ALGORITHM FOR SELECTING A SET OF INFORMATIVE SIGNS BASED ON THE INFORMATION MEASUREMENT CRITERIA OF THE IMPORTANCE OF OBJECTS

Gulomjon Primovych Jo'rayev,
Doctor of Philosophy (PhD) in Technical Sciences, Kashkadarya Regional Center of Pedagogical Excellence , Uzbekistan
gpjurayev@mail.ru

Saidkul Khujamurodovich Saparov ,
PhD student , Muhammad al- Khorazmi Tashkent Information in the name technologies University, Uzbekistan
saidqul80 @ bk. ru

Uktamjon Bektashovich Allayarov,
assistant Tashkent Medicine Academy Termez branch
criptolione7777@gmail.com

Dilfuza Elmurodovna Rashidova,
assistant Samarkand economy and service Institute, Uzbekistan
dilfuzarashidova23@gmail.com

Abstract:
In this work describes the algorithm developed based on information measurement criteria for evaluating the importance of objects. This algorithm is aimed at selecting informative symbols by determining the degree of similarity of objects. Based on the similarity coefficient and important information indicators, the differences between class objects are analyzed and the data volume is optimized. This approach has been proposed as an effective tool that can be applied in the fields of machine learning and data processing.

Keywords: information measure, object recognition, similarity coefficient, informative symbols, class object, machine learning, data analysis.

I. Introduction

Today, with the rapid development of information technologies, the issues of object recognition and their classification are taking an important place in many fields. From security systems to medical diagnostics, industrial automation and many processes based on artificial intelligence continue to depend on the quality of object recognition. Especially when working with a large amount of data, there is a growing need to make effective decisions based on the

identification of important features of objects and the identification of mutual hidden laws between them [1, 2, 3, 4] .

The use of information measurement criteria to evaluate the importance of objects is important in improving the efficiency of such systems. Information measures are mathematical criteria and algorithms developed to determine the degree of similarity between objects, and these criteria help to identify the main characteristics and informative features of objects. In the study, the issues of classification of class objects according to the level of similarity and selection of their characteristic features were considered using information measurement criteria.

Determining the degree of similarity of objects is an important task in many practical fields. For example, in medicine, there is a need to compare different images in the diagnosis of diseases, and in industry, there is a need to automatically classify products or identify faults. In these processes of objects important informative signs choose not only system efficiency increases , perhaps analysis to do also speeds up the process. That's why for of objects to himself characteristic and their informative signs to determine algorithms analysis systems basis is considered

II. Main part

This is it in research objects between similarity level to determine through informative characters choose algorithms work released of objects importance in determining information measure criteria use through class of objects to himself special signs sure to determine and them classification opportunity was created . Information measure criteria of objects which features their in identification important role to play evaluation enable gives

This of approach important aspects one is that it is big in volume data with at work the most necessary characters Select information flows optimizes . For example , various properties have objects from among only the most the important ones to determine and they are based on decision acceptance to do process acceleration opportunity creates This process is especially data analysis , artificial intellect and with a car teaching in technologies big important have

This is it research at work objects importance information measure criteria based on evaluation and informative characters complex of choice efficient algorithm work exit mean caught That 's it in accordance with this , to solve the evaluation of the degree of similarity of the objects of information using the criteria of information measurement of the importance of objects, and in this way to develop an algorithm for selecting a set of informative characters is as follows done is increased [5, 6, 7] :

Suppose we are given two objects of a class $x_{pi}, x_{pk} \in X_p$ in the nominal data space X_p.

To determine the similarity of these objects in terms of nominal characters, $\mu(x_{pi}, x_{pk}) = (\mu^1(x_{pi}, x_{pk}), \mu^2(x_{pi}, x_{pk}), \dots, \mu^N(x_{pi}, x_{pk}))$ we enter a Boolean vector and its components are formed based on the following:

$$\mu^j(x_i, x_k) = \begin{cases} 1 \quad \text{if} \quad x_i^j = x_k^j, j = \overline{1,N}. \\ \\ 0 \text{ else.} \end{cases}$$

where $\mu^j(x_{pi}, x_{pk}) = 1$, if the two corresponding object components $x_i^j = x_k^j$are equal to each other, otherwise $\mu^j(x_{pi}, x_{pk}) = 0$. So, for the optional two objects of the study sample $\mu(x_i, x_k)$vector can be defined as one-valued.

by the size of x_{pi}, x_{pk}these objects $\varkappa(x_{pi}, x_{pk}) = \sum_{j=1}^{N} \mu^j(x_{pi}, x_{pk})$. This quantity x_{pi}, x_{pk}indicates the number of identical components of objects. It is determined by the degree of similarity of both objects $\nu(x_{pi}, x_{pk})$and it is calculated as a percentage

$$\nu(x_{pi}, x_{pk}) = \frac{\varkappa(x_{pi}, x_{pk}) * 100\%}{N}.$$

such δ number that there is $\tilde{X}_p$a degree of similarity between objects of the newly formed class $\forall x_{pi}, x_{pk}: \nu(x_{pi}, x_{pk}) \geq \delta$.

That is$\tilde{X}_p = \{\forall x_{pi}, x_{pk}: \nu(x_{pi}, x_{pk}) \geq \delta$ be greater than the number, $i \neq k, p = \overline{1,r}\}$ that is, finding a solution to the above-mentioned problem (3.1) provides a solution to the problems of determining the degree of similarity of class objects and choosing a set of informative symbols. The solution only $X_p$applies to class objects. The meaning of this mathematical expression is the degree of similarity between the objects of each class being formed, regardless of how many of them there are. δnot less than the number $\lambda \in \Lambda^{\ell}$required in the set. As a result, class objects are formed on demand and a set of informative characters is selected. should not be less than the given number δ.

Suppose that two objects of class $X_p$and $X_q$are differentiating$\boldsymbol{\lambda}$ which provides selection of vector components $\boldsymbol{J}(\boldsymbol{\lambda}, X_p, X_q)$let the criterion be given. Given N- dimensional character space such thatℓ $(\ell \ll N)$to the dimensional space so that the objects of two classes are clearly distinguished from each other in the resulting space of symbols.

To solve this problem, we introduce the following definitions:

1. - $\boldsymbol{I}(\boldsymbol{\lambda}, X_p)$the average of the similarity levels between all objects of the class **with respect to the vector** , $X_p$through this function ;$\boldsymbol{\lambda}$

2. - $\boldsymbol{I}(\boldsymbol{\lambda}, X_q)$the average of the mutual similarity levels of all objects of the class **with respect to the vector** $X_q$through the functional ;$\boldsymbol{\lambda}$

3 . $\boldsymbol{I}(\boldsymbol{\lambda}, X_p, X_q)$let's determine the average of the closeness between the objects of the classes through the criterion .X_p va X_q

Algorithm for selecting a set of informative symbols based on the criterion of information measurement of the importance of objects

Below we propose an algorithm based on the evaluation of the importance of information in

the process of selecting variables or features. Such algorithms are mainly used in the fields of data analysis, artificial intelligence, data processing and statistical modeling.

The main steps of this algorithm can be as follows:

Step 1. The data will be analyzed. First, the existing database is analyzed. In this process, the type, size and quality of data are taken into account. So , N- in the dimension space $x_i \in X$, $i = \overline{1,M}$,objects are given. In this case, $x_i = (x_i^1, x_i^2, \dots, x_i^N)$, $i = \overline{1,M}$,the objects are divided into clusters, which Nbelong to one of the classes in the - dimensional character space . X_pEach class $m_p$consists of $X = \bigcup_{p=1}^{r} X_p$objects $x_{p1}, \dots, x_{p_{m_p}}$. Usually, in this step, the data is pre-processed.

Step 2. In this step, first, a criterion for determining the degree of similarity of objects is used, in which a boolean vector is used to determine the similarity of objects $\mu(x_{pi}, x_{pk}) = (\mu^1(x_{pi}, x_{pk}), \mu^2(x_{pi}, x_{pk}), \dots, \mu^N(x_{pi}, x_{pk}))$:

$$\mu^j(x_i, x_k) = \begin{cases} 1 \quad \text{if} \quad x_i^j = x_k^j, j = \overline{1,N}. \\ 0 \text{ else.} \end{cases}$$

where $\mu^j(x_{pi}, x_{pk}) = 1$, if the two corresponding object components $x_i^j = x_k^j$are equal to each other, otherwise $\mu^j(x_{pi}, x_{pk}) = 0$. Then, the criterion for evaluating the character size is calculated. In this case, the coefficient of similarity of objects is determined $\varkappa(x_{pi}, x_{pk}) = \sum_{j=1}^{N} \mu^j(x_{pi}, x_{pk})$by size x_{pi}, x_{pk}. It is determined by the degree of similarity of $\nu(x_{pi}, x_{pk})$both objects $x_{pi}, x_{pk}$and it is calculated as a percentage

$$\nu(x_{pi}, x_{pk}) = \frac{\varkappa(x_{pi}, x_{pk}) * 100\%}{N}.$$

A number is given by which $\tilde{X}_p$class δ objects are formed. Let $\forall x_{pi}, x_{pk}: \nu(x_{pi}, x_{pk}) \geq \delta$it be. Yes$\tilde{X}_p = \{\forall x_{pi}, x_{pk}: \nu(x_{pi}, x_{pk}) \geq \delta$ be greater than the number, $i \neq k, p = \overline{1,r}\}$ class is formed.

Step 3. A set of informative symbols is defined. The most important and influential characters were selected from the data. This process is done by sorting the data according to various measurement criteria. Here is the following

$$\begin{cases} \tilde{X}_p = \{\forall x_{pi}, x_{pk}, \lambda: \nu(x_{pi}, x_{pk}, \lambda) \geq \delta, i \neq k, p = \overline{1,r}\} \\ \lambda \in \Lambda^{\ell} = \left\{ \lambda: \sum_{j=1}^{N} \lambda^j = \ell, \ \lambda^j \in \{0,1\}, j = \overline{1,N} \right\} \end{cases}$$

the optimization problem is solved. Here $\varkappa(x_{pi}, x_{pk}, \lambda) = \sum_{j=1}^{N} \mu^j(x_{pi}, x_{pk})\, \lambda^j$, $\nu(x_{pi}, x_{pk}, \lambda) = \frac{\varkappa(x_{pi}, x_{pk}, \lambda) * 100\%}{N}$. So, such a vector λ should be chosen, in which the number

of classes is reduced, $\tilde{X}_p$ the similarity between class objects is given δ The dan kam bo'lmasligi kerak.detected λ vector components provide a set of informative symbols. For example, we can use the complete selection algorithm to find this solution.

Step 4. Evaluation and selection of characters is carried out. Based on the selected criteria, the importance of the signs is evaluated. In this step, statistical methods and data analysis algorithms are used. Typically, the following optimization problem is solved.

$$\begin{cases} v(x_{pi}, x_{pk}, \lambda) = \max\limits_{\lambda \in \Lambda^{\ell}} \sum\limits_{i \neq k=1}^{m_p} \sum\limits_{k=i+1}^{m_p} \dfrac{\varkappa(x_{pi}, x_{pk}, \lambda) * 100\%}{N} \\ \lambda \in \Lambda^{\ell} = \left\{ \lambda : \sum\limits_{j=1}^{N} \lambda^j = \ell \,, \;\; \lambda^j \in \{0,1\}, j = \overline{1,N} \right\}. \end{cases}$$

The obtained scientific and practical results are used to find solutions to specific problems. Based on the selected characters, the data is analyzed and used in modeling. The block diagram of the above algorithm is given below (see Figure 1).

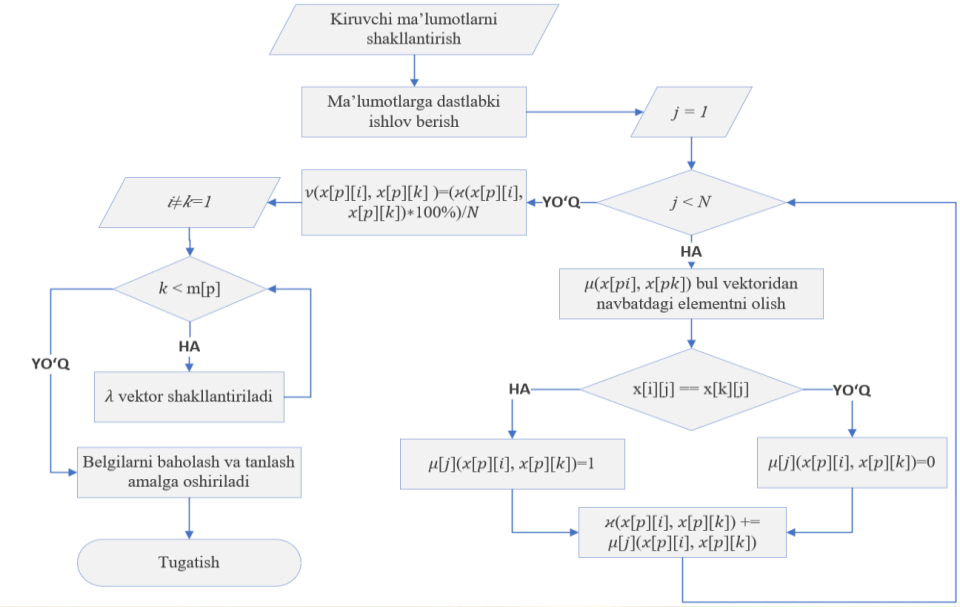

Figure 1 . Algorithm for selecting a set of informative symbols based on the criterion of information measurement of the importance of objects

III. Summary

According to the results of the research, the selection of informative signs with the help of algorithms for evaluating the importance of objects based on information measurement criteria allows to increase the efficiency of systems. In this case, the similarity coefficients between class objects are calculated and a set of informative characters is selected, which allows to reduce the volume of information flows and simplify analysis processes. This is the approach systems optimization and of resources efficient to use help gives

Research at work classes between of objects similarity coefficient and differences to determine through of information the most important signs choose opportunity was created . Of this as a result data volume optimized only necessary and useful information part separate get of the process to criteria based on algorithm work developed This algorithm application reach through one by one field information analysis to do and them again at work big important have , especially this artificial intellect and with a car teaching algorithms work exit in the fields of systems speed and accuracy such as qualities efficiency in raising important importance occupation is enough

References

1. Bishop, CM (2006). Pattern Recognition and Machine Learning. Springer.
2. Duda , RO, Hart, PE, & Stork, DG (2000). Pattern Classification. Wiley- Interscience .
3. Murphy, KP (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
4. Goodfellow , I., Bengio , Y., & Courville , A. (2016). Deep Learning. MIT Press.
5. Saparov SX Brain cancer early in classification informative characters Complex choose algorithm .// "Muhammad al- Khorazmi descendants " scientific and practical and information-analytical magazine . #3(25). Tashkent-2023. - B. 57-63
6. Saparov SX, Allayarov UB, Gudratov HB Brain cancer the disease early in classification informative characters complex choose algorithm // " People economy in the field advanced technologies application on the subject of " problems ". territorial scientific and technical conference lectures collection . Nukus-2023. -B.159-163.
7. Saparov SX, Allayarov UB, Gudratov HB Brain cancer early in classification objects importance to determine algorithm // " People economy in the field advanced technologies application on the subject of " problems ". territorial scientific and technical conference lectures collection . Nukus-2023. - B.164-166.