SELECTION OF INFORMATIVE FEATURES USING HEURISTIC CRITERIA OF FISHER TYPE

MAMATOV N. S

Centre for the development of software and hard warily-program complex of Tashkent

University of Information technologies, Uzbekistan

ABSTRACT

At present the most popular criteria are of informative heuristic criteria associated with the estimation of separability given classes and based on the fundamental pattern recognition compactness hypothesis: with increasing distance between the classes improved their separability. "Good" are those features that maximize the distance.

Such heuristic criteria, although are widely used in solving practical problems of classification, but in theory are scarcely explored.

KEYWORDS: Informative Features, Fisher Type Functional, Heuristic Criteria, Space of Features

INTRODUCTION

Formation of the feature space in problems of classification can be divided into two stages: selection of the initial description of objects and the formation of an informative description of the objects on the basis of reducing the dimension of the initial description of the space.

The first stage selects the original characteristics of the system, useful in varying degrees, to separate a predetermined alphabet of images, which can obtain a priori information necessary to describe the language of image features. This step is the least developed in the data analysis problems, where there are currently no formalized methods of its implementation. When determining the initial signs of a system widely used prior knowledge, intuition and experience of the subject field. In this case, you should also take into account the important fact related to the fact that every real object is an infinite number of different properties, reflecting his hand. Naturally, in each case they are not all essential properties, but only a limited set of their defining features really allow classification. Finding such features always requires careful examination of the content essence of classified objects using experimental data on the properties of the objects under consideration. To solve this problem may be a useful software data analysis tools, such as exploratory analysis tools, knowledge discovery and verification of various systems of signs. In this case, great help can provide structural data processing methods, in which we study the structural relationship of geometric configurations point objects in multidimensional space descriptions. Analysis of the data structure helps the researcher understand what properties of objects contribute to the separation of the images, to evaluate the information content of the individual features.

Assigning of the second stage of the above - to determine the most useful to classify a set of attributes of the objects. Necessity the implementation of this phase is due to the following circumstance.

When the initial system features selected, it usually turns out to be quite excessive. There are arguments "for" and "against" on the conservation of such redundancy. The argument "for" is that increasing the number of features allows

more fully describe the objects. The argument "against": an increase in the number of features increases the "noise" in the data makes it difficult to process and leads to additional time for its implementation.

Consequently, the argument "for" comes mainly from the statistical 'preconditions, while the argument "against" is pushed mainly by non-statistical. If practical motives are almost always important, the conditions "when running statistics are executed much less frequently than expected. In [1] includes such criteria for the applicability of statistical methods:

- you can repeatedly repeat the experiment in the same conditions;

- it is impossible to predict the outcome of the experiment due to the influence of a large number of random factors;

- With an increasing number experimental results converge to certain values.

Moreover, the authors of [1] pointed out that the strict mathematical methods to verify the fulfillment of these conditions, in a particular case does not exist. They secrete sociology, demography, theory of reliability and quality as the sampling area, where these conditions are fulfilled in most cases. Very often, however, they are violated - wholly or partly - usually due to the fact that it is not carried out the second part of claim 1 criterion, i.e. not complied with the same experimental conditions..

In connection with the search for an answer to the question: how many objects should be taken in compliance with the conditions of a statistical ensemble and how signs should be measured (in terms of statistics and not domain) to produce a result with a given accuracy, it is advisable to refer to the results of studies on the evaluation of recognition errors at different teaching sample volumes m and the number of signs of N [1-5]. You can draw the following conclusions:

- the error increases rapidly with an increasing number signs of N and slowly decreases with increasing number of objects $m;$

- Increasing the number of signs requires a significant increase in the volume of teaching sample to achieve the same error.

Therefore, paramount importance to choose the number of signs is to play non-statistical considerations arising from the nature of the problem being solved and the features of the subject area. Only when performing the static ensemble of conditions that are usually very difficult to check, you can be guided by the findings of the statistical evidence of the required amount to ensure the accuracy of the result.

When the classification process is realized in the conditions of small volume of training sample, the reduction of the dimension of the original feature space becomes crucial. Usually, such a conversion feature space is reduced to the determination of a relatively small number of features that have the greatest information content in accordance with the selected criterion.

In general, speaking about the transformation of the feature space and the choice of in formativeness criterion, it should be borne in mind that signs of transformation, carried out due to the classification of quality, resulting in the problem of representation of the original data in the space of smaller dimension. Thus obtained is determined by a set of features optimized some functions of criteria ignores the divide objects into classes. If the symptoms are selected to improve the performance of classifying system, the criterion for this choice is linked to the separability of classes. In

accordance with these goals two approaches to reduce the original space of dimension attributes are commonly used in applied researches.

In the first approach, new features are determined without regard to the classification of the quality - the task of presenting data. This problem occurs when processing large amounts of information when necessary to replace the original system features $x = (x^1,...,x^N)$ a set of auxiliary variables significantly smaller dimension $z(x) = (z^1(x),...z^l(x)),\ (l < N)$. According to [2], it is the most accurate recovery ($m \times N$) values of the initial features $x_j^1, x_j^2...,x_j^N$ at a substantially smaller number of ($m \times l$) values of auxiliary variables $z_j^1,.z_j^2,...,z_j^l;\ j = \overline{1,m}$, where m - number of objects in the given sample. If such a replacement is possible, then it leads to this problem of representation of the original data in the space of smaller dimension.

In the second approach search attributes associated with the evaluation classification quality. In this case, the specification of the feature space is performed, that is the definition of a set of informative signs, which are selected to adequately address the problem of classification.

It is the development of an approach based on the use of heuristic criteria of in formativeness of the symptoms associated with the evaluation of separability of classes given training sample, the subject of this article.

STATEMENT OF A PROBLEM AND THE CONCEPT OF THE PROBLEM DECISION

Considered below informativeness criterion, being a heuristic based on an assessment of the measures of separability of objects given training sample using the Euclidean metric

For example, the teaching sample sets the object $x_{11}, x_{12},...,x_{1m_1}, x_{21}, x_{22},...,x_{2m_2},....,x_{r1}, x_{r2},...,x_{rm_r}$, for which it is known that each group of objects $x_{p1}, x_{p2},...,x_{pm_p}$ belongs to a particular class $X_p,\ p = \overline{1,r}$.

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

For a given training sample objects $x_{p1}, x_{p2},...,x_{pm_p} \in X_p,\ p = \overline{1,r}$, where x_{pi} -the vector in the N-dimensional feature space, we introduce the vector $\lambda = \left(\lambda^1, \lambda^2,...,\lambda^N\right)$, $\lambda^k \in \{0;1\}$, $k = \overline{1,N}$, which, as noted in the previous section, uniquely characterizes the particular subsystem features. The vector components λ equal to one indicate the presence of the relevant signs in this subsystem, and zero components show no signs of the relevant.

Space of features $\left\{x = \left(x^1, x^2,...,x^N\right)\right\}$ will assume Euclidean and is denoted by R^N.

Definition 1: Truncation of space $R^N = \left\{x = \left(x^1, x^2,...,x^N\right)\right\}$ on the λ call space $R^N\big|_\lambda = \left\{x\big|_\lambda = \left(\lambda^1 x^1, \lambda^2 x^2,...,\lambda^N x^N\right)\right\}$.

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

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

Definition 2: A vector λ is ℓ - informative, if the sum of the components is equal ℓ, i.e.

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

For each subsystem, a given ℓ -informative vector λ, defined its ℓ -dimensional attributive subspace. In each of these spaces, we introduce some norms regarding the truncation on λ, for simplicity choose Euclidean norm:

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

We denote

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

Where $\overline{x}_p$ - the average object of a class X_p.

We introduce the function

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

The function $S_p(\lambda)$ describes the average spread of objects of the class X_p in the subset of features defined by the vector λ. We define criteria of informativeness content in the form of functional subsystems

$$I_1(\lambda) = \frac{\sum_{p,q=1}^{r} \left\| \overline{x}_p - \overline{x}_q \right\|_\lambda^2}{\sum_{p=1}^{r} S_p^2(\lambda)} \, . \tag{1}$$

This functional is a generalization of Fisher's functional [3]. We denote

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

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

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

Then the functional (1) reduces to the form

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

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

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

Next criteria given by the form of the functional (2) is called the Fisher criterion informativeness and designate it as $I_1(\lambda)$. This criterion was studied in [2, 5], where it identified particularly evaluated the efficacy and proposed methods for selection of informative features based on the maximization of the functional (2).

Were developed a lot of methods for determining the set of informative features of on the basis of a simple type of Fisher criterion. One of them is the method of "Orderings", this method does not always provide the best solution against Fisher criterion.

For example: For $a = (5, 10, 10, 1)$, $b = (1, 50, 50, 19)$ и $N = 4$, $\ell = 2$ an optimal solution is $\lambda = (1, 0, 0, 1)$ and a vector, the vector $\lambda = (1, 1, 0, 0)$ is not an optimal solution.

The following are optimal conditions for the "Orderings" of the method.

Consider the following optimization problem:

$$\begin{cases} I(\lambda) = \dfrac{(a, \lambda)}{(b, \lambda)} \to \max, \\ \lambda \in \Lambda^l, \lambda_i = \{0, 1\}, i = \overline{1, N}, \\ a, b \in R^N, a_i \geq 0, b_i > 0, i = \overline{1, N}, \end{cases} \tag{3}$$

Where Λ^l - the l-dimensional information space of features:

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

The main aim is to determine when the method of "Orderings" of against vectors a and b are gives the optimal solution of the problem (3).

Let the vectors a and b components are arranged respectively as follows:

$$\frac{a_1}{b_1} \geq \frac{a_2}{b_2} \geq \ldots \geq \frac{a_N}{b_N}. \tag{4}$$

Here are some of the data needed to find the optimal solution.

Let there be given real numbers a, b ва $c \geq 0, d > 0$ $(a + c \geq 0, b + d > 0)$. Then we have one of the following

lemmas:

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

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

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

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

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

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

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

We introduce the following notation:

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

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

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

Theorem 1: Order that a vector $\lambda^0 = \left(\underbrace{1,1,...1}_{l}, \underbrace{0,0,...,0}_{N-l} \right)$, selected using an ordered sequence (4), was the best

solution of the problem (3), if and only if the absence of relations $a = \Delta a_{ij}, \ b = \Delta b_{ij}$ in the conditions of Lemmas 2 and 4.

Proof

Adequacy: Let selected $\forall \lambda \in \Lambda^l$. Then the expression $\begin{cases} A^* = (a, \lambda) = \sum\limits_{i=1}^{N} a_i \lambda_i, \\ B^* = (b, \lambda) = \sum\limits_{i=1}^{N} b_i \lambda_i \end{cases}$ can be written as

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

To preserve the l informativeness of the vector λ, f and k are defined as follows:

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

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

For the A^* and B^* correspondingly following equality holds

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

Where A_k and B_k - sum of Δa_{ij} and Δb_{ij}, satisfying the conditions of k-lemmas $\left(k = \overline{1, 6}\right)$.

From the lemmas 6 it follows $\dfrac{A_1 + A_2 + A_3 + A_4 + A_5 + A_6}{B_1 + B_2 + B_3 + B_4 + B_5 + B_6} \leq \dfrac{A_1 + A_2 + A_3 + A_4 + A_5}{B_1 + B_2 + B_3 + B_4 + B_5}$.

Since satisfies the condition (4) the amount A_5 and B_5 is zero, and the amount A_2, B_2 and A_4, B_4 condition of the theorem.

So $\begin{cases} A^* = A + A_1 + A_3 \\ B^* = B + B_1 + B_3 \end{cases}.$

Summands of the sums A_3 and B_3 satisfies the conditions of lemma 3, then from $\begin{cases} A_3 < 0 \\ B_3 < 0 \end{cases}$ it follows

$\dfrac{A}{B} < \dfrac{A_3}{B_3}$. from lemma 3 we obtain $\dfrac{A + A_3}{B + B_3} < \dfrac{A}{B}$.

Summands of the sums A_1 and B_1 satisfies the conditions of lemma 1, then from $\begin{cases} A_1 > 0 \\ B_1 > 0 \end{cases}$ it

follows $\dfrac{A}{B} > \dfrac{A_1}{B_1}$. From lemma 3 we obtain $\dfrac{A_4}{B_4} < \dfrac{A}{B} < \dfrac{A_2}{B_2}$ (5)

As $\begin{cases} A_4 < 0 \\ B_4 < 0 \end{cases}$, then from $\dfrac{A_4}{B_4} < \dfrac{A_2}{B_2}$ and lemma -4 we obtain

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

As $\begin{cases} A > 0 \\ B > 0 \end{cases}$, then from $\dfrac{A_2 + A_4}{B_2 + B_4} > \dfrac{A}{B}$ and lemma -1 we obtain

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

From $\dfrac{A}{B} < \dfrac{A + A_2 + A_4}{B + B_2 + B_4}$ we have that the vector $\lambda^0 = \left(\underbrace{1,1,1,...,1}_{l}, \underbrace{0,0,...,0}_{N-l} \right)$ corresponding to the value

$I(\lambda) = \dfrac{A}{B}$ is not optimal.

The theorem is proved.

If the vector λ selected by the sequence (4) is not an optimal solution of the problem (3), then to find such a solution it is necessary to carry out the replacement on the basis of Lemmas 2 and 4. This process continues until all and until they are exhausted Δa_{ij} and Δb_{ij} satisfying the conditions of Lemma 2 and 4. in this case, according to theorem 1, the solution found is optimal.

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

Suppose for Δa_{ij} and Δb_{ij} are taking place Lemma 2 and 4. In this case, having the $\dfrac{A + \Delta a_{ij}}{B + \Delta b_{ij}} > \dfrac{A}{B}$ and

swapping components i and j of the vector λ , we obtain the corresponding λ value of the functional equal

$$\frac{A + \Delta a_{ij}}{B + \Delta b_{ij}} \ .$$

Based on this method, the following algorithm is designed to work denoted as A_1.

Step 1: Is accepted

$$\lambda = \{ \underbrace{1,1,...,1}_{\ell}, 0, 0, ... 0 \} \ .$$

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

Step 3: Implementation of the appropriation $i = 1$, $j = N$; $A_1 = A, B_1 = B$.

Step 4: The calculation of values Δa_{ij} and Δb_{ij} .

Step 5: Check the conditions of lemma 4. If Δa_{ij} and Δb_{ij} satisfy these conditions, the values of i-th and j-th component of the vector λ interchanged and after calculation $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$, you can skip to step 7, otherwise - go to the next step.

Step 6: Check the conditions of Lemma 2. If Δa_{ij} and Δb_{ij} not satisfy the conditions of the lemma, proceed to the next step, otherwise the i-th and j-th component of the vector λ and interchanged after calculating $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$ moves to the next step.

Step 7: Checking the condition $j > \ell$. If it holds, then carried out the assignment $j = j - 1$ and go to step 4, otherwise - go to the next step.

Step 8: Checking the condition $i < \ell$. If it holds, then carried out the assignment $i = i + 1$ and go to step 4, otherwise - go to the next step.

Step 9: Checking the condition $A_1 = A$ and $B_1 = B$. If they are performing, then the vector λ is the best solution and the process ends, otherwise, a transition to step 3.

The realized given algorithm method is based on Theorem 1, and the method is called "Delta-1".

In general, Theorem 1 gives the possibility of determining the optimum results based on the methods of "Ordering".

In many cases, pre-selected vector λ can provide the optimal solution of problem (3). Therefore the following theorem to determine the conditions under which this can happen

Let the chosen $\forall \lambda \in \Lambda^l$.

Theorem 2: To the chosen vector λ is provided the optimal solution of problem (3) if and only if the lack of $a = \Delta a_{ij}$ and $b = \Delta b_{ij}$ $\left(i = \overline{1,l}, \ j = \overline{l+1,N} \right)$, satisfying the conditions of Lemmas 2, 4 and 5.

Proof

Adequacy: Let the selected $\forall \lambda \in \Lambda^l$. Then, the expression $\begin{cases} A^* = (a, \lambda) = \sum_{i=1}^{N} a_i \lambda_i, \\ B^* = (b, \lambda) = \sum_{i=1}^{N} b_i \lambda_i \end{cases}$ can be written as

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

To preserve the l informativeness of the vector λ , f and k are defined as follows:

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

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

For the A^* and B^* correspondingly following equality holds

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

Where A_k and B_k - sum of Δa_{ij} and Δb_{ij}, satisfying the conditions of k-lemmas $\left(k = \overline{1,\ 6}\right)$.

From the lemmas 6 it follows $\dfrac{A_1 + A_2 + A_3 + A_4 + A_5 + A_6}{B_1 + B_2 + B_3 + B_4 + B_5 + B_6} \le \dfrac{A_1 + A_2 + A_3 + A_4 + A_5}{B_1 + B_2 + B_3 + B_4 + B_5}$.

By conditions of the theorem sums A_2, B_2 A_4, B_4 and A_5, B_5 are zero.

So $\begin{cases} A^* = A + A_1 + A_3 \\ B^* = B + B_1 + B_3 \end{cases}.$

Every corresponding element of the set A_3 and B_3 are satisfy the conditions of Lemma 3, then from $\begin{cases} A_3 < 0 \\ B_3 < 0 \end{cases}$

follows $\dfrac{A}{B} < \dfrac{A_3}{B_3}$. From Lemma 3 we obtain $\dfrac{A + A_3}{B + B_3} < \dfrac{A}{B}$.

Every corresponding element of the set A_1 and B_1 are satisfy the conditions of Lemma 1, then from $\begin{cases} A_1 > 0 \\ B_1 > 0 \end{cases}$

follows $\dfrac{A}{B} > \dfrac{A_1}{B_1}$. From Lemma 3 we obtain $\dfrac{A + A_1}{B + B_1} < \dfrac{A}{B}$.

From $\dfrac{A}{B} < \dfrac{A_3}{B_3}$ and $\dfrac{A + A_1}{B + B_1} < \dfrac{A}{B}$ follows

$$\dfrac{A + A_1}{B + B_1} < \dfrac{A}{B} < \dfrac{A_3}{B_3} \tag{7}$$

So $\begin{cases} A_3 < 0 \\ B_3 < 0 \end{cases}$, then from $\dfrac{A + A_1}{B + B_1} < \dfrac{A_3}{B_3}$ and lemma -3 we obtain

$$\dfrac{A + A_1 + A_3}{B + B_1 + B_3} < \dfrac{A + A_1}{B + B_1} \tag{8}$$

From (7) and (8) follows $\dfrac{A + A_1 + A_3}{B + B_1 + B_3} < \dfrac{A}{B}$.

Adequacy

Suppose there Δa_{ij} and Δb_{ij} satisfying the conditions of Lemma 2 and Lemma 4, for the result of Lemma 2 and Lemma 4 we have $\dfrac{A + A_2 + A_4}{B + B_2 + B_4} > \dfrac{A}{B}$.

Every corresponding elements of the set A_1 and B_1 are satisfy the conditions of Lemma 2, then from $\begin{cases} A_2 > 0 \\ B_2 > 0 \end{cases}$

follows $\dfrac{A}{B} < \dfrac{A_2}{B_2}$. From Lemma 2 we obtain $\dfrac{A + A_2}{B + B_2} > \dfrac{A}{B}$.

Every corresponding element of the set A_4 and B_4 are satisfy the conditions of Lemma 4, then from $\begin{cases} A_4 < 0 \\ B_4 < 0 \end{cases}$

follows $\dfrac{A}{B} > \dfrac{A_4}{B_4}$. From Lemma 4 we obtain $\dfrac{A + A_4}{B + B_4} > \dfrac{A}{B}$.

From $\dfrac{A}{B} < \dfrac{A_2}{B_2}$ и $\dfrac{A}{B} > \dfrac{A_4}{B_4}$ follows

$$\frac{A_4}{B_4} < \frac{A}{B} < \frac{A_2}{B_2} \tag{9}$$

So $\begin{cases} A_4 < 0 \\ B_4 < 0 \end{cases}$, then from $\dfrac{A_4}{B_4} < \dfrac{A_2}{B_2}$ and lemmas -4 we obtain

$$\frac{A_2 + A_4}{B_2 + B_4} > \frac{A_2}{B_2} > \frac{A}{B} \ . \tag{10}$$

So $\begin{cases} A > 0 \\ B > 0 \end{cases}$, then from $\dfrac{A_2 + A_4}{B_2 + B_4} > \dfrac{A}{B}$ and lemmas -1 we obtain

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

Every corresponding element of the set A_5 and B_5 are satisfy the conditions of Lemma 5,

then $\dfrac{A + A_2 + A_4 + A_5}{B + B_2 + B_4 + B_5} > \dfrac{A + A_2 + A_4}{B + B_2 + B_4} > \dfrac{A}{B}$.

From $\dfrac{A + A_2 + A_4 + A_5}{B + B_2 + B_4 + B_5} > \dfrac{A}{B}$ we have that the vector λ corresponding to the value $I(\lambda) = \dfrac{A}{B}$ is not optimal.

If the vector λ is not an optimal solution of the problem (3), we carried out a replacement on the basis of Lemmas 2, 4 and 5.

The replacement process is continued until the exhausted Δa_{ij} and Δb_{ij} are satisfying the conditions of Lemmas 2, 4 and 5, and at the same time, in accordance with Theorem 2 the solution found is optimal.

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

Let us suppose, for Δa_{ij} and Δb_{ij} one of the lemmas 2, 4 and 5 is true. In this case, in accordance with these lemmas $\dfrac{A + \Delta a_{ij}}{B + \Delta b_{ij}} > \dfrac{A}{B}$ and the relation values of the components i and j of vector λ are changing mutually

Process of successive interchange continues until, until you have the conditions of Theorem 2.

This method is called the method of "Delta-2" and implemented by the algorithm, denoted as A_2 and represented as follows:

Step 1: Is accepted

$$\lambda = \{\underbrace{1, 1, \ldots, 1}_{\ell}, 0, 0, \ldots 0\}.$$

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

Step 3: Implementation of the appropriation $i = 1$, $j = N$; $A_1 = A, B_1 = B$.

Step 4: The calculation of values Δa_{ij} and Δb_{ij}.

Step 5: Check the conditions of lemma 4. If Δa_{ij} and Δb_{ij} satisfy these conditions, the values of i-th and j-th component of the vector λ interchanged and after calculation $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$, you can skip to step 9, otherwise - go to the next step.

Step 6: Check the conditions of Lemma 2. If Δa_{ij} and Δb_{ij} not satisfy the conditions of the lemma, the values of i-th and j-th component of the vector λ interchanged and after calculation $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$, you can skip to step 9, otherwise - go to the next step.

Step 7: Check the conditions of Lemma 5. If Δa_{ij} and Δb_{ij} not satisfy the conditions of the lemma, the values of i-th and j-th component of the vector λ interchanged and after calculation $A = A + \Delta a_{ij}, B = B + \Delta b_{ij}$, you can skip to step 9, otherwise - go to the next step.

Step 8: Checking the condition $j > \ell$. If it holds, then carried out the assignment $j = j - 1$ and go to step 5, otherwise - go to the next step.

Step 9: Checking the condition $i < \ell$. If it holds, then carried out the assignment $i = i + 1$ and go to step 5, otherwise - go to the next step.

Step 10: Checking the condition $A_1 = A$ and $B_1 = B$. If they are performing, then the vector λ is the best solution and the process ends, otherwise, a transition to step 3.

CONCLUSIONS

In the article determined the optimal conditions for the method of "ordering" and the selected vector. With the help of these theorems we developed a new method for selecting informative features using heuristic criteria of Fisher type.

REFERENCES

1. Fazilov Sh., Mamatov N.S., Method of "Deltas" for building of informative features' space. Uzbek Journal "The Problem of Informatics and Energetic". 2005, *#6*, pp. 11-16.

2. Mamatov N.S. Working out the methods and algorithms for forming of informative features space with the help of к-types criteria. Uzbek Journal "The Problem of Informatics and Energetic". 2006, #2-3, pp.14-19.

3. Fazilov Sh., Mamatov N.S., Gradient method for forming a space of informative features based on uniform criteria with positive degree. Cybernetics and Informatics: reports of the II International Conference, Bishkek, 2007. pp 8-9

4. Mamatov N.S., Working out the methods and algorithms for forming of informative features space with the help of к-types criteria. Reports of the Academy of Sciences of Republic of Uzbekistan.-2007.-№4. pp. 50-53.

5. Fazilov Sh., Mamatov N.S., Gradient method for forming a space of informative features based on uniform criteria with positive degree. Reports of the Academy of Sciences of Republic of Uzbekistan.-2008.-№2. pp. 12-18.