherwise - to step 8.

Step 7. Assigning $\phi_{max} = \phi_i$, $m = i$.

Step 8. $i = i + 1$.

Step 9. If $i \leq N$, then go to step 5. Otherwise, go to the last step.

Step 10. Assigning $\mu_m = 1, k = k + 1$.

Step 11. If $k \leq \ell$, then go to step 4, otherwise $\lambda_0 = \mu$ and the procedure ends.

In conclusion, we note that the results obtained apply to all the criteria for informativeness of features given by functionals, which in principle can be reduced to form (4).

Suppose that the components of vector a are positive and the vectors b are strictly greater than zero, i.e.

$$a = (a^1, a^2, ..., a^N); \ a^j \geq 0, \ b = (b^1, b^2, ..., b^N); \ b^j > 0, \ j = \overline{1, N}$$
,

а также

$$\frac{a^1}{b^1} \geq \frac{a^2}{b^2} \geq ... \geq \frac{a^N}{b^N}. \qquad (8)$$

On the set Λ^ℓ, consider the problem (5).

From the definition of the set of ℓ-informative vectors Λ^ℓ it follows that the cardinality of the considered set is equal to C_N^ℓ. From this it follows that in order to find the maximum value of the functional $I(\lambda)$ for a given ℓ using the brute force method [9], calculations are performed C_N^ℓ times. Hence we have that for every given ℓ there is an ℓ-informative vector $\lambda_\ell \in \Lambda^\ell$, providing

$$I(\lambda_\ell) = \max \lambda \in \Lambda^\ell, \qquad (9)$$

where $\ell = \overline{1, N}$.

Regarding the function $I(\lambda_\ell)$, $\ell = \overline{1, N}$, the result is known, which we use in the future.

For all ℓ $\left(\ell = \overline{1, N} \right)$, the numbers a^1 and b^1 are involved in the values of the functional $I(\lambda_\ell)$, that is, if inequalities (8) are used as the initial data, then the y ℓ-optimal vector is the first coordinate equal to one.

Using this result, to solve the problem (5), we can propose a search method, according to which the maximum of the function $I(\lambda)$ is found over all $\lambda = (\lambda^1, \lambda^2, ..., \lambda^N)$, $\lambda^k \in \{0,1\}$, for which $\sum_{k=1}^{N} \lambda^k = \ell$.

To organize an efficient search, strictly order all possible ℓ-informative vectors λ. First, in this order, we will consider a vector having the first ℓ components equal to one, the others - zero. And for the last vector, the last ℓ components are equal to one, the rest ones are zero.

Let some $\lambda_r = (\lambda_r^1, \lambda_r^2, ..., \lambda_r^N)$-informative vector be given, which is in a certain place of the considered order. We determine the rule for choosing the immediately following vector $\lambda_{r+1} = (\lambda_{r+1}^1, \lambda_{r+1}^2, ..., \lambda_{r+1}^N)$.

If
$$\lambda_r^1 = *, \ \lambda_r^2 = *,..., \lambda_r^k = *, \lambda_r^{k+1} = 1, \lambda_r^{k+2} = 0, \lambda_r^{k+3} = 0,..., \lambda_r^N = 0$$
, then
$$\lambda_{r+1}^1 = \lambda_r^1, \ \lambda_{r+1}^2 = \lambda_r^2,..., \lambda_{r+1}^k = \lambda_r^k, \lambda_{r+1}^{k+1} = 0, \lambda_{r+1}^{k+2} = 1, \lambda_{r+1}^{k+3} = \lambda_r^{k+3},..., \lambda_{r+1}^N$$
.

If
$$\lambda_r^1 = 1, \ \lambda_r^2 = 1,..., \lambda_r^k = 1, \lambda_r^{k+1} = 0, \lambda_r^{k+2} = 0,..., \lambda_r^j = 0, \lambda_r^{j+1} = 1,..., \lambda_r^N =$$
, then
$$\lambda_{r+1}^1 = \lambda_r^1, \ \lambda_{r+1}^2 = \lambda_r^2,..., \lambda_{r+1}^{k-1} = \lambda_r^{k-1}, \lambda_{r+1}^{k+1} = 1, \lambda_{r+1}^{k+2} =$$
$$= 1,..., \lambda_{r+1}^{k+N-j+1} = 1, \ \lambda_r^{k+N-j+2} = 0,..., \lambda_{r+1}^N = 0$$
.

Here the sign "*" means "zero" or "one".

Schematically, this rule is represented as

$$\left(*, *, ... *, 1, 0, 0, ..., 0 \right), \left(*, *, ..., *, 1, 0, 0, ..., 0, 1, 1, ..., 1 \right)$$

.

In [7], it was proved that the above rule guarantees the enumeration of all possible options, starting with the first and ending with the last.

The principle of operation of this algorithm is used to solve the problem (1). Only in this case, as the last vector λ, we take the vector $\lambda = (0,1,1,...1,0,0,...,0)$, not $\lambda = (0,0,...,0,1,1,...1)$.

The number of calculations of the functional $I(\lambda)$ in this method for a given $\ell \geq 2$ is equal to

$$C_N^\ell - C_{N-1}^\ell = \frac{N!}{(N-\ell)!\ell!} - \frac{(N-1)!}{(N-\ell-1)!\ell!} = \frac{\prod\limits_{i=1}^{\ell}(N-\ell+i) - \prod\limits_{i=1}^{\ell}(N-\ell-1+i)}{\ell!}$$

$$= \frac{\ell \prod\limits_{i=1}^{\ell-1}(N-\ell+i)}{\ell!}.$$

The proposed algorithm gives the optimal result for solving the problem (1), and the number of iterations of calculating the functional in this method is significantly less than in the exhaustive search method [8].

From the results of the proposed scheme, it is easy to formulate a simple algorithm for maximizing the criterion of informativeness of feature subsystems by the method of partial enumeration. This algorithm is implemented as follows:

Step 1. We assume $\lambda = \left(\underbrace{1,1,...1}_{\ell},0,0,...,0 \right)$.

Step 2. Calculation $I(\lambda)$.

Step 3. We assume $I_{max} = I(\lambda), \lambda_{max} = \lambda$.

Step 4. Determination of $\lambda = Q(\lambda)$ (according to the following rule).

Step 5. Calculation $I(\lambda)$.

Step 6. Check: if $I_{max} < I(\lambda)$, then assign $I_{max} = I(\lambda), \lambda_{max} = \lambda$. Otherwise, proceed directly to the next step.

Step 7. Check: if $\lambda = \left(0,\underbrace{1,1,...1}_{\ell},0,0,...,0 \right)$, then the procedure ends and λ_{max} sets the best subsystem of features. Otherwise, go to step 4.

The proposed algorithms were tested on examples of solving a number of practical problems, one of which is related to the evaluation of the operational state of a steeply inclined conveyor installed in a quarry of a mining and metallurgical complex.

This conveyor may be in one of three states: dangerous, pre-emergency, safe. The signals characterizing the dangerous state of the complex were the first class of objects (K_1), the signals characterizing the pre-emergency state - the second class of objects (K_2), the signals characterizing the safe state - the third class of objects (K_3). The number of signs characterizing each object was 9.

Each class contains the same number of objects, equal to 5056. Thus, each object can be represented as a vector $x_{ij} = (x_{ij}^1, x_{ij}^2, ... , x_{ij}^9)$, where x_{ij}^k is the $k-$ sign of the $j-$ th object of the $i-$ th class, where $k = \overline{1,9}; j = \overline{1,5056}; i = \overline{1,3}$.

The task of determining the main indicators characterizing the state of the pipeline, was reduced to solving the optimization problem

$$\begin{cases} I(\lambda) = \dfrac{(a,\lambda)}{(b,\lambda)} \to opt, \\ \lambda \in \Lambda^\ell. \end{cases}$$

After determining the ℓ-informative feature sets ($\ell = \overline{1,9}$), the recognition problem was solved using the "k-nearest neighbors" method, which allowed us to estimate the degree of "usefulness" of each of these feature sets in terms of their influence on the recognition quality of control sample objects.

As a result of solving this problem, the most informative set of features x_2, x_5, x_8 was determined, where x_2, x_5 and x_8 represent the values of the signals coming from the vertical components of three-component sensors. On the basis of these results, specialists in the subject area developed recommendations for preventing emergencies during the operation of a conveyor complex.

III. CONCLUSION

The task of reducing the dimension of the original feature space is reduced to an optimization problem, which consists in finding the features in the N-dimensional source system, such an l-dimensional (l <N) feature subsystem, which provides an extreme value of a given measure of informativeness of features. The solution of this optimization problem serves as an informative description of objects with the smallest number of features.

For the criterion of informative Fisher proposed methods for the selection of informative features. It is shown that for solving the optimization problem of selecting informative features a regular method uses a vector function that indicates the direction of the fastest growth of the functional. For the case of the fulfillment of the ordering condition, a recurrent method was developed, which allows, in contrast to the regular method, to obtain an optimal solution in a fixed number of steps. This makes it possible to solve the optimization problem by the method of partial enumeration.

REFERENCES

1. Ayvazyan S.A., Buchstaber V.M., Enyukov I.S., Meshalkin L.D. Applied statistics: Classification and reduction of dimension. - M .: Finance and Statistics, 1989. - 607 p.
2. Aleksandrov V.V., Fazylov Sh.H. Integrated approach to data analysis // Automation of research based on information-computing network. –L .: LNIVTS USSR Academy of Sciences. 1984. -C. 4-10.
3. Gorelik A.L., Skripkin V.A. Building recognition systems. - M .: Sov.radio, 1974. - 224 p.
4. Zhuravlev Yu.I., Kamilov M.M. Tulyaganov Sh.E. Algorithms for calculating estimates and their application. - Tashkent: Fan, 1974. - 124 p.
5. Kutin G.I. Methods of ranking complexes of features. Review // Foreign Radioelectronics. -1981, -№9. -WITH. 54-70.

6. Methods, criteria and algorithms used in the conversion, selection and selection of features in data analysis / K.A.Cheponis, D.A. articles. - Vilnius, 1988. - 150 p.
7. Fazilov Sh.X., Mamatov N.S. Developing methods and algorithms for forming of informative features' space on base k-type uniform criteria// Eighth World Conference "Intelligent Systems for Industrial Automation", WCIS-2014,25-27 November 2014, Tashkent, Uzbekistan.
8. Fazilov Sh.X., Mamatov N.S. Selection features using heuristic criteria// Ninth World Conference "Intelligent Systems for Industrial Automation", WCIS-2016,25-27 October 2016, Tashkent, Uzbekistan.
9. Маматов Н.С. Mamatov N.S. Heuristic criteria for the informativeness of signs. Materials of the XVII International Scientific and Methodical Conference "Informatics: Problems, Methodology, Technologies". -Voronezh. 9-10 February 2017. Volume 3, P.114-120.

.