International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Selection Of Informative Features using The Modified Version Of The Delta Method

Narzillo Mamatov, Nilufar Niyozmatova, Abdurashid Samijonov, Zafar Yuldashev, Musokhon Dadakhanov

Abstract**: *Currently, the most common criteria for informative features are heuristic criteria related to assessing the separability of given classes and based on the compactness hypothesis fundamental in pattern recognition: with increasing distance between classes, their separability improves. "Good" are those signs that maximize this distance.Although heuristic criteria, although they are widely used in solving practical problems of classification, however, in theoretical terms they are little studied.***

Keywords**: *vector, class, criteria, informative features, object, function, optimization problem.***

I. INTRODUCTION

Currently, there are many methods for selecting features that are focused on the use of a specific information content criterion. One of these methods was proposed and studied in [1-10]. Its essence is to use the measure of importance of the initial feature, which is a properly processed degree of reduction of the so-called "votes" when removing this feature. The authors of the method called this measure the information weight of the attribute. To determine the information weights, special computational algorithms are used that allow you to take into account the relationship of signs.An interesting approach to the selection of informative features was considered in [2–4]. As part of this approach, researchers have proposed many methods for determining informative sets of features, taking into account the limitations associated with the cost of creating technical means of measuring these features. The measure of informativeness of the latter is set by criteria based on the use of Euclidean distance. If the signs are deterministic, the definition of informative sets is reduced to solving optimization problems, which can be described as follows.

To determine informative features in pattern recognition problems, one can use the informativeness criterion specified in the form of the inverse Fisher functional.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author

Mamatov Narzillo*, Tashkent University Information Technologies named after Al-Kharezmi, Tashkent, Uzbekistan.

Niyozmatova Nilufar, Tashkent University Information Technologies named after Al-Kharezmi, Tashkent,

Samijonov Abdurashid, Bauman Moscow State Technical University, Moscow, Russia Federation.

Yuldashev Zafar, Tashkent University Information Technologies named after Al-Kharezmi, Tashkent,

Dadakhanov Musokhon, Namangan State University, Namangan,

In this case, as shown in [6-10], the problem of forming the space of informative features is solved by reducing to the minimization problem of the functional.In this paper, we consider the condition under which the selected set of features is the optimal solution to this problem.

II. STATEMENT OF THE PROBLEM

The following information content criteria are heuristic and, using the Euclidean metric, are based on an assessment of the separability measure for the objects of this training sample.

Let the learning set be defined by objects $x_{11}, x_{12}, ..., x_{1m_1}, x_{21}, x_{22}, ..., x_{2m_2}, ..., x_{r1}, x_{r2}, ..., x_{rm_r}$, and also suppose that each group of objects $x_{p1}, x_{p2}, ..., x_{pm_p}$ belongs to some class X_p, $p = \overline{1, r}$.

Objects of x_{pi} are N-dimensional numerical vectors, i.e. $x_{pi} = \left(x_{pi}^1, x_{pi}^2, ..., x_{pi}^N\right)$.

For this training sample of objects $x_{p1}, x_{p2}, ..., x_{pm_p} \in X_p$, $p = \overline{1, r}$, here x_{pi} is a vector in the N-dimensional space of attributes, the vector $\delta = \left(\delta^1, \delta^2, ..., \delta^N\right)$, $\delta^k \in \{0;1\}$, $k = \overline{1, N}$ is introduced, which characterizes a certain subsystem of attributes. If the components of the vector are equal to unity, then the corresponding signs exist in this subsystem, and if they are equal to zero, then they are absent.

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

By truncating the space $R^N = \left\{x = \left(x^1, x^2, ..., x^N\right)\right\}$ by δ we mean the space $R^N\big|_\delta = \left\{x\big|_\delta = \left(\delta^1 x^1, \delta^2 x^2, ..., \delta^N x^N\right)\right\}$.

The truncated distance between objects $x, y \in R^N$ is the Euclidean distance from $x\big|_\delta$, $y\big|_\delta$ to $R^N\big|_\delta$

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

If the sum of the components of the vector δ is l, then we call it l -informative, i.e. $\sum_{i=1}^{N} \delta^i = l$.

Defined by l -informative vector δ each subsystem has its own l -dimensional subspace of features.

In these subspaces we introduce the Euclidean norm with respect to truncation in δ.

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

Denote

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

where $\bar{x}_p$ is an averaged object of class X_p.

We introduce the function

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

The $S_p(\delta)$ function characterizes the average scatter of class X_p objects in a subset of the attributes specified by the δ vector. We define a criterion of informativeness of subsystems in the form of a functional

$$I_1(\delta) = \frac{\sum_{p,q=1}^{r}\left\|\bar{x}_p - \bar{x}_q\right\|_\delta^2}{\sum_{p=1}^{r} S_p^2(\delta)} \qquad (1)$$

Functional (1) is a generalization of the Fisher functional [5]. We denote

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

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

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

Then (1) comes down to the form

$$I(\delta) = \frac{(a,\delta)}{(b,\delta)}, \qquad (2)$$

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

The coefficients a^j, b^j are independent of and δ are calculated in advance. To calculate the functional $I(\delta)$ for each δ, about N operations are required.

Further, the criterion defined in the form of functional (2) will be called the Fisher information criterion and denoted by $I_1(\delta)$. This criterion was studied in [5, 9], where its features were identified, its effectiveness was evaluated, and methods for choosing informative features based on maximizing functional (2) were proposed.

Using the functional (2), the article solves the problem of choosing informative attributes, by eliminating non-informative attributes from the original attribute space.

Let natural numbers N and l $(l \le N)$ be given, as well as N-dimensional vectors $a, b \in R^N$.

According to [5, 6], we consider the optimization problem:

$$\begin{cases} I(\delta) = \dfrac{(a,\delta)}{(b,\delta)} \to \min \\ \delta \in \Omega^l, \delta_i = \{0,1\},\ i=\overline{1,N} \\ a,b \in R^N,\ a_i \ge 0, b_i > 0,\ i=\overline{1,N} \end{cases} \qquad (3)$$

where Ω^l - space l - dimensional informative features:

$$\Omega^l = \left\{\delta = (\delta_1, \delta_2, ..., \delta_N) \middle| \delta_i = \{0,1\}, i=\overline{1,N}, \sum_{i=1}^{N}\delta_i = l\right\}$$

It is necessary to determine in which case $\forall\delta \in \Omega^l$ will be the optimal solution to problem (3). To do this, we use the lemmas presented in [7]:

III. SOLUTION OF THE PROBLEM

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

Let real numbers a_1, a_2 and $a_3 \ge 0$, $a_4 > 0$ $(a_1 + a_3 \ge 0, a_2 + a_4 > 0)$ be given. For these numbers holds one of the following lemma:

Lemma 1. If $\begin{cases} a_1 > 0 \\ a_2 > 0 \end{cases}$ and $\frac{a_3}{a_4} > \frac{a_1}{a_2}$, then holds the relation $\frac{a_1}{a_2} < \frac{a_1 + a_3}{a_2 + a_4} < \frac{a_3}{a_4}$.

Lemma 2. If $\begin{cases} a_1 > 0 \\ a_2 > 0 \end{cases}$ and $\frac{a_3}{a_4} < \frac{a_1}{a_2}$, then holds the relation $\frac{a_1}{a_2} > \frac{a_1 + a_3}{a_2 + a_4} > \frac{a_3}{a_4}$.

Lemma 3. If $\begin{cases} a_1 < 0 \\ a_2 < 0 \end{cases}$ and $\frac{a_3}{a_4} < \frac{a_1}{a_2}$, then holds the relation $\frac{a_1}{a_2} > \frac{a_1 + a_3}{a_2 + a_4} < \frac{a_3}{a_4}$.

Lemma 4. If $\begin{cases} a_1 < 0 \\ a_2 < 0 \end{cases}$ and $\frac{a_3}{a_4} > \frac{a_1}{a_2}$, then holds the relation $\frac{a_1}{a_2} < \frac{a_1 + a_3}{a_2 + a_4} > \frac{a_3}{a_4}$.

Lemma 5. If $\begin{cases} a_1 \ge 0 \\ a_2 \ge 0 \end{cases}$, then holds the relation $\frac{a_1 + a_3}{a_2 + a_4} \ge \frac{a_3}{a_4}$.

Lemma 6. If $\begin{cases} a_1 \le 0 \\ a_2 \le 0 \end{cases}$, then holds the relation $\frac{a_1 + a_3}{a_2 + a_4} \le \frac{a_3}{a_4}$.

Next introduced the notation:

$$A = \sum_{i=1}^{l} a_i,\ B = \sum_{i=1}^{l} b_i,\ \begin{cases} \Delta a_{ij} = a_j - a_i \\ \Delta b_{ij} = b_j - b_i, i=\overline{1,l},\ j=\overline{l+1,N} \end{cases},$$

$$\delta^0 = \left(\underbrace{1,1,\dots,1}_{l\,ma},\underbrace{0,0,\dots,0}_{N-l\,ma}\right).$$

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

Let $\forall\delta\in\Omega^l$ be selected.

Theorem 1. For the vector δ to be the optimal solution of the (1), the absence 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 1, 3, 6 is necessary and sufficient.

Based on this theorem, a modified version of the Delta informative feature selection method, considered in [6-7], is proposed.

The theorem is proved as in [7].

If the vector δ is not an optimal solution of the (3), then replacements are made based on Lemmas 1, 3, and 6.

The replacement process continues until all Δa_{ij} and Δb_{ij} are satisfied that satisfy the conditions of Lemmas 1, 3, 6, and based on the theorem, the resulting solution is optimal.

The values of the functional and components of the vector λ are determined based on the method as follows.

Let one of Lemmas 1, 3, and 6 holds for Δa_{ij} and Δb_{ij}. In accordance with the above lemmas, the relation $\frac{A+\Delta a_{ij}}{B+\Delta b_{ij}}<\frac{A}{B}$ is satisfied and the values of the components i and j of the vector λ are interchanged.

IV. ALGORITHM

Based on this method has been proposed the following algorithm:

Step 1. Initialize $\delta=\{\underbrace{1,1,\dots}_{l}1,0,0,\dots 0\}$

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

Step 3. We carry out tasks $i=1,\ j=N;\ A_1=A, B_1=B$.

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

Step 5. Checking the conditions of Lemma 1. If Δa_{ij} and Δb_{ij} satisfy these conditions, then, in accordance with it, the values of the i-th and j-th components of the vector δ are carried out, the calculation $A=A+\Delta a_{ij}, B=B+\Delta b_{ij}$ and go to step 9, otherwise - go to the next step.

Step 6. We verify the conditions of Lemma 3. If Δa_{ij} and Δb_{ij} do not satisfy the conditions of this lemma, then in accordance with it the values of the *i*-th and *j*-th components of the vector δ are replaced, the $A=A+\Delta a_{ij}, B=B+\Delta b_{ij}$ is calculated, and go to step 9, otherwise - go to the next step.

Step 7. We verify the conditions of Lemma 6. If Δa_{ij} and Δb_{ij} do not satisfy the conditions of this lemma, then in accordance with it the values of the *i*-th and *j*-th components of the vector δ are replaced, the calculation of $A=A+\Delta a_{ij}, B=B+\Delta b_{ij}$ and go to step 9, otherwise - go to the next step.

Step 8. We verify the condition $j>l$. If it is fulfilled, then $i=i+1$ is assigned and go to step 5; otherwise, go to the next step.

Step 9. We verify the condition $i<l$. If it is fulfilled, then $i=i+1$ is assigned and go to step 5; otherwise, go to the next step.

Step 10. We verify the conditions $A_1=A$ and $B_1=B$. If they are satisfied, then the vector δ is the optimal solution and the process ends, otherwise go to step 3 is carried out.

V. RESULT ANALYSIS

The results of experimental studies that were carried out for three classes of objects K_1, K_2 and K_3 defined in R^N feature space. The dimension N is selected as a multiple of some given small number l, i.e. $N=l*h\ (h\in N)$. The set of features $x=(x_1,x_2,\dots,x_N)$ is arranged in groups of $\left(x_1,\dots,x_l,x_{l+1},\dots,x_{2l},x_{2l+1},\dots,x_{(h-1)l+1},\dots,x_{hl}\right)$, each of which contains l features. The classes are chosen so that from the point of view of their separability, the groups of characters are mutually independent, and within the same group the features are highly dependent, i.e. the removal of any sign from its group entails a sharp deterioration in the information content of this group as a whole. The experiments were related to solving the optimization problem for functional (1) using the proposed method and exhaustive search. The results of the determination of non-informative feature sets obtained using this method completely coincided with the results obtained on the basis of the exhaustive search method. In this case, the number of iterations in the exhaustive search was C_N^l, and for the proposed method, the maximum number of iterations was l.

VI. CONCLUSION

Experimental researches of the proposed algorithm for the selection of informative features were carried out on the example of solving the model problem. The results obtained using the proposed algorithm completely coincide with similar results obtained using the exhaustive search method.

REFERENCES

1. Kutin G.I. Methods for ranking feature complexes. Overview // Foreign electronics. -1981, -№9. -FROM. 54-70.
2. Zhuravlev Yu.I. On recognition algorithms with representative sets (on logical algorithms) // Journal of Computational Mathematics and Physics. 2002. Vol. 42 No. 9. - S.1425-1435.
3. Zagoruyko N.G. Methodology for evaluating the informative efficiency of independent parameters of a speech signal // Computational systems. -1964. -Vyp. 10. - S. 13-19.
4. Zagoruyko N.G. Cognitive data analysis. Novosibirsk: Academic Publishing House GEO, 2013.186 p.
5. Duda R., Hart P. Pattern recognition and scene analysis: Transl. From English. - M.: 2004. - from. 512.
6. Mamatov N, Samijonov A, and Yuldashev Z, "Selection of features based on relationships," J. Phys. Conf. Ser., vol. 1260, no. 10, p. 102008, 2019.

7. Fazilov Sh and Mamatov N, "Formation an informative description of recognizable objects," in Journal of Physics: Conference Series, 2019.
8. 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.
9. Mamatov N, "Selection of informative features using heuristic criteria of fisher type," BEST Int. J. Manag. Inf. Technol. Eng. (BEST IJMITE), 2016.
10. 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.