

10-19-2018

Formation an informative description of objects.

Sh.X Fazilov

Science and innovation center for information and communication technologies at the Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Address: Амир Темур, 108, 100095, Tashkent city, Republic of Uzbekistan

N.S Mamatov

Science and innovation center for information and communication technologies at the Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Address: Амир Темур, 108, 100095, Tashkent city, Republic of Uzbekistan, m_narzullo@mail.ru

Follow this and additional works at: <https://uzjournals.edu.uz/ijctcm>

 Part of the [Engineering Commons](#)

Recommended Citation

Fazilov, Sh.X and Mamatov, N.S (2018) "Formation an informative description of objects.," *Chemical Technology, Control and Management*: Vol. 2018 : Iss. 3 , Article 22.

DOI: <https://doi.org/10.34920/2018.4-5.94-102>

Available at: <https://uzjournals.edu.uz/ijctcm/vol2018/iss3/22>

This Article is brought to you for free and open access by 2030 Uzbekistan Research Online. It has been accepted for inclusion in Chemical Technology, Control and Management by an authorized editor of 2030 Uzbekistan Research Online. For more information, please contact sh.erkinov@edu.uz.

Formation an informative description of objects.

Cover Page Footnote

Tashkent State Technical University, SSC «UZSTROYMATERIALY», SSC «UZKIMYOSANOAT», JV «SOVPLASTITAL», Agency on Intellectual Property of the Republic of Uzbekistan

**FORMATION AN INFORMATIVE DESCRIPTION OF OBJECTS****Sh.X.Fazilov¹, N.S.Mamatov²**

^{1,2} *Science and innovation center for information and communication technologies at the Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi*
Address: Amup Temyp, 108, 100095, Tashkent city, Republic of Uzbekistan
E-mail: ¹m_narzullo@mail.ru

Abstract: *In the article considered the problem of formation an informative description of objects in classification tasks. To solve this problem, proposed the methods and algorithms based on heuristic criteria of private and generalized species.*

Keywords: *features space, pattern recognition, informative features set, optimal solution.*

Introduction

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.

At the first stage, the initial system of characteristics is selected, useful, in one way or another, for dividing the given alphabet of images, by which it is possible to obtain a priori information necessary for describing images in the language of these characteristics. This stage is the least developed in the problems of data analysis, where there are currently no formalized methods for its implementation [1]. In determining the initial system of attributes, a priori knowledge, intuition and experience of specialists in the relevant subject area are widely used. In this case, one should also take into account the important circumstance related to the fact that each real object represents an infinite number of different properties that reflect its sides [2]. Naturally, in

each specific case, not all properties are important, but only their limited set, which determines the characteristics that really allow classification. Finding such characteristics requires always a careful study of the substantive essence of the classified objects with the use of experimental data on the properties of these objects under consideration. To solve this problem, software tools for analyzing data, in particular, the means for exploratory analysis, extraction of knowledge and verification of various attribute systems, may prove useful. In this case, structural methods of data processing [3, 4] can be of great help, under which the structural interrelations of geometric configurations of points-objects in the multidimensional description space are investigated. The analysis of the data structure helps the researcher to understand what properties of objects contribute to the separation of images, to evaluate the informativeness of individual features.

The purpose of the second of the above stages is to determine the most useful set of characteristics for the objects under study. The need to implement this stage is due to the following circumstance.

When the initial system of signs is chosen, then, as a rule, it turns out to be very excessive. There are arguments "pros" and "cons" regarding the preservation of such redundancy. The argument "pros" is that the increase in the number of attributes allows us to more fully describe the objects. The "cons" argument: increasing the

number of attributes increases the "noise" in the data, makes processing more difficult and leads to additional time costs for its implementation.

Consequently, the argument "pros" comes mainly from statistical assumptions, where as the argument "cons" is repelled mainly from the non-statistical. If practical motives are important almost always, then the conditions when statistics work, turn out to be performed much less often than expected. In [5] such criteria of applicability of statistical methods:

1) it is possible to repeat the experiment many times under the same conditions;

2) the outcome of the experiment can not be predicted in advance because of the influence of a large number of random factors;

3) with increasing number of experiments, the results converge to certain values.

And the authors of [5] note that strict mathematical methods that allow to verify the fulfillment of these conditions do not exist in the concrete case. They highlight sociology, demography, reliability theory and selective quality control as areas where these conditions are met in most cases. Very often, however, they are violated - in whole or in part - usually due to the fact that the second part of the criterion is not fulfilled, i.e. the same experimental conditions are not observed.

In connection with the search for the answer to the question: how many objects need to be taken under the conditions of the statistical ensemble and how many characteristics should be measured (from the statistical point of view, rather than the subject domain) to obtain the result with a given accuracy, it is advisable to refer to the results of studies on estimating the recognition error different volumes of the training selection m and the number of features N . Let us draw the following conclusions:

- the error increases rapidly with increasing number of features and decreases slowly with increasing number of objects m ;

- an increase in the number of characteristics requires a significant increase in the volume of the training sample in order to achieve the same error.

Therefore, the paramount importance for the choice of the number of features should be played by the extra-statistical considerations arising from the essence of the problem being solved and the features of the subject domain. Only when the conditions of the statistical ensemble are satisfied, which are usually very difficult to check, can one be guided by the statistics conclusions about the required number of characteristics to ensure the accuracy of the result.

When the classification processing is realized in conditions of a small volume of the training sample, the decrease in the dimension of the original indicative space acquires a decisive role. Typically, such a transformation of the feature space is reduced to the definition of a relatively small number of features that have the greatest informativeness in accordance with the chosen criterion.

In general, when speaking about the transformation of a feature space and the choice of a criterion of informativeness, it should be borne in mind that the transformation of characteristics, performed outside the connection with the quality of classification, leads to the problem of presenting the initial data in a space of smaller dimension. The resulting set of attributes is determined by the optimization of some criterion function that does not take into account the partitioning of objects into classes. If the characteristics are chosen to improve the characteristics of the classification system, then the criterion for such a choice is related to the separability of classes. In accordance with these problems, in applied research, two approaches are commonly used to reduce the dimension of the original feature space.

In the first approach, new features are determined without taking into account the quality of classification - the task of presenting data. This problem arises when processing large arrays of information, when it is necessary to replace the system of initial features $x = (x^1, \dots, x^N)$ with a set of auxiliary variables of significantly smaller dimensions $z(x) = (z^1(x), \dots, z^l(x))$, ($l < N$). According to [6], this means the most accurate

restoration ($m \times N$) of the values of the original features $x_j^1, x_j^2, \dots, x_j^N$ by a significantly smaller number ($m \times l$) of the values of the auxiliary variables $z_j^1, z_j^2, \dots, z_j^l; j = \overline{1, m}$, where m is the number of objects in the sample. If such a substitution is possible, it leads to the indicated problem of presenting the initial data in a space of smaller dimension.

In the second approach, the search for attributes is related to the assessment of the quality of the classification. In this case, the character space is concretized, i.e. determination of an informative set of characteristics that are chosen to adequately solve the problem of classification.

1. Statement of a problem

Reducing the dimension of the source feature space in data analysis problems essentially represents a transition from the original system of attributes $x = (x^1, x^2, \dots, x^N)$ to a new system $z = (z^1, z^2, \dots, z^l)$ that includes fewer features ($l < N$) than the original system. Usually new features are formed in the form of functions from the initial features, i.e. $z = F(x)$, by solving the optimization problem. The latter consists in finding a system of characteristics Z in which

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

Here $I(z)$ - a given measure of the informativeness of a l -dimensional system of features Z , and F - the class of permissible transformations of the initial characteristics x^1, x^2, \dots, x^N , which in the general case can be represented by one of the following types of transformations: linear, nonlinear, discrete, continuous, logical. According to [6, 7], the concrete choice of the informative measure $I(z)$ and the class of admissible transformations F leads to a specific method of diminishing the dimension.

Thus, the decrease in the dimension of the original feature space can be interpreted as the mapping of the N -dimensional vector X into an l

-dimensional vector z , which in the general case can be represented as $z = F(x)$, where F is the permissible transformation, $l \leq N$.

Usually, two variants of the transformation of the original feature space.

The first variant of the transformation can be represented by a system of functions

$$\begin{cases} z^1 = f_1(x^1, x^2, \dots, x^N), \\ z^2 = f_2(x^1, x^2, \dots, x^N), \\ \dots, \\ z^l = f_l(x^1, x^2, \dots, x^N). \end{cases} \tag{2}$$

Where the new features z^1, z^2, \dots, z^l have the form of functions (usually linear) from the original ones x^1, x^2, \dots, x^N . Where $l < N$.

The second variant can be represented by a system of functions

$$z^i = f_i(x^i); i = \overline{1, N}, \tag{3}$$

$$f(x^i) = \begin{cases} 0, & \text{if the feature } x^i \text{ is excluded;} \\ 1, & \text{if the feature } x^i \text{ is not excluded.} \end{cases}$$

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

Methods based on the use of type conversions (2) and (3) are called methods to reduce the dimension of the original feature space.

The criteria for informativity considered below, being heuristic, are based on an estimation of the measure of separability of objects of a given training selection using the Euclidean metric.

2. The solution of the problem

Let's say that the training sample is specified by objects

$$x_{11}, x_{12}, \dots, x_{1m_1}, x_{21}, x_{22}, \dots, x_{2m_2}, \dots, x_{r1}, x_{r2}, \dots, x_{rm_r}$$

for which it is known that each group of objects $x_{p1}, x_{p2}, \dots, x_{pm_p}$ belongs to a certain class

$$X_p, p = \overline{1, r}.$$

Each object x_{pi} is an N -dimensional vector of numerical characteristics, i.e.

$$x_{pi} = (x_{pi}^1, x_{pi}^2, \dots, x_{pi}^N).$$

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

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

Definition 1. The truncation of space $R^N = \{x = (x^1, x^2, \dots, x^N)\}$ by λ we call a space $R^N|_\lambda = \{x|_\lambda = (\lambda^1 x^1, \lambda^2 x^2, \dots, \lambda^N x^N)\}$.

By the truncated distance between two objects $x, y \in R^N$, we mean the Euclidean distance between $x|_\lambda, y|_\lambda$ in $R^N|_\lambda$, i.e.

$$\|x - y\|_\lambda = \sqrt{\sum_{k=1}^N \lambda^k (x^k - y^k)^2}.$$

Definition 2. We call a 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 specified by the ℓ -informative vector λ , its own ℓ -dimensional feature subspace is defined. In each of these subspaces we introduce the Euclidean norm with respect to the truncation by λ

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

We denote by

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

where \bar{x}_p - averaged object of class X_p .

We introduce the function

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

The function $S_p(\lambda)$ characterizes the average spread of class objects X_p in a subset of characteristics specified by a vector λ . We define a criterion for the informativeness of subsystems in the form of a functional

$$I(\lambda) = \frac{\sum_{p,q=1}^r \|\bar{x}_p - \bar{x}_q\|_\lambda^2}{\sum_{p=1}^r S_p^2(\lambda)}. \quad (4)$$

This functional is some generalization of the Fisher functional. We denote by

$$a = (a^1, a^2, \dots, a^N); b = (b^1, b^2, \dots, b^N),$$

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

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

Then the functional (4) reduces to the form

$$I_1(\lambda) = \frac{(a, \lambda)}{(b, \lambda)}, \quad (5)$$

where $(*, *)$ - scalar product of vectors.

The coefficients a^j, b^j do not depend on λ and are computed in advance. To calculate the functional $I(\lambda)$, each requires λ an order of N operations.

Further, the criterion given in the form of the functional (5) will be called the Fisher's informative criterion and denote it as $I_1(\lambda)$.

Consider the following form of the quality functional:

$$I_2(\lambda) = \max_{p,q=1,r} \frac{R_{pq}^2(\lambda)}{G_p(\lambda)G_q(\lambda)}, \quad (6)$$

where

$$G_p(\lambda) = \sqrt{\frac{2}{m_p} \cdot \frac{1}{m_p - 1} \sum_{i=1}^{m_p} \sum_{k=1}^{m_p} \sum_{j=1}^N \lambda^j (x_{pi}^j - x_{pk}^j)^2}, \quad (7)$$

$$R_{pq}(\lambda) = \sqrt{\frac{1}{m_p} \cdot \frac{1}{m_q} \sum_{i=1}^{m_p} \sum_{k=1}^{m_p} \sum_{j=1}^N \lambda^j (x_{pi}^j - x_{pk}^j)^2} \quad (8)$$

The value $G_p(\lambda)$ characterizes the mean-square scatter of objects within the class p, the value $R_{pq}(\lambda)$ characterizes the mean square distance between classes X_p and X_q in the subspace of the subsystem of characteristics specified by the vector λ . It follows that the use of the functional (8) to determine an informative set of attributes of a given capacity is associated with the maximization of this functional. The number of operations for calculating a single value of the functional is of the order $r \times r \times N$.

Formulas (7), (8) can be reduced to the form

$$G_p(\lambda) = \sqrt{(b_p, \lambda)}, \quad R_{pq}(\lambda) = \sqrt{(a_{pq}, \lambda)},$$

where

$$(b_p, \lambda) = \sum_{j=1}^N b_p^j \lambda^j; \quad (a_{pq}, \lambda) = \sum_{j=1}^N a_{pq}^j \lambda^j,$$

$$b_p^j = \frac{2}{m_p} \cdot \frac{1}{m_p - 1} \sum_{i=1}^{m_p} \sum_{k=1}^{m_p} (x_{pi}^j - x_{pk}^j)^2,$$

$$a_{pq}^j = \frac{1}{m_p} \cdot \frac{1}{m_q} \sum_{i=1}^{m_p} \sum_{k=1}^{m_p} (x_{pi}^j - x_{qk}^j)^2.$$

Then the functional (6) takes the form

$$I_2(\lambda) = \max_{p,q=1,r} \frac{(a_{pq}, \lambda)}{\sqrt{(b_p, \lambda)} \cdot \sqrt{(b_q, \lambda)}} \quad (9)$$

Further, the criterion given by a functional of the form (9) will be called Gorelik's informative criterion.

In [9,10], the basic properties of particular types of the functional (9), which have the form

$$I_3(\lambda) = \frac{(a, \lambda)}{\sqrt{(b, \lambda)(c, \lambda)}}, \quad (10)$$

$$I_4(\lambda) = \frac{(a, \lambda)^2}{(b, \lambda)(c, \lambda)}, \quad (11)$$

$$I_5(\lambda) = \frac{(a, \lambda)}{(b, \lambda)(c, \lambda)}, \quad (12)$$

where $(*,*)$ - scalar product of vectors $a^j = a_{pq}^j, b^j = b_p^j, c^j = b_q^j, j = \overline{1, N}$.

There are also heuristic criteria for the informativity of attributes, given by functionals

$$I_6(\lambda) = \frac{(a, \lambda) + (d, \lambda)}{(b, \lambda)(c, \lambda)}, \quad (13)$$

$$I_7(\lambda) = \frac{(a, \lambda)(d, \lambda)}{(b, \lambda)(c, \lambda)}, \quad (14)$$

where

$$(b, \lambda) = \sqrt{\frac{1}{m_p} \sum_{i=1}^{m_p} \|x_{pi} - \bar{x}_p\|_\lambda^2};$$

$$(c, \lambda) = \sqrt{\frac{1}{m_q} \sum_{i=1}^{m_q} \|x_{qi} - \bar{x}_q\|_\lambda^2};$$

$$(a, \lambda) = \sqrt{\frac{1}{m_q} \sum_{i=1}^{m_q} \|x_{qi} - \bar{x}_p\|_\lambda^2};$$

$$(d, \lambda) = \sqrt{\frac{1}{m_p} \sum_{i=1}^{m_p} \|x_{pi} - \bar{x}_q\|_\lambda^2}.$$

Here (b, λ) and (c, λ) characterizes the mean-square scatter of the objects in the classes X_p and X_q , respectively, and with respect to the truncation by λ , and $(a, \lambda) + (d, \lambda)$ and $(a, \lambda) \cdot (d, \lambda)$ determine the measure of proximity between classes X_p and X_q with respect to the truncation by λ for the functionals (13) and (14), respectively.

The coefficients b^j, c^j, a^j, d^j do not depend on, so they are calculated in advance using the formulas:

$$b^k = \frac{1}{m_p} \sum_{i=1}^{m_p} (x_{pi}^k - \bar{x}_p^k)^2; \quad c^k = \frac{1}{m_q} \sum_{i=1}^{m_q} (x_{qi}^k - \bar{x}_q^k)^2;$$

$$a^k = \frac{1}{m_q} \sum_{i=1}^{m_q} (x_{qi}^k - \bar{x}_p^k)^2; \quad d^k = \frac{1}{m_p} \sum_{i=1}^{m_p} (x_{pi}^k - \bar{x}_q^k)^2.$$

The undoubted merit of the functionals considered above is that the scatter characteristics within the class and the interclass distance $R_{pq}(\lambda)$ are practically not affected by the random measurement errors contained in the training sample. The fundamental difference between the functionals (9) - (14) and the functional (5) is that

they do not sum the particular criteria, but select only one (maximum or minimum). Due to this, on the one hand, the reliability of class separation increases, however, on the other hand, there are significant difficulties associated with maximizing the functional.

In [10], the functionals presented in the form

$$I_8(\lambda) = \frac{(a, \lambda)}{(b, \lambda)} + \frac{(a, \lambda)^2}{(b, \lambda)(c, \lambda)}, \quad (15)$$

where $(*, *)$ – scalar product of vectors $a^j = b_{pq}^j, b^j = a_p^j, c^j = a_q^j$ and

$$I_9(\lambda) = \frac{(a, \lambda)}{(b, \lambda)} + \frac{(u, \lambda)^2}{(w, \lambda)(v, \lambda)}, \quad (16)$$

where the components of the vectors a and b are calculated in the same way as for the functional $I_1(\lambda)$, and the components of the vectors w, v, u are as for $I_4(\lambda)$.

Criteria of the form (15) and (16) are combined, since the first term is a criterion of Fisher's information, and the second is Gorelik.

If the corresponding components of the vectors b and c are relatively close to each other, which in practice is not so rare, it is more expedient to maximize the functional

$$\tilde{I}(\lambda) = \frac{(a, \lambda)}{(d, \lambda)},$$

where $d = (b + c)/2$.

This functional is identical in form to the functional (5).

The above replacement of the functionals $I(\lambda)$ and $\tilde{I}(\lambda)$ can be justified by the following reasoning. If the vectors b and c are componentwise close to each other, then, independently of the λ scalar products (b, λ) and (c, λ) are relatively close. In this case, their mean geometric value $\sqrt{(b, \lambda)(c, \lambda)}$ is close to their arithmetic mean $\frac{1}{2}[(b, \lambda) + (c, \lambda)] = (d, \lambda)$ and therefore we assume $I(\lambda) \approx \tilde{I}(\lambda)$.

In addition, $(d, \lambda) \geq \sqrt{(b, \lambda)(c, \lambda)}$ and, consequently, $I(\lambda) \approx \tilde{I}(\lambda)$.

Hence the maximum value $\tilde{I}(\lambda)$ will give a lower bound for the maximum value $I(\lambda)$. Since, $I(\lambda) \approx \tilde{I}(\lambda)$ vector λ the maximizing functional $\tilde{I}(\lambda)$ is taken as the optimal one.

Since all the above criteria are formed on the basis of the "compactness" hypothesis, there arises the problem of bringing them to one system and developing a single method for them.

The general criterion for the formation of an informative feature space, proposed in [11], has the following form

$$I(\lambda) = \frac{(a, \lambda)^r}{\prod_{j=1}^r (b^{(j)}, \lambda)}. \quad (17)$$

Criteria (5), (14), (15) and (16) are particular types of criterion (17).

Criterion (17) is a generalized type of the criterion of the Fisher type, represented by a homogeneous zero-order functional.

The permissible transformation of the source feature space when an IFS is selected is given by the system (3), which defines a new set of attributes z as a subsystem of the initial system of variables $x = (x^1, x^2, \dots, x^N)$, i.e.

$z(x) = (x^{i_1}, x^{i_2}, \dots, x^{i_\ell}), \ell < N$. Moreover, the INP is determined on the basis of the optimization condition (with respect to i_1, i_2, \dots, i_ℓ) the informality criterion associated with the separability of classes.

• **A regular method for determining informative features sets**

The proposed method of selecting an IFS is based on the use of the criteria of informativeness given by functionals reducible to the form (5). In this case, the solved optimization problem is formulated as

$$\begin{cases} I(\lambda) = \frac{(a, \lambda)}{(b, \lambda)} \rightarrow \max; \\ \lambda \in \Lambda^\ell, \end{cases} \quad (18)$$

where Λ^ℓ - set of all ℓ -information features.

To solve the problem (18), we introduce the vector function

$$\varphi(\lambda) = a(b, \lambda) - b(a, \lambda), \quad (19)$$

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

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

We introduce the (follower) operator

$$\mu: \Lambda^\ell \rightarrow \Lambda^\ell$$

such that

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

The operator μ has an obvious constructive idea. If we order the components of the vector $\varphi(\lambda)$, i.e. find a set of pairwise different indices j_1, j_2, \dots, j_N such that $\varphi^{j_1}(\lambda) \geq \varphi^{j_2}(\lambda) \geq \dots \geq \varphi^{j_N}(\lambda)$, then the components of the vector $\mu(\lambda)$ will be defined as

$$\begin{aligned} \mu^{j_1}(\lambda) = 1, \mu^{j_2}(\lambda) = 1, \dots, \mu^{j_r}(\lambda) = 1, \\ \mu^{j_{r+1}}(\lambda) = 0, \mu^{j_{r+2}}(\lambda) = 0, \dots, \mu^{j_N}(\lambda) = 0 \end{aligned}$$

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

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

$$(\varphi(\lambda), \mu(\lambda)) = \max \{(\varphi(\lambda), \eta) | \eta \in \Lambda^\ell\}. \quad (20)$$

Property 1. For arbitrary $\lambda (\lambda \in \Lambda^\ell)$ $(\varphi(\lambda), \mu(\lambda)) \geq 0$ is true.

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

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

We note that Theorem 2 guarantees the optimality of the solution obtained; i.e. value of the functional $I(\lambda)$ with the solution found reaches λ its maximum on the set Λ^ℓ .

On Theorems 1 and 2, the proposed method of maximizing the functional (5) based on an iterative procedure is based. And at the first step an arbitrary ℓ -informative vector λ is chosen, for example, $\lambda = \left(\overbrace{1, 1, \dots, 1}^\ell, 0, 0, \dots, 0 \right)$.

Then, at each iteration, the new vector λ is determined from the previous one by the successor operator $\mu(\lambda)$, i.e. just assignment $\lambda = \mu(\lambda)$ is made.

The iterative process continues as long as the functional $I(\lambda)$ grows. In the event that growth ceases, i.e. $I(\lambda) = I(\mu(\lambda))$, then λ - the optimal solution. Usually this solution is achieved, as shown by experiments, at 3-4 steps.

It should be noted that the results obtained extend to all criteria for the informativity of features, given by functionals, which in principle can be reduced to the form (5).

To date, various methods for selecting informative features have been developed for various particular types of Fisher criterion. In theoretical and practical terms, the method of selecting informative attribute sets, based on the use of Fisher's criterion of generalized information, is of great interest.

Consider the functional

$$I(\lambda) = \frac{\sum_{p,q=1}^r \|\bar{x}_p - \bar{x}_q\|_\lambda^2}{\sum_{p=1}^r S_p^2(\lambda)}, \quad (21)$$

which represents a generalized form of Fisher's criterion and is used in the case when the number of classes is equal r . In the space of given features we consider the N -dimensional vectors a and $b^{(1)}, b^{(2)}, \dots, b^{(r)}$, whose components are defined as

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

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

Using these notations, we reduce the functional (21) to the following form:

$$I(\lambda) = \frac{(a, \lambda)^r}{\prod_{i=1}^r (b^{(i)}, \lambda)}, \quad (22)$$

where $(*, *)$ – scalar product of vectors.

Using the concepts and definitions, consider the task of forming a space ℓ -informative features on the basis of the general form of the Fisher criterion

We represent this problem as

$$\left\{ \begin{array}{l} I(\lambda) = \frac{(a, \lambda)^r}{\prod_{j=1}^r (b^{(j)}, \lambda)} \rightarrow \max, \\ \lambda \in \Lambda^\ell, \lambda_i = \{0, 1\}, i = \overline{1, N}, \\ a, b^{(j)} \in R^N, a_i \geq 0, b_i^{(j)} > 0, \\ j = \overline{1, r}, i = \overline{1, N}, \end{array} \right. \quad (23)$$

where Λ^ℓ – set of ℓ -informative features.

Preliminarily choose a vector λ in the form

$$\lambda^0 = \left(\underbrace{1, 1, \dots, 1}_\ell, \underbrace{0, 0, \dots, 0}_{N-\ell} \right) \text{ and introduce the}$$

following notation:

$$A = \sum_{i=1}^l a_i, B_j = \sum_{i=1}^l b_i^{(j)}, j = \overline{1, r}.$$

Now, to solve the problem (23), we introduce an auxiliary vector C , whose components are calculated as follows:

$$c_i = \frac{a_i}{A} - \sum_{j=1}^r \frac{b_i^{(j)}}{B_j}.$$

Suppose, given $\forall \mu \in \Lambda^\ell$ and in accordance with λ^0 the formed vector C .

Theorem 3. In order for the chosen vector λ to provide the optimal solution of problem (23), it is necessary and sufficient that there is no vector μ satisfying the condition $(C, \mu) > 0$.

Theorem 3 is based on a method that we call

Delta and its essence is as follows.

We will form the vector μ as follows. If the components of the vector C are ordered in descending order, i.e. find a set of pairwise distinct indices j_1, j_2, \dots, j_N , such that

$$c_{j_1} \geq c_{j_2} \geq c_{j_3} \geq \dots \geq c_{j_N}, \quad (24)$$

then the components of the vector μ will be defined as

$$\mu_{j_1} = 1, \mu_{j_2} = 1, \dots, \mu_{j_\ell} = 1, \mu_{j_{\ell+1}} = 0, \mu_{j_{\ell+2}} = 0, \dots, \mu_{j_N} = 0. \quad (25)$$

In other words, the components of the vector μ corresponding to the first ℓ -maximal components of the vector C are equal to unity, the remaining components to zero.

Obviously, μ is also an ℓ -informative vector, and $(C, \mu) = \max \{ (C, \eta) \mid \eta \in \Lambda^\ell \}$.

At the first step, a vector λ^0 is selected and then, at each iteration, the new vector λ is determined from the previous vector μ by an assignment vector $\lambda = \mu$.

The iterative process continues until equality is reached $I(\lambda)$ and $I(\mu)$, which indicates the optimality μ .

Conclusion

The practical task to be solved is connected with the study of vital clinical signs (signs) of coronary heart disease.

As the object of the study was chosen pathology of the ventricle of the heart, represented by the following three states:

- Stenocardia (S) representing the 1st class of objects (K_1);
- Average state (O) representing the 2nd class of objects (K_2);
- Myocarditis infarction (I), representing the 3rd class of objects (K_3).

The initial data were presented in the form of three tables. Each type of disease means one class and the number of features that characterize each class of objects is equal to 82.

Class K_1 contained 35 objects, K_2 - 27 objects and K_3 - 21 objects. Formally, every medical object can be represented as a vector

$x_{ij} = (x_{ij}^1, x_{ij}^2, \dots, x_{ij}^{82})$. Here x_{ij}^k is the k – feature of the j – object of the i – class; m_i – the number objects of the i – class ($i = \overline{1,3}; j = \overline{1, m_i}; k = \overline{1, 82}$).

The task was solved

$$\begin{cases} I(\lambda) = \frac{(a, \lambda)}{(b, \lambda)(c, \lambda)(d, \lambda)} \rightarrow \max \\ \lambda \in \Lambda^\ell, (\ell \leq 82). \end{cases}$$

The verification of the belonging of each object to its class was carried out using the "k-nearest neighbors" method.

The analysis of the results shows that with the $\ell = 67$ found informative set of characteristics provides an error of classification of the objects of the training sample, equal to 9.8%. At the same time, this set of characteristics ensured the accuracy of the results of classification of the objects of the control sample, equal to 80%.

A meaningful interpretation of the results obtained, carried out by specialists of the center of emergency medical care, showed that these results mainly confirm their assumptions about the importance of the studied indicators, revealed by clinical means.

REFERENCES

1. Verhagen K., Dain R., Grun F. et al. Pattern Recognition: Status and Prospects. -M.: Radio and Communication, 1985. -104 p.
2. Gorsky ND, Fazilov Sh.X. Data analysis: basic stages and computational experiment. -Tashkent, 1987. - 28 pp. Dep. in VINITI on Aug. 18. 1987, No. 6057-B87.
3. Alexandrov V.V, Gorsky N.D. Algorithms and programs of the structural method of data processing. -L.: Science, 1983. -208 p.
4. Alexandrov V.V, Fazilov Sh.X. Integrated approach to data analysis // Automation of research on the basis of the information and computer network. -L.: LNIVTS AS of the USSR. 1984.-C. 4-10.
5. Aivazyan SA, Enyukov IS, Meshalkin LD Applied Statistics: Basics of Modeling and Primary Data Processing. - Moscow: Finances and Statistics, 1983. - 472 p.
6. Ayvazyan S.A., Buchstaber V.M., Enukov I.S., Meshalkin L.. Applied Statistics: Classification and Dimension Reduction. - Moscow: Finance and Statistics, 1989. - 607 p.
7. Vapnik V.N., Chervonenkis A.Ya. Theory of pattern recognition. - Moscow: Nauka, 1974. - 416 p.
8. 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.
9. 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.
10. 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.