

Determination of non-informative features based on the analysis of their relationships

N Mamatov, A Samijonov¹, N Niyozmatova

Tashkent University Information Technologies named after Al-Kharezmi, Tashkent, Uzbekistan

¹Bauman Moscow State Technical University, Moscow, Russia

Abstract. In solving most of the practical problems associated with classification, a heuristic approach is widely used. Currently, many heuristic methods and algorithms based on various heuristic criteria have been developed. The most popular criteria are heuristic informational criteria. These criteria are related to the assessment of the separability of these classes and are based on the fundamental hypothesis of recognition compactness, i.e. with an increase in the relationship between classes, their separability improves. If traits maximize relationships, they are called "good." Such heuristic criteria are widely used and give good results in solving practical problems, but they are poorly studied in theoretical terms. At present, the method for choosing non-informative features, taking into account the relationship of features based on heuristic criteria, has not yet been developed. The article considers the problem of determining informative features by eliminating uninformative ones.

1. 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. In classification problems, when forming a feature space, the initial description of the objects is first selected, and then, based on the reduction in the dimensionality of the space of the initial description, an informative description of the objects is formed.

At the first stage, to separate a given alphabet of images, the initial system of signs is selected, where you can get a priori information necessary to describe the images in the language of these signs. This stage is little studied in the problems of data analysis, and currently there are no formalized methods for its implementation. In determining the initial system of signs, a priori knowledge, intuition and experience of specialists in the corresponding subject field are widely used. In this case, one should also take into account the important circumstance associated with the fact that each real object represents an infinite number of different properties that reflect its sides.

The purpose of the second stage is to determine the most useful for classifying a set of features of the studied objects. The need to implement this stage is due to the following circumstance.

When the initial system of features is selected, then, as a rule, it turns out to be very redundant. There are pros and cons for maintaining such redundancy. The argument for is that an increase in the number of features allows a more complete description of objects. The argument is "against": an increase in the number of signs increases the "noise" in the data, complicates the processing and leads to additional time costs for its implementation.

Consequently, the argument "for" comes mainly from statistical assumptions, while the argument "against" is based primarily on non-statistical ones. If practical motives are almost always important, then the conditions when the statistics work turn out to be fulfilled much less frequently than expected.

In [5], the following criteria of applicability of statistical methods are distinguished:

- 1) the experiment can be repeated many times under the same conditions;
- 2) it is impossible to predict the outcome of the experiment in advance due to the influence of a large number of random factors;
- 3) with an increase in the number of experiments, the results converge to some values.

Moreover, the authors of [5] note that there are no strict mathematical methods to verify that these conditions are satisfied. They distinguish sociology, demography, the theory of reliability and



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

selective quality control as areas where these conditions are in most cases fulfilled. Very often, however, they turn out to be violated – in whole or in part – usually because the second part of criterion 1 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 should be taken under the conditions of the statistical ensemble and how many features should be measured (from the point of view of statistics, and not the subject area) to obtain a result with a given accuracy, it is advisable to refer to the results of studies evaluating recognition errors for different sizes of the training sample m and the number of attributes N . We draw the following conclusions:

- the error increases rapidly with increasing number of signs N and slowly decreases with increasing number of objects m ;
- an increase in the number of signs requires a significant increase in the volume of the training sample in order to achieve the same error.

Therefore, non-statistical considerations arising from the essence of the problem being solved and the features of the subject area should play the primary role in choosing the number of features. Only when the conditions of the statistical ensemble are fulfilled, which are usually very difficult to verify, can one be guided by the conclusions of statistics on the required number of attributes to ensure the accuracy of the result.

When classification processing is implemented in conditions of a small volume of the training sample, the decrease in the dimension of the initial feature space acquires a decisive role. As a rule, such a transformation of feature space is reduced to determining a relatively small number of features that have the most information content in accordance with the selected criterion.

In the general case, when speaking about the transformation of the attribute space and the choice of the criterion of informativeness, it should be borne in mind that the transformation of attributes, carried out without regard to the quality of classification, leads to the problem of representing the initial data in a space of lower dimension. The resulting set of features is determined by the optimization of some function of the criterion, which does not take into account the division 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 connected with the separability of classes. In accordance with these tasks, applied research usually uses two approaches to reducing the dimension of the original feature space [1-4].

In the first approach, new features are determined without regard to the quality of classification - the task of presenting data. This problem arises when processing large amounts 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 lower dimension $z(x) = (z^1(x), \dots, z^l(x))$, ($l < N$).

According to [6–8], this means the most accurate recovery ($m \times N$) of the values of the initial features $x_j^1, x_j^2, \dots, x_j^N$ from a significantly smaller number ($m \times l$) of values of 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 under consideration. If such a replacement is possible, then it leads to the indicated problem of representing the initial data in a space of lower dimension.

In the second approach, the search for attributes is associated with an assessment of the quality of classification. In this case, the specification of the attribute space is performed, i.e. definition of an informative set of features that are selected adequately for the classification problem to be solved.

It is the development of an approach based on the use of heuristic criteria for the informativeness of attributes associated with assessing the separability of classes according to a given training sample, is the subject of this article.

The informational content criteria considered below, being heuristic, are based on an assessment of the separability measure of objects of a given training sample using the Euclidean metric.

2. Statement of a problem and the concept of the problem decision

Let the training set be given by $x_{11}, x_{12}, \dots, x_{1m_1}, x_{21}, x_{22}, \dots, x_{2m_2}, \dots, x_{r1}, x_{r2}, \dots, x_{rm_r}$ objects, 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 signs, 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 attribute space, we introduce the vector $\lambda = (\lambda^1, \lambda^2, \dots, \lambda^N), \lambda^k \in \{0; 1\}, k = \overline{1, N}$, which uniquely characterizes a certain subsystem of attributes. The components of the vector λ , equal to unity, indicate the presence of corresponding signs in this subsystem, and the zero components indicate the absence of corresponding signs.

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

Definition 1 By truncating the space $R^N = \{x = (x^1, x^2, \dots, x^N)\}$ by λ we mean the space $R^N|_{\lambda} = \{x|_{\lambda} = (\lambda^1 x^1, \lambda^2 x^2, \dots, \lambda^N x^N)\}$.

As the displayed relationship between objects $x, y \in R^N$, we take some function $r_{\lambda}(x, y)$ of proximity between objects $x|_{\lambda}, y|_{\lambda}$ in space $R^N|_{\lambda}$.

Definition 2 We call the vector λ ℓ -uninformative if the sum of its components is ℓ , i.e. $\sum_{i=1}^N \lambda^i = \ell$.

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 averaged class object X_p .

We introduce the function

$$r_p(x_p, x) = \sqrt{\frac{1}{m_p} \sum_{i=1}^N r(x_p, x_i)}$$

The $r_p(x_p, x)$ function characterizes the average scatter of class X_p objects in a subset of the attributes specified by the λ vector. We set the criterion for the informativeness of the subsystems in the form of a functional

$$I(\lambda) = \frac{r(\bar{x}, \bar{y})}{\sum_{p=1}^h r_p(x_p, x)} \tag{1}$$

This functional is some generalization of the Fisher functional.

Denote

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

$$a_i = r(\bar{x}_i, \bar{y}_i), b_i = \sum_{j=1}^N r(x_i, x_j), i = \overline{1, N}.$$

Then functional (1) reduces to the form

$$I_1(\lambda) = \frac{(a, \lambda)}{(b, \lambda)} \tag{2}$$

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

The odds a^j, b^j are independent of λ and are calculated in advance. To calculate the functional $I(\lambda)$ for each λ , an order of N operations is required.

Further, the criterion specified in the form of functional (2) will be called the Fisher noninformativity criterion and denote it as $I_1(\lambda)$.

A method has been developed for determining the totality of non-informative features based on a

simple type of Fisher criterion, called the "Ordering" method, this method does not always provide the best solution according to the Fisher criterion. The following are the optimal conditions for this method.

This method can be used to determine non-informative features based on the solution of the following optimization problem:

$$\begin{cases} I(\lambda) = \frac{(a, \lambda)}{(b, \lambda)} \rightarrow \min, \\ \lambda \in \Lambda^l, \lambda_i \in \{0, 1\}, i = \overline{1, N}, \\ a, b \in R^N, a_i \geq 0, b_i > 0, i = \overline{1, N}, \end{cases} \quad (3)$$

where Λ^l is the space of l – non-informative vectors.

The main goal is to determine when the "Ordering" method with respect to the ratio of vectors a and b gives an optimal solution to problem (3).

Suppose that the vectors a and b , respectively, of the components are ordered as follows:

$$\frac{a_1}{b_1} \leq \frac{a_2}{b_2} \leq \dots \leq \frac{a_N}{b_N}. \quad (4)$$

To solve this problem, the following lemmas and theorems will be useful in what follows.

Let real numbers a, b and $c \geq 0, d > 0$ ($a + c \geq 0, b + d > 0$) be given. Then one and the following lemmas hold:

Lemma 1 If $\begin{cases} a > 0 \\ b > 0 \end{cases}$ и $\frac{c}{d} > \frac{a}{b}$, then the following relation holds $\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$.

Lemma 2 If $\begin{cases} a > 0 \\ b > 0 \end{cases}$ и $\frac{c}{d} < \frac{a}{b}$, then the following relation holds $\frac{a}{b} > \frac{a+c}{b+d} > \frac{c}{d}$.

Lemma 3 If $\begin{cases} a < 0 \\ b < 0 \end{cases}$ и $\frac{c}{d} < \frac{a}{b}$, then the following relation holds $\frac{a}{b} > \frac{a+c}{b+d} < \frac{c}{d}$.

Lemma 4 If $\begin{cases} a < 0 \\ b < 0 \end{cases}$ и $\frac{c}{d} > \frac{a}{b}$, then the following relation holds $\frac{a}{b} < \frac{a+c}{b+d} > \frac{c}{d}$.

Lemma 5 If $\begin{cases} a \geq 0 \\ b \leq 0 \end{cases}$, then the following relation holds $\frac{a+c}{b+d} \geq \frac{c}{d}$.

Lemma 6 If $\begin{cases} a \leq 0 \\ b \geq 0 \end{cases}$, then the following relation holds $\frac{a+c}{b+d} \leq \frac{c}{d}$.

As proof of the above lemma is very simple, not shown.

We introduce the following notation:

$$A = \sum_{i=1}^l a_i, B = \sum_{i=1}^l b_i, \begin{cases} \Delta a_{ij} = a_j - a_i \\ \Delta b_{ij} = b_j - b_i, i = \overline{1, l}, j = \overline{l+1, N} \end{cases}, \lambda^0 = \left(\underbrace{1, 1, \dots, 1}_{lma}, \underbrace{0, 0, \dots, 0}_{N-lma} \right).$$

If $a = \Delta a_{ij}, b = \Delta b_{ij}, c = A, d = B$ is adopted in the above lemmas, then for $\forall i, j$ ($i = \overline{1, l}, j = \overline{l+1, N}$),

taking into account $\begin{cases} A + \Delta a_{ij} \geq 0, \\ B + \Delta b_{ij} > 0 \end{cases}$, one of these lemmas will take place.

Theorem 1 For the vector $\lambda^0 = \left(\underbrace{1, 1, \dots, 1}_{lma}, \underbrace{0, 0, \dots, 0}_{N-lma} \right)$ chosen using the ordered sequence (4) to be the

optimal solution to problem (3), it is necessary and sufficient that there are no $a = \Delta a_{ij}$, $b = \Delta b_{ij}$ relations under the conditions of Lemmas 2 and 4.

Proof.

Adequacy.

Let any vector $\lambda \in \Lambda^l$ be selected. Then the expressions

$$\begin{cases} A^* = (a, \lambda) = \sum_{i=1}^N a_i \lambda_i; \\ B^* = (b, \lambda) = \sum_{i=1}^N b_i \lambda_i \end{cases}$$

can be written as

$$\begin{cases} A^* = A + \sum_{i=1}^p \Delta a_{fk}^{(k)}; \\ B^* = B + \sum_{i=1}^p \Delta b_{fk}^{(k)}. \end{cases}$$

To preserve ℓ - the non-informational content of the vector λ the indices f and k are defined as follows.

1. If $\lambda_i = 0$ and $\lambda_j = 1$, then $f = j$ and $k = i$ ($i = \overline{1, l}, j = \overline{l+1, N}$).
2. If $\lambda_i = 1$ and $\lambda_j = 0$, then $f = i$ and $k = j$ ($i = \overline{1, l}, j = \overline{l+1, N}$).

For A^* and B^* there are equalities

$$\begin{cases} A^* = A + A_1 + A_2 + A_3 + A_4 + A_5 + A_6; \\ B^* = B + B_1 + B_2 + B_3 + B_4 + B_5 + B_6, \end{cases}$$

where A_k and B_k are the sums of Δa_{ij} and Δb_{ij} , respectively, satisfying the conditions of Lemma k, where $k = \overline{1, 6}$.

From Lemma 5

$$\frac{A_1 + A_2 + A_3 + A_4 + A_5 + A_6}{B_1 + B_2 + B_3 + B_4 + B_5 + B_6} \geq \frac{A_1 + A_2 + A_3 + A_4 + A_6}{B_1 + B_2 + B_3 + B_4 + B_6}.$$

The sums of A_6 and B_6 are equal to zero, since condition (3) is satisfied, and the sums of A_1, B_1 and A_3, B_3 are equal to zero by the hypothesis of the theorem.

Thus, there is a

$$\begin{cases} A^* = A + A_2 + A_4; \\ B^* = B + B_2 + B_4. \end{cases}$$

The terms of the sums A_2 and B_2 satisfy the conditions of Lemma 2, then $\begin{cases} A_2 > 0 \\ B_2 > 0 \end{cases}$ follows from

$$\frac{A}{B} < \frac{A_2}{B_2}.$$

From Lemma 2

$$\frac{A_2}{B_2} > \frac{A + A_2}{B + B_2} > \frac{A}{B}.$$

The summands of the sums A_4 and B_4 satisfy the conditions of Lemma 4, then from $\begin{cases} A_4 < 0 \\ B_4 < 0 \end{cases}$

follows $\frac{A}{B} > \frac{A_4}{B_4}$.

From Lemma 4

$$\frac{A + A_4}{B + B_4} > \frac{A}{B}.$$

From the inequalities

$$\begin{cases} \frac{A_4}{B_4} < \frac{A}{B} < \frac{A_2}{B_2}; \\ \frac{A + A_2}{B + B_2} > \frac{A}{B} > \frac{A_4}{B_4} \end{cases}$$

follow

$$\frac{A + A_2}{B + B_2} > \frac{A}{B} > \frac{A_4}{B_4}. \quad (5)$$

Since $\begin{cases} A_4 < 0 \\ B_4 < 0 \end{cases}$, then, given the inequality

$$\frac{A + A_2}{B + B_2} > \frac{A_4}{B_4}$$

and Lemma 4, we obtain

$$\frac{A + A_2 + A_4}{B + B_2 + B_4} > \frac{A + A_2}{B + B_2}. \quad (6)$$

From inequalities (5) and (6)

$$\frac{A + A_2 + A_4}{B + B_2 + B_4} > \frac{A}{B}.$$

Necessity. Suppose there exist Δa_{ij} and Δb_{ij} satisfying the conditions of Lemmas 1 and 3. According to these lemmas, the inequality

$$\frac{A + A_1 + A_3}{B + B_1 + B_3} < \frac{A}{B}.$$

The summands of the sums A_1 and B_1 satisfy the conditions of Lemma 1, then from $\begin{cases} A_1 > 0 \\ B_1 > 0 \end{cases}$

follows $\frac{A}{B} > \frac{A_1}{B_1}$, and from Lemma 1 it follows

$$\frac{A + A_1}{B + B_1} < \frac{A}{B}.$$

The summands of the sums A_3 and B_3 satisfy the conditions of Lemma 3, then from $\begin{cases} A_3 < 0 \\ B_3 < 0 \end{cases}$

follows $\frac{A}{B} < \frac{A_3}{B_3}$, and from Lemma 3 it follows

$$\frac{A + A_3}{B + B_3} < \frac{A}{B}.$$

From $\frac{A}{B} > \frac{A_1}{B_1}$ and $\frac{A}{B} < \frac{A_3}{B_3}$ follows

$$\frac{A_1}{B_1} < \frac{A}{B} < \frac{A_3}{B_3}. \quad (7)$$

Since $\begin{cases} A_3 < 0 \\ B_3 < 0 \end{cases}$, then, taking into account the inequality $\frac{A_1}{B_1} < \frac{A_3}{B_3}$ and Lemma 3, we obtain

$$\frac{A_1 + A_3}{B_1 + B_3} < \frac{A_1}{B_1} < \frac{A}{B}. \quad (8)$$

Since $\begin{cases} A > 0 \\ B > 0 \end{cases}$, then, taking into account the inequality $\frac{A_1 + A_3}{B_1 + B_3} < \frac{A}{B}$ and Lemma 2, we obtain

$$\frac{A_1 + A_3}{B_1 + B_3} < \frac{A + A_1 + A_3}{B + B_1 + B_3} < \frac{A}{B}.$$

From

$$\frac{A}{B} > \frac{A + A_1 + A_3}{B + B_1 + B_3}$$

it follows that the vector $\lambda^0 = \left(\underbrace{1, 1, 1, \dots, 1}_\ell, \underbrace{0, 0, \dots, 0}_{N-\ell} \right)$ corresponding to the value $I(\lambda) = \frac{A}{B}$ is not

optimal.

The theorem is proved.

If the vector λ selected using the sequence (3) is not an optimal solution to the problem (2), then to find such a solution, it is necessary to make replacements based on Lemmas 2 and 4. This process continues until all Δa_{ij} and Δb_{ij} that satisfy the conditions of Lemmas 2 and 4 are exhausted. Moreover, in accordance with Theorem 1, the solution found will be optimal.

In this method, the value of the functional and the components of the vector λ are formed on the basis of the indicated lemmas as follows.

Let Lemmas 2 and 4 hold for Δa_{ij} and Δb_{ij} . In this case, having $\frac{A + \Delta a_{ij}}{B + \Delta b_{ij}} > \frac{A}{B}$ and interchanging the components i and j of the vector λ , we get the corresponding λ value of the functional equal to $\frac{A + \Delta a_{ij}}{B + \Delta b_{ij}}$.

Based on this method, the following algorithm is developed.

Step 1 We set the vector $\lambda = \{ \underbrace{1, 1, \dots, 1}_\ell, 0, 0, \dots, 0 \}$.

Step 2 We calculate the values of A and B, i.e. $A = (a, \lambda)$, $B = (b, \lambda)$.

Step 3 Assign $i = 1$, $j = N$; $A_1 = A$, $B_1 = B$.

Step 4 We calculate the values Δa_{ij} and Δb_{ij} .

Step 5 We check the conditions of Lemma 4. If Δa_{ij} and Δb_{ij} satisfy these conditions, then the values of the i -th and j -th components of the vector λ are interchanged and then we calculate $A = A + \Delta a_{ij}$, $B = B + \Delta b_{ij}$, go to step 7, otherwise, to the next step.

Step 6 We check the conditions of Lemma 2. If Δa_{ij} and Δb_{ij} do not satisfy the conditions of this lemma, go to the next step, otherwise the i -th and j -th components of the vector λ are interchangeable and calculate $A = A + \Delta a_{ij}$, $B = B + \Delta b_{ij}$ go to the next step.

Step 7 Check condition $j > \ell$. If it is satisfied, then assign $j = j - 1$ and go to step 4, otherwise - to the next step.

Step 8 Check condition $i < \ell$. If it is satisfied, then assign $i = i + 1$ and go to step 4, otherwise - to the

next step.

Step 9 We check the conditions $A_1 = A$ and $B_1 = B$. If they are satisfied, then the vector λ is the optimal solution and we complete the process, otherwise, go to step 3.

The method implemented by this algorithm is based on Theorem 1, and is called the “Delta-1” method in the work.

In general, Theorem 1 makes it possible to determine optimal results based on “ordering” methods.

In many cases, a pre-selected vector λ can provide an optimal solution to problem (3). therefore, the following theorem allows us to determine under what conditions this can happen.

Let selected $\forall \lambda \in \Lambda^l$.

Theorem 2 In order for the selected vector λ to provide an optimal solution to problem (3), it is necessary and sufficient that there are $a = \Delta a_{ij}$ and $b = \Delta b_{ij}$ ($i = \overline{1, l}, j = \overline{l+1, N}$) that satisfy the conditions of Lemmas 2, 4, and 5.

If the vector λ is not an optimal solution to problem (3), then replacements are made based on Lemmas 2, 4, and 5.

The replacement process continues until all Δa_{ij} and Δb_{ij} are satisfied that satisfy the conditions of Lemmas 2, 4, and 5 and, in accordance with Theorem 1, the solution found is optimal.

In this method, the values of the functional and components of the vector λ are determined as follows.

Suppose that one of Lemmas 2, 4, and 5 is valid for Δa_{ij} and Δb_{ij} . In this case, in accordance with these lemmas, the relation $\frac{A + \Delta a_{ij}}{B + \Delta b_{ij}} > \frac{A}{B}$ holds and the values of the components i and j of the vector

λ are interchanged.

The process of successive interchange is continued until the conditions of Theorem 1 are satisfied.

This method in the work is called the “Delta-2” method and is implemented using the algorithm presented as follows.

Step 1 Set the initial vector $\lambda = \{1, \underbrace{1, \dots, 1}_\ell, 0, 0, \dots, 0\}$.

Step 2 We calculate the values of A and B, i.e. $A = (a, \lambda)$, $B = (b, \lambda)$.

Step 3 Assign $i = 1, j = N$; $A_1 = A, B_1 = B$.

Step 4 We calculate the values Δa_{ij} and Δb_{ij} .

Step 5 We check the conditions of Lemma 4. If Δa_{ij} and Δb_{ij} satisfy these conditions, then, in accordance with it, we replace the values of the i -th and j -th components of the vector λ , calculate $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$ and go to step 9, otherwise, to the next step.

Step 6 We check the conditions of Lemma 2. If Δa_{ij} and Δb_{ij} do not satisfy the conditions of this lemma, then, in accordance with it, we replace the values of the i -th and j -th components of the vector λ , calculate $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$ and go to step 9, otherwise on the next step.

Step 7 We check the conditions of Lemma 5. If Δa_{ij} and Δb_{ij} do not satisfy the conditions of this lemma, then, in accordance with it, we replace the values of the i -th and j -th components of the vector λ , calculate $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$ and go to step 9, otherwise on the next step.

Step 8 Check condition $j > \ell$. If it is, then we assign $j = j - 1$ and go to step 5; otherwise, go to the next step.

Step 9 Check condition $i < \ell$. If it is, then we carry out the assignment $i = i + 1$ and go to step 5, otherwise - to the next step.

Step 10 We check the conditions $A_1 = A$ and $B_1 = B$. If they are satisfied, then the vector λ is the optimal solution and complete the process, otherwise go to step 3.

3. Conclusion

The paper proposes a new approach for the selection of non-informative signs, taking into account the relationship of signs based on heuristic criteria, defines the optimal conditions for the method of "ordering" and the selected vector. Using the proved theorems, a new method for selecting non-informative features using heuristic criteria of Fisher type has been developed.

References

- [1] Ayvazyan S Buchstaber V Eukov I and Meshalkin L 1989 *Applied Statistics: Classification and Dimension Reduction* (Moscow: Finance and Statistics) p 607
- [2] Alexandrov V and Fazilov Sh 1984 *Proc. Automation of research on the basis of the information and computer network* (Leningrad: LNIVTS AS of the USSR) pp 4-10
- [3] Gorelik A and Skripkin V 1974 *Building recognition systems* (Moscow: Sov. radio) p 224
- [4] Zhuravlev Yu Kamilov M and Tulyaganov Sh 1974 *Algorithms for calculating estimates and their application* (Tashkent: Fan) p 124
- [5] Kutin G 1981 *Proc. International radioelectronics* vol 9 pp 54–70
- [6] Cheponis K 1988 *Methods, criteria and algorithms used in the conversion, selection and selection of features in data analysis* (Lithuania: Vilnius) p 150
- [7] Fazilov Sh, Mamatov N *IOP Conf. Series: Journal of Physics: Conf. Series 1210 (2019) 012043 doi:10.1088/1742-6596/1210/1/012043*
- [8] Fazilov Sh and Mamatov N 2007 *Reports of the II International Conference Problems of Management and Informatics* (Kirgizstan: Bishkek) pp 59-68