

Viktoriia STRILETS¹, Volodymyr DONETS¹, Mykhaylo UGRYUMOV¹,
Sergii ARTIUCH², Roman ZELENSKYI³, Tamara GONCHAROVA⁴

¹ V. N. Karazin Kharkiv National University, Kharkiv, Ukraine

² State Organization "Grigoriev Institute for Medical Radiology and Oncology
of the National Academy of Medical Sciences of Ukraine", Kharkiv, Ukraine

³ Medical Institute Kharkiv Medical Academy of Postgraduate Education, Kharkiv, Ukraine

⁴ National University of Civil Defense of Ukraine, Kharkiv, Ukraine

AGENT-ORIENTED DATA CLUSTERING FOR MEDICAL MONITORING

Medical data processing is one of the priority machine learning areas. Usually, data obtained in the process of medical patient monitoring are complex and have a different nature. Solving the problem of clustering, classification, or forecasting problem these data requires the creation of new methods or improvement of existing methods to improve the decision accuracy and effectiveness. The classical clustering approaches and the c-means fuzzy clustering method were analyzed. Based on the multiagent systems theory, it is proposed to use in the c-means method the separate rules for selecting elites when forming clusters and selecting the best of them in accordance with the chosen intra-cluster distance measures. The result of solving such a problem is the number of clusters, as well as the number of elements in them.

The method quality was tested on Fisher iris data set using three measures of intra-cluster distance: Mahalanobis distance, Mahalanobis distance considering the membership function, and Kullbak-Leibler entropy. The highest accuracy of 98% was obtained for the distance measured by the Kullbak-Leibler entropy. Therefore, this measure was chosen to solve the clustering problem of medical monitoring data for prostate disease. Medical monitoring data were divided into four classes of patient states: "healthy persons", "non-metastatic patients", "metastatic patients" and "hormone-resistant patients". The accuracy of clustering according to medical data was 95,6%. In addition to accuracy, the confusion matrix, ROC- and LF-curves were used to assess the method quality. The minimum value of the ROC-curve was 0.96 for Fisher's irises and 0.95 for medical monitoring data, which characterizes the high quality of the proposed clustering method. The loss function value is also quite small (-0.056 and -0.0176 for each considered data set), which means that the optimal cluster number and the distribution of data over them are obtained.

Based on the obtained results analysis, the proposed method can be recommended for use in medical information and diagnostic decision support systems for clustering monitoring data.

Keywords: clustering; fuzzy clustering; agent-based approach; intraclass distance; medical diagnostic.

Introduction

Medical institutions generate considerable data sets, which contain important information about patients. Apart from the structural data, medical institutions yield big amount of unstructured texts which hold important information about patients' health. They include anamneses, inspection results, inspection result descriptions, such as ultrasonic, electrocardiographic, radiologic etc. Nowadays the clinical test analysis is a fast developing scientific school, which is at the junction of information technologies and medicine [1, 2].

Intelligent methods of treating collected data arrays can automate solving many tasks in the clinical practice, which will improve quality of medical help [3]. Intelligent analysis systems of medical data are used for decision support in disease diagnosis and taking medical steps to control the work of medical staff and to prevent potentially dangerous changes in health status of patients [4]. The developed information medical systems

are based on artificial intelligence and machine learning methods [5, 6], which are used for clustering, classifying and predicting data.

The problem of stratified medicine is important for correctly grouping patients based on disease risk or response to therapy.

In the medical field, clustering has been proven to be a powerful tool to solve problem of the stratification [2, 3]. However when clustering, errors can occur due to an incorrect assumption about the number of classes and the distribution of elements among them.

Therefore, the **goal** of the work is to improve the quality of patient stratification in medical monitoring systems by developing an effective method for data clustering. The **scientific task** of the research is to develop an effective method of data clustering based on the agent-based approach. The method will reduce the probability of incorrect determination of the patients state when monitoring them. Thus, the error of the third kind in the clustering of states will be minimized.

General clustering method analysis

Any method of data clustering [7] builds a mathematical model by which it divides a set of data into clusters. Therefore, the word clustering method will be assumed as a method and its mathematical model. There are two groups of cluster analysis methods: hierarchical and non-hierarchical [8]. Hierarchical methods involve the sequential integration of objects into clusters according to the degree of their proximity to each other or, conversely, the sequential division of object sets into smaller clusters. In this case, the cluster solution is a hierarchical structure of nested clusters. Non-hierarchical methods allow us to find and identify "condensation" of objects in the space of variables. There is also the clustering "with learning", which assumes that the number of classes is known in advance, and there is a learning sample – a set of objects which are known to which classes they belong. The remaining objects are classified according to the degree of their proximity to the objects in the educational sample.

The main hierarchical cluster analysis methods [9] are the neighbor method, the agglomeration method, the full bond method, the King medium bond method, and the Ward method. The most universal is the latter. There are also centroid methods [10] and methods that use the median [10]. The work [11] describes negative aspects of using this type of methods.

Non-hierarchical methods [12] are iterative methods of fragmentation of the original group. In the process of distribution, new clusters are formed till the stopping rule is fulfilled. The methods differ in the choice of starting point, the rule of forming new clusters and the stopping rule. Fast cluster analysis, also called k-means algorithm, is most often used [13 – 15]. When using this algorithm, the analyst enters the number of clusters in the resulting partition in advance.

There are also methods for selecting the number of clusters [16]. One of the most common is the elbow method [17]. The method consists of constructing the variation as a function of the number of clusters and choosing the curve bend representing the number of clusters used. The same method can be used to select the number of parameters in other data-driven models, such as the number of main components to describe a data set.

When choosing between hierarchical and non-hierarchical methods, you should pay attention to the following points [9]. Non-hierarchical methods show higher stability for emissions, incorrect choice of metrics, inclusion of insignificant variables into the base for clustering, etc. The researcher must record in advance the resulting number of clusters, the stopping rule and, if necessary, the initial center of the cluster. The latter condition significantly affects the efficiency of the algorithm. If there

is no reason to intentionally set this condition, it is recommended to use hierarchical methods. We note an essential characteristic for both groups of algorithms: clustering of all observations is not always the right solution. It may be most optimal to clean the sample from emissions firstly, and then continue the analysis. You can also avoid setting a very high stop criterion (you can make a stop, for example, when clustered more than 90% of observations).

We should note that to choose a specific method of clustering, the analyst must be acquainted with the nature and prerequisites of the methods, otherwise the results will resemble the "patients' average temperature in the hospital." To make sure that the selected method is really effective in this case, the following procedure is generally used: several a priori different groups are considered and their representatives are randomly mixed with each other. Then the clustering procedure is carried out in order to restore the output division into groups. An indicator of the effectiveness of the method will be the object coincidence cleaning in the identified and output groups.

Some authors [10, 13, 16] affirm that the choice of metrics and clustering method is not the main point in cluster analysis. This statement is valid in the following cases. Firstly, it applies only to ruder non-hierarchical methods. Secondly, in all cases, it is necessary to choose metrics so that they do not contradict the idea of the chosen method of cluster combining. Particular attention should be paid to the choice of metric if the variables are dependent. An adequate metric can be the solution to this problem.

Clustering problem statement

Let the input data sample be given as:

$$\tilde{X} = (x_i : x_{im} \in X), m = \overline{1, M}, i = \overline{1, n}, \quad (1)$$

where $x_i \in (x_{i1}, x_{i2}, \dots, x_{im})$, X is the input data matrix.

Y is the set of cluster numbers. The function of arrangement between objects $\rho(x, x^1)$ is set. We need to divide the sample into disparate subsets, called clusters, so that each cluster would consist of objects close in metrics, and objects of different clusters differ significantly, wherein each object $x_i \in X^m$ is assigned a cluster number y_i .

The clustering algorithm [7] is a function $a: X \rightarrow Y$, which matches the cluster number $x \in X$ for any object $y \in Y$. The set Y in some cases is known in advance, but more often the problem is to determine the optimal number of clusters, in terms of a criterion for the quality of clustering.

The solution of the clustering problem is fundamentally ambiguous, and there are several reasons for this [7]:

- there is definitely no best criterion for clustering quality. There are a number of heuristic criteria, as well as a number of algorithms that do not have a clear criterion, but carry out a fairly reasonable clustering "to build". They can all give different results;
- the number of clusters is usually unknown in advance and is set according to some subjective criterion;
- the result of clustering significantly depends on the metric, the choice of which is usually also subjective and is determined by the expert.

We determine the distance between the objects as a metric for cluster analysis. The measure of proximity (likeness) of objects may be introduced as a reverse value of the inter-object spacing. Numerous works dedicated to the cluster analysis describe over 50 different methods of calculating inter-object spacing. Apart from the term "spacing" the term "metric", which implies the method of calculating a particular distance is used in many works. The widely used method in terms of quantitative signs is the "Euclidean distance" or "Euclidean metric":

$$d(X_i, X_j) = \left(\sum_{m=1}^M (x_{im} - x_{jm})^2 \right)^{\frac{1}{2}}. \quad (2)$$

The Minkowski distance [9] is generalized in accordance with the Euclidean distance; it uses magnitude p instead of 2. We obtain the usual Euclidean distance at $p = 2$. The expression (3) for the generalized Minkowski metric is shown below.

$$d(X_i, X_j) = \left(\sum_{m=1}^M |x_{im} - x_{jm}|^p \right)^{\frac{1}{p}}. \quad (3)$$

Metrics such as Manhattan distance (distance of city quarters or city-block), metrics of "dominance" (Sup-metric or Supplement-norm) are also used. The Minkowski's metrics are a large family of metrics, including the most popular metrics. However, there are methods for calculating the distance between objects, fundamentally different from the Minkowski's metrics. The most important of these is the Mahalanobis distance [18].

The work [19] deals with the metrics with nominal and ordinal attributes: ordinal – the Spearman, Kendall metrics; nominal – the Jacquard, Russell-Rao, Bravais, Yula metrics. As mentioned earlier, these are not all existing metrics. Accordingly, if the researcher has attributes of other types in addition to quantitative attributes, then the problem arises to find a metric between different types of characteristics measured in different scales. In this case, it is necessary to solve the problem of bringing all different types of scales to one common scale [20].

C-means clustering method

The c-means method of fuzzy clustering (from English fuzzy clustering, soft k-means, c-means) allows us to make a fuzzy classification (breakdown) of elements of the original set with power N into a given number of subsets – "classes". The fuzzy clustering method of c-means differs from the former k-means method in determining the object's belonging to a possible class [21].

The algorithm of c-means [22] includes a sequence of operations:

1) at the first stage, the k centers of clusters $c_j, j = 1, \dots, k$ are randomly set;

2) the matrix of membership of elements to clusters w_{ij} is calculated. At this stage, there are various variations of the method of determining the distance between the possible center and the element of the subset. In the case of normal data distribution, the calculation of the membership matrix will look like this

$$w_{ij} = \frac{N(d(x_i, c_j) | \mu = 0, \sigma_j)}{\sum_{i=1}^{P_j} N(d(x_i, c_j) | \mu = 0, \sigma_j)}, \quad (4)$$

where x_i is the i^{th} element of the set, $i = 1, \dots, P_j$, c_j is the center of the j cluster, $d(x_i, c_j)$ is the distance between the points x_i and c_j , $N(d(x_i, c_j) | \mu = 0, \sigma_j)$ is the probability density of the normal distribution at the point $d(x_i, c_j)$;

3) need to move the centers of the clusters, taking into account the membership matrix

$$c_j \leftarrow \frac{\sum_{i=1}^{P_j} w_{ij} x_i}{\sum_{i=1}^{P_j} w_{ij}}; \quad (5)$$

4) need to calculate the loss function (at this stage you can use the principle of maximum similarity). In the case of a normal distribution, the loss function has the form

$$\text{LossFunc} = \sum_{j=1}^k \sum_{i=1}^{P_j} d(x_i, c_j)^2 w_{ij} \rightarrow \min; \quad (6)$$

5) if the value of the loss function has not decreased, it is necessary to go to point 2. Similarly, there may be a stop function, according to which the output can be achieved earlier at a suitable condition (for example, when achieving a certain number of iterations).

The method of fuzzy clustering of c-means also has a number of significant disadvantages limiting its application. For example, if the target cluster has a more complex shape than just an m-dimensional sphere, or in the case of cluster intersection and high data noise, the fuzzy clustering method can't determine the precise element affiliation to a particular cluster [21, 23].

Most real data sets have big scatters of possible values of state variables of elements of set. It determines the complex M-dimensional form of a possible cluster [23], which makes it impossible to determine the affiliation of elements. There are also difficulties with a large number of clusters and a large number of state variables. Therefore, there is a need to modify the existing c-means method [24, 28]. There are several ways of their modifications, one of the most obvious is the modification of the method of estimating the distance between the center of the cluster and the intended element (or estimating the affiliation of this element of the cluster).

In [25, 28], a modification of the cluster element membership estimation function and a modification of the membership matrix calculation were used. They used the assumption of the Cauchy distribution; to determine the affiliation, the Mahalanobis distances were used:

$$d^2(x_i, c_j) = MD^2(x_i, c_j) = (x_i - c_j)^T \hat{\Sigma}_j^{-1} (x_i - c_j), \quad (7)$$

where M is the dimension of the state space, $\hat{\Sigma} = \Sigma + \lambda E$ is a regularized covariance matrix, λ is a constant, greater than zero.

Then, taking into account the Cauchy distribution assumption, the membership matrix is calculated by the expression [25]

$$w_{ij} = \frac{\rho(x_i, c_j)}{\sum_{i=1}^I \rho(x_i, c_j)}, \quad (8)$$

$$\text{where } \rho(x_i, c_j) = \frac{1}{\pi \eta \left[1 + \frac{MD^2(x_i, c_j)}{\eta^2} \right]}.$$

Similarly, the problem of clustering m-dimensional ellipses is eliminated using the Mixture models and Gaussian mixture models methods, which are similar to the above method and which were successfully used in determining the serviceability / failure state of the device [26] or through the use of derivative information in determining the distance [20].

The experiments showed improvement of clustering quality, but also still showed the dependence on the nature of the sample data distribution, i.e. the new method

did not allow to sufficiently model the dependencies in the data set [22, 25]; the use of a derivative in a modified distance can minimize the sum of square distances within clusters [27].

The idea of considering the nature of data distribution in samples taking into account the relative entropy in fuzzy clustering of c-means was also proposed, according to which the algorithm for calculating clusters was changed [21, 26], but the former measure of calculating location (distance) between elements and centers was left in the form of Euclidean distance, which did not take into account the data entropy [26].

To solve the problems of the basic method of c-means, as well as its modifications Mixture models, Gaussian mixture models [29] and the idea, which was inspired by an attempt to take into account the data entropy [26] and Kulbak-Leibler information distance rating [29] was proposed.

The Kulbak-Leibler distance is an asymmetric measure (it does not satisfy the triangle inequality) of the information difference between the two probability distributions, the method is well established in applied statistics, fluid mechanics and machine learning. In contrast to the considered modifications, the Kulbak-Leibler distance will allow us to estimate the degree of information difference [29].

Before considering the proposed method, we introduce some symbols: $X = \{x_m\}$ is the vector of state variables, which characterizes the record in the set $m = 1, \dots, M$, M is the dimension of the vector of state variables; $F = \{f_i\}$, $i = 1, \dots, I$ is the vector of objective functions, where I is the vector dimension of the objective functions. Take $M_\alpha[f_i]$ and $M_\alpha[x_m]$ as mathematical expectation of f_i and x_m , then D_{f_i} and D_{x_m} is the dispersion of f_i and x_m , σ_{f_i} and σ_{x_m} is the standard deviation f_i and x_m .

The variance and standard deviation f_i for $x_m = \text{var}$ are determined by expressions

$$D_{f_i|x_m} = \text{var}[M_\alpha[f_i[X_{-m}]]], \quad (9)$$

$$\forall n = 1, \dots, M, n \neq m, x_n = \text{const},$$

$$\sigma_{f_i|x_m} = \text{sqrt}[D_{f_i|x_m}]. \quad (10)$$

The following formulas were used to assess the state variable informativeness. Using formula (9), we obtain the informativeness coefficient

$$\beta_{f_i} = \frac{D_{f_i|x_m}}{E_{f_i}}, \quad (11)$$

where E_{f_i} is the signal energy. The signal-to-noise ratio is deduced from the expression (10), which is nothing but the influence coefficient

$$\varphi_{im} = \text{SNR}_{f_i|x_m} = \frac{\sigma_{f_i|x_m}}{\sigma_{x_m}}. \quad (12)$$

The Kullbak-Leibler entropy is determined by the expression [29]

$$D_{KL}(f_i, X) = \sum_{m=1}^M \rho(x_m|f_i) \log_2 \left[\frac{\rho(x_m|f_i)}{\rho(x_m)} \right]. \quad (13)$$

Mutual information is defined by the expression

$$H_{im} = \frac{1}{2} \log_2 \left[\text{SNR}_{f_i|x_m}^2 \right] = \frac{1}{2} \log_2 \left[\beta_{im} \frac{E_{f_i}}{D_{x_m}} \right]. \quad (14)$$

Next, to assess the quality of data clustering, we replace formula (6) with an expression for evaluating reciprocal information, which will be a function of losses in the proposed modification of the c-means method:

$$H(X, M) = - \frac{1}{\sum_{j=1}^k P_j} \sum_{j=1}^k \left[P_j \left(M_j^{(t+1)} \right) \times \sum_{i=1}^{P_j} D_{KL} \left(x_i, M_j^{(t+1)} \right) \right] \rightarrow \min, \quad (15)$$

where M_j are the state variables belonging to the j cluster.

Agent-based model

According to the developed model, it is proposed to divide the population into four states: C – cluster centers, X – agents-elements of the dataset, X_j – agents-elements of clusters, Z – agents-clusters.

The agent-based approach is to consider elements and clusters as agents that choose centers or clusters according to a given measure. That is, the agent-elements choose the cluster that has the center closest to them, according to the intra-cluster distance. And the number of cluster-agents is formed according to the inter-cluster distance so as to ensure a minimum loss function.

The traditional method of c-means does not completely solve the problem of taking into account the complex structure of clusters and reducing the error of clustering of the third kind. Therefore, the advanced clustering method based on an agent-oriented approach has been developed that will solve these problems.

Data clustering method based on agent-based approach

The c-means clustering method was chosen as the basis.

Let the sample of data be considered as $X = \{X_j\}$; $X_j = \{x_{jp}^0\}$ and $N = \sum_{j=1}^{k^*} P_j^*$ is the total number of elements, where P_j^* is the number of elements in the j^{th} cluster; k^* is the number of clusters.

It is necessary to find $\{k_t, P_j\}$, where P_j is the number of elements in the cluster, $j = 1, \dots, k$.

Four measures of intra-cluster distance were used as a metric for the clustering data:

$$d(x_{jp}^0, c_j^0) = \begin{cases} d_1(x_{jp}^0, c_j^0) \text{ (I),} \\ MD^2(x_{jp}^0, c_j^0) \text{ (II),} \\ w_{jp}^{-1} MD^2(x_{jp}^0, c_j^0) \text{ (III),} \\ -D_{KL}(x_{jp}^0, c_j^0) \text{ (IV),} \end{cases} \quad (16)$$

where I is the Manhattan distance, II is the Mahalanobis distance, III is the Mahalanobis distance with membership function and IV is the Kullbak-Leibler entropy.

Let us define the average measure of the intra-cluster distance:

$$M(c_j^0) = \frac{1}{P_j} \sum_{p=1}^{P_j} d(x_{jp}^0, c_j^0). \quad (17)$$

Also the Loss Function (LF) is defined:

$$LF(X) = \frac{1}{k_t} \sum_{j=1}^{k_t} M(c_j^0). \quad (18)$$

Then the statement of the research problem will take the form:

$$\begin{cases} Z = \{k_t, \{P_j\}\} \\ \hat{Z} = \arg \min_{X \in D_X \in R^M} LF(X). \end{cases}$$

It is necessary to determine the number of clusters and distribute the data among clusters so that the value of the loss function is minimal.

To correct the centers of clusters, we use the expression

$$c_j^0 = \frac{\sum_{p=1}^{P_j} w_{jp} x_{jp}^0}{\sum_{p=1}^{P_j} w_{jp}}, \quad (19)$$

where $w_{jp} = \frac{\rho(x_{jp}^0, c_j^0)}{\sum_{p=1}^{P_j} \rho(x_{jp}^0, c_j^0)}$ is the membership matrix

with the Cauchy distribution $\rho(x_{jp}^0, c_j^0) = \frac{1}{\pi \eta \left[1 + \frac{MD^2(x_{jp}^0, c_j^0)}{\eta^2} \right]}$.

Data clustering algorithm based on agent-based approach:

1. Set $k_t^{(n)} > k^*$, $P_t^{(n)} = \text{int} \left(\frac{N}{k_t^{(n)}} \right)$ and randomly

generate cluster centers $\{c_j^0\}$.

2. Using the selected measure with cluster distance $d_1(x_{jp}^0, c_j^0)$ (16), choose $P_t^{(n)}$ the nearest neighbors for each j^{th} cluster.

3. For each j^{th} cluster, using the $\{w_{jp}\}$ and $\left\{ \rho(x_{jp}^0 | M_j) \right\}$, the cluster centers $\{c_j^0\}$ are corrected.

4. For each j^{th} cluster, using the selected measure within the cluster distance $d(x_{jp}^0, c_j^0)$ (16) and $P_t^{(n)}$ the nearest neighbors is choose. Delete the points that are duplicated: $P_t^{(n)} \rightarrow P_j^{(n)}$.

5. For each j^{th} cluster the average measure of the intra-cluster distance $M(c_j^0)$ is calculated, and also $LF(X)$ is calculated.

6. Elite selection. Find the cluster with the largest $M(c_j^0)$ and delete it.

7. $k_t^{(n+1)} = k_t^{(n)} - 1$; $P_t^{(n+1)} = \text{int} \left(\frac{N}{k_t^{(n+1)}} \right)$.

8. Back to step 2, if $k_t^{(n+1)} > 1$.

The method application for toy data

The proposed method was verified on a standard dataset that is used in classification and clustering – a dataset based on Fisher's irises. This dataset consists of data on 150 specimens of irises, 50 specimens each from three species – Iris setosa, Iris virginica and Iris versicolor. For each specimen, four characteristics were measured: the length of the sepal length; the width of the outer sepal width; the length of petal length; the width of the inner

lobe of the petal width. Thus, it is necessary to divide the dataset into three clusters (classes), and one of the classes (Iris setosa) is linearly separable from the other two.

The method was applied to cluster data using three measures of intra-class distance (16): Mahalanobis distance (II), Mahalanobis distance taking into account the membership function (III) and Kullbak-Leibler entropy (IV). The following indicators of the proportion of correct answers of the accuracy algorithm were obtained, the results are given in Table 1.

Table 1
Clustering results for Fisher irises using three measures of intra-cluster distance

	Distance		
	II	III	IV
Accuracy	0.8	0.913	0.98

Based on the results of calculations for each variant, the Confusion Matrix, ROC- and LF-curves were built. For example, for the measure of the intra-class distance in the Kullbak-Leibler entropy form on Table 2 and Fig. 1, 2.

Table 2
Clustering results for Fisher's irises by Kullbak-Leibler entropy

		Predicted class		
		Iris setosa	Iris virginica	Iris versicolor
Actual class	Iris setosa	50	0	0
	Iris virginica	0	48	2
	Iris versicolor	0	1	49

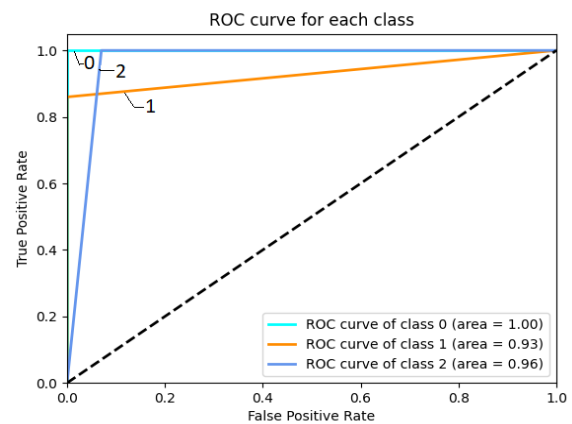


Fig. 1. ROC-curves for Fisher's Iris classes by clustering with Kullbak-Leibler entropy

It was found that one class was clustered completely correctly, and for two classes 3 elements were assigned to other classes by mistake. The value of total loss was reached -0.0565 (Fig. 2).

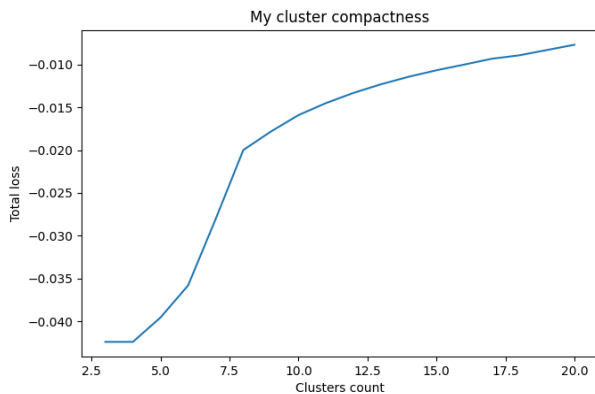


Fig. 2. LF-curves for Fisher's Iris classes by clustering with Kulbak-Leibler entropy

Based on the comparison of the obtained data, it can be concluded that the use of Kullbak-Leibler entropy as a measure of the intra-class distance gave the highest Accuracy and ROC-curves. Thus, the proposed method can be used to cluster real data.

The developed method was verified by comparison with the existing c-means – Python Scikit-learn (sklearn) library in Python (fcmeans: FCM), where the Euclidean distance was used as a measure of intra-cluster distance. The cluster recognition accuracy for the dataset based on Fisher's irises was 0.893, which is lower than in accordance with the proposed method (using measure of intra-class distance – Mahalanobis distance taking into account the membership function) – 0.913.

The advantage of the developed method is the ability to determine not only the number of clusters, but also the number of elements in them, which is especially important in cases where the number of elements in clusters is different. This method, in comparison with analogues, provides a decrease in the probability of incorrect determination of the state of patients (errors of the third kind in the clustering of system states) when observing them.

However, it does not allow dividing a set into connected subsets if they are nested inside each other.

To implement the proposed method, we used Python 3.6.9 and open-source math and data science libraries for Python3: (NumPy – for all math operations, pandas – for loading and processing datasets, SciPy – for calculating distances (e.g. Mahalanobis, Manhattan, Euclidean), sklearn – for calculating some metrics (e.g. AUC-curves, confusion matrix), matplotlib – for drawing graphics). Each part of our software is separated by scripts for specific purposes (implemented as a library): for data processing, for calculating of probability densities of the distribution and membership matrices, for cluster updating, for calculating loss functions, for using the fitted model, and so on. Mixing of these parts can be used in the Python script or in Jupyter-Notebook for clustering on some data with different configurations.

The method application for medical data clustering

The developed method was applied to cluster medical monitoring data for prostate diseases.

Based on a systematic analysis of the process of monitoring patients with prostate diseases, a hierarchy of stages of diagnostics was identified: laboratory, visual diagnosis and controlled variables of patients' state, corresponding to each state of diseases. An experimental dataset of controlled variables characterizing the state of patients observed for the selected disease was formed. The developed method was applied to cluster analysis of collected dataset.

The dimension of the dataset of state variables was 24.

The case histories of 180 patients with prostate diseases were analyzed retrospectively: 50 with benign prostatic hyperplasia and 130 with prostate cancer.

An experimental sample of registered controlled variables was formed, characterizing the state of the observed patients for the selected disease. The sample was divided into 4 classes:

- «Healthy» (benign formations) – 50 people (I);
- «Non-metastatic» – 45 people (II);
- «Metastatic» – 52 people (III);
- «Hormone-resistant» – 33 people (IV).

The level of disease progression was chosen as a classification criterion when dividing the total sample into classes.

According to the calculation results for the variant using the Kullbak-Leibler entropy intra-class distance as a measure, Confusion Matrix, ROC-and Loss function-curves were obtained (see Table 3 and Fig. 3, 4). Accuracy reached 0.956 and total loss was -0.0176.

Table 3

Results of clustering medical data by Kullbak-Leibler entropy

		Predicted class			
		I	II	III	IV
Actual class	I	50	0	0	0
	II	1	42	2	0
	III	1	0	50	1
	IV	0	1	2	30

For each cluster, several elements were identified incorrectly by belonging to the original, but the total number of such elements is small. In general, it can be argued about the high accuracy of the clustering results obtained using the developed method.

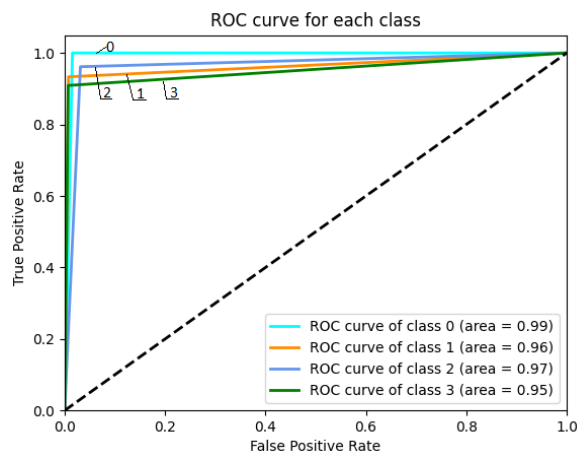


Fig. 3. ROC-curves for medicine data classes by clustering with Kullbak-Leibler entropy

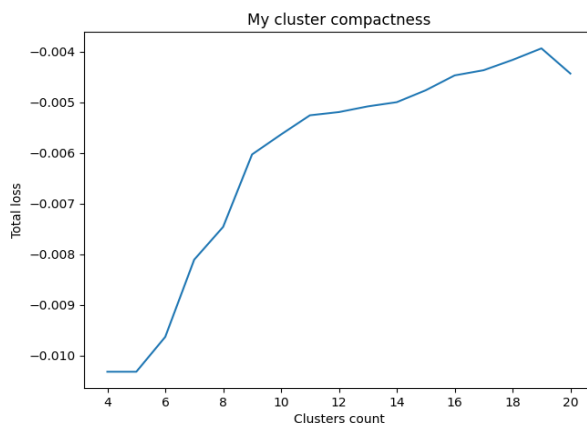


Fig. 4. LF-curves for medicine data classes by clustering with Kullbak-Leibler entropy

Conclusions

Information technologies are increasingly being introduced into all spheres of human life, and medicine is no exception. Medical monitoring data is one of the main areas of machine learning.

The paper considers the problem of data clustering and analyzes two popular methods for its solution, the k-means and c-means method. A data clustering method based on an agent-based approach is proposed, again using the c-means method.

Based on the agent-based approach and in accordance with the chosen measures of intra-cluster distance, permissive elite selection rules are proposed for the formation of clusters, the selection of the best of them, and also for the selection of elements into clusters in the process of solution synthesis.

The result of solving such a problem is the number of clusters, as well as the number of elements in them.

Clustering of data was carried out for two data sets:

Fisher's irises and the results of medical monitoring of indicators for diseases of the prostate gland.

Based on the analysis of the results of clustering data on Fisher's irises, it was found that the proposed method gives the best performance using the Kullbak-Leibler entropy intra-cluster distance (98%) as a measure.

When solving the problem of clustering medical monitoring data, the proposed clustering method based on the agent-based approach with the Kullbak-Leibler entropy measure gave an accuracy of 95.6%.

Based on the analysis of the results obtained, the proposed method can be recommended for use in medical information and diagnostic decision support systems.

Contribution of authors: review and analysis of references, analysis of clustering method – **V. Strilets**; selection and application of software and hardware tools for implementation of clustering method, presentation of clustering results of medical monitoring data – **V. Donets**; formulation of the purpose and tasks of research, development of clustering methods – **M. Ugryumov**; collection and description of medical monitoring data on prostate diseases – **S. Artiukh**; development of conceptual provisions and methodology of research, analysis of the research results – **R. Zelenskiy**; analysis of using the machine learning methods in the medicine, formulation of conclusions – **T. Goncharova**. All authors have read and agreed to the published version of the manuscript.

References (GOST 7.1:2006)

1. Berka, P. *Data mining and medical knowledge management: Cases and applications* [Text] / P. Berka, S. Rauch, D. Zighed. – N.Y. : Hershey, 2009. – 464 p. DOI: 10.4018/978-1-60566-218-3.
2. Karahoca, A. *Data mining applications in engineering and medicine* [Text] / A. Karahoca. – London : IntechOpen, 2012. – 338 p. DOI: 10.5772/2616.
3. Kountchev, R. *Advances in intelligent analysis of medical data and decision support systems* [Text] / R. Kountchev, B. Iantovics. – Springer, 2013. – 247 p. DOI: 10.1007/978-3-319-00029-9.
4. Giannopoulou, E. G. *Data Mining in Medical and Biological Research* [Text] / E. G. Giannopoulou. – London : IntechOpen, 2008. – 332 p. DOI: 10.5772/95.
5. *Adaptive Probabilistic Neuro-Fuzzy System and its Hybrid Learning in Medical Diagnostics Task* [Text] / Ye. Bodyanskiy, A. Deineko, I. Pliss, O. Chala // *The Open Bioinformatics Journal*. – 2021. – Vol. 14. – P. 123-129. DOI: 10.2174/18750362021140100123.
6. *A Method for Assessing the Risks of Complications in Chemoradiation Treatment of Squamous Cell Carcinoma of the Head and Neck* [Text] / V. Starenkiy, S. Artiukh, M. Ugryumov, V. Strilets, S. Chernysh, D. Chumechenko // *The Open Bioinformatics Journal*. –

2021. – Vol. 14. – P. 138-143. DOI: 10.2174/18750362021140100138.

7. Schlesinger, M. *Ten lectures on statistical and structural pattern recognition [Text]* / M. Schlesinger, V. Hlavac. – Springer, Dordrecht, 2002. – 522 p. DOI: 10.1007/978-94-017-3217-8.

8. Wishart, D. *Exploiting the graphical user interface in statistical software: the next generation [Text]* / D. Wishart // *Interface '98. Computing Science and Statistics.* – 1998. – Vol. 30. – P. 257-263.

9. *Data clustering: algorithms and applications [Text]* / edited by Charu C. Aggarwal, Chandan K. Reddy. – CRC Press, Taylor & Francis Group, 2014. – 622 p.

10. Manning, Ch. D. *Introduction to Information Retrieval [Text]* / Ch. D. Manning, P. Raghavan, H. Schutze. – Cambridge University Press, 2008. – 506 p.

11. Sneath, P. H. A. *Numerical Taxonomy: The Principles and Practice of Numerical Classification [Text]* / P. H. A. Sneath, R. R. Sokal. – San Francisco : W. H. Freeman and Company, 1973. – 573 p.

12. Giordani, P. *Non-Hierarchical Clustering* / P. Giordani, M.B. Ferraro, F. Martella // *An Introduction to Clustering with R. Behaviormetrics: Quantitative Approaches to Human Behavior.* – 2020. – Vol 1. – P. 75-109. DOI: 10.1007/978-981-13-0553-5_3.

13. *A Handbook of Soil Terminology, Correlation and Classification [Text]* / P. Krasilnikov, J.-J.I. Marti, R. Arnold, S. Shoba. – London, Earthscan, 2009. – 449 p.

14. Wang, Sh. *On the Efficiency of K-Means Clustering: Evaluation, Optimization, and Algorithm Selection [Text]* / Sh. Wang, Y. Sun, Z. Bao // *Proceedings of the VLDB Endowment.* – 2020. – Vol. 14. – P. 163-175. DOI: 10.14778/3425879.3425887.

15. Тулякова, Н. О. *Локально-адаптивная фильтрация нестационарного шума в длительных электрокардиографических сигналах [Текст]* / Н. О. Тулякова, А. Н. Трофимчук // *Радіоелектронні і комп'ютерні системи.* – 2020. – № 4(96). – С. 16–33. DOI: 10.32620/reks.2020.4.02.

16. MacKay, D. *Chapter 20. An Example Inference Task: Clustering [Text]* / D. MacKay // *Information Theory, Inference and Learning Algorithms.* – Cambridge University Press, 2003. – P. 284–292.

17. *K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based on Mean and Median [Text]* / E. Umargono et al. // *Proceedings of the International Conferences on Information System and Technology.* – 2018. – P. 234-240. DOI: 10.5220/0009908402340240.

18. Kinlaw, W. B. *A New Index of the Business Cycle [Text]* / W. B. Kinlaw, M. P. Kritzman, D. Turkington // *Macroeconomics: Prices.* – 2020. – 30 p. DOI: 10.2139/ssrn.3521300.

19. Гайдышев, И. П. *Анализ и обработка данных. Специальный справочник [Текст]* / И. П. Гайдышев. – СПб. : «Пумер», 2001. – 752 с.

20. Finn, V. K. *The Heuristics of Detection of Empirical Regularities by JSM Reasoning [Text]* /

V. K. Finn, O. P. Shesternikova // *Automatic Documentation and Mathematical Linguistics.* – 2018. – Vol. 52, Iss. 5. – P. 215–247. DOI: 10.3103/S0005105518050023.

21. Bakumenko, N. *Application of the C-Means Fuzzy Clustering Method for the Patient's State Recognition Problems in the Medicine Monitoring Systems [Text]* / N. Bakumenko, V. Strilets, M. Ugryumov // *CEUR Workshop Proceedings of 3rd International Conference on Computational Linguistics and Intelligent Systems, COLINS 2019.* – 2019. – Vol. I: Main Conference. – P. 218-227. Available at: <https://www.researchgate.net/publication/338819685> (accessed 15.02.2021).

22. Winkler, R. *Problems of Fuzzy c-Means Clustering and Similar Algorithms with High Dimensional Data Sets [Text]* / R. Winkler, F. Klawonn, R. Kruse // *Challenges at the Interface of Data Analysis, Computer Science, and Optimization.* – 2012. – P. 79–87. DOI: 10.1007/978-3-642-24466-7_9.

23. Amruthnath, N. *Fault Class Prediction in Unsupervised Learning using Model-Based Clustering Approach [Text]* / Nagdev Amruthnath, Tarun Gupta // *2018 International Conference on Information and Computer Technologies (ICICT).* – 2018. – P. 5-12. DOI: 10.13140/RG.2.2.22085.14563.

24. Askari, S. *Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development [Text]* / S. Askari // *Expert Systems with Applications.* – 2020. – vol. 165. – Article no. 113856. DOI: 10.1016/j.eswa.2020.113856.

25. Zarinbala, M. *Relative entropy fuzzy c-means clustering [Text]* / M. Zarinbala, M. H. Fazel Zarandia, I.B. Turksen // *Information Sciences.* – 2014. – Vol. 260. – P. 74-97. DOI: 10.1016/j.ins.2013.11.004.

26. *A new distance with derivative information for functional k-means clustering algorithm [Text]* / Yinfeng Menga, Jiye Liangb, Fuyuan Caob, Yijun He // *Information Sciences.* – 2018. – Vol. 463–464. – P. 166-185. DOI: 10.1016/j.ins.2018.06.035.

27. Siddique, A. B. *Implementation of Fuzzy C-Means and Possibilistic C-Means Clustering Algorithms, Cluster Tendency Analysis and Cluster Validation [Text]* / Md. Abu Bakr Siddique, Rezoana Bente Arif, Mohammad Mahmudur Rahman Khan, Zahidun Ashrafi // *ArXiv e-Journal.* – 2019. – 8 p. Available at: <https://arxiv.org/abs/1809.08417v3> (accessed 27.02.2021).

28. Nielsen, F. *On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means [Text]* / F. Nielsen // *Sony Computer Science Laboratories, Entropy.* – 2019. – Vol. 21, Iss. 5. – Article no. 485. DOI: 10.3390/e21050485.

29. Møllersen, K. *On Data-Independent Properties for Density-Based Dissimilarity Measures in Hybrid Clustering [Text]* / K. Møllersen, S. Dhar, F. Godtliebsen // *Applied Mathematics.* – 2016. – Vol. 7, no. 15. – P. 1674-1706. DOI: 10.4236/am.2016.715143.

References (BSI)

1. Berka, P., Rauch, S., Zighed, D. *Data mining and medical knowledge management: Cases and applications*. N.Y., Herskey, 2009. 464 p. DOI: 10.4018/978-1-60566-218-3.
2. Karahoca, A. *Data mining applications in engineering and medicine*. London, IntechOpen, 2012. 338 p. DOI: 10.5772/2616.
3. Kountchev, R., Iantovics, B. *Advances in intelligent analysis of medical data and decision support systems*. Springer, 2013. 247 p. DOI: 10.1007/978-3-319-00029-9.
4. Giannopoulou, E. G. *Data Mining in Medical and Biological Research*. London, IntechOpen, 2008. 332 p. DOI: 10.5772/95.
5. Bodyanskiy, Ye., Deineko, A., Pliss, I., Chala, O. Adaptive Probabilistic Neuro-Fuzzy System and its Hybrid Learning in Medical Diagnostics Task. *The Open Bioinformatics Journal*, 2021, vol. 14, pp. 123-129. DOI: 10.2174/18750362021140100123.
6. Starenkiy, V., Artiukh, S., Ugryumov, M., Strilets, V., Chernish, S., Chumachenko, D. A Method for Assessing the Risks of Complications in Chemoradiation Treatment of Squamous Cell Carcinoma of the Head and Neck. *The Open Bioinformatics Journal*, 2021, vol. 14, pp. 138-143. DOI: 10.2174/18750362021140100138.
7. Schlesinger, M., Hlavac, V. *Ten lectures on statistical and structural pattern recognition*. Springer, Dordrecht, 2002. 522 p. DOI: 10.1007/978-94-017-3217-8.
8. Wishart, D. Exploiting the graphical user interface in statistical software: the next generation. *Interface '98. Computing Science and Statistics*, 1998, no. 30, pp. 257-263.
9. Charu, C. Aggarwal., Chandan, K. Reddy (ed.). *Data clustering: algorithms and applications*. CRC Press, Taylor & Francis Group, 2014. 622 p.
10. Manning, Ch. D., Raghavan, P., Schutze, H. *Introduction to Information Retrieval*. Cambridge University Press, 2008. 506 p.
11. Sneath, P. H. A., Sokal, R. R. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. San Francisco, W.H. Freeman and Company, 1973. 573 p.
12. Giordani, P., Ferraro, M. B., Martella, F. Non-Hierarchical Clustering. *An Introduction to Clustering with R. Behaviormetrics: Quantitative Approaches to Human Behavior*, 2020, vol. 1, pp. 75-109. DOI: 10.1007/978-981-13-0553-5_3.
13. Krasilnikov, P., Marti, J.-J. I., Arnold, R., Shoba, S. *A Handbook of Soil Terminology, Correlation and Classification*. London, Earthscan, 2009. 449 p.
14. Wang, Sh., Sun, Y., Bao, Z. On the Efficiency of K-Means Clustering: Evaluation, Optimization, and Algorithm Selection. *Proceedings of the VLDB Endowment*, 2020, vol. 14, pp. 163-175. DOI: 10.14778/3425879.3425887.
15. Tulyakova, N. O., Trofimchuk, A. N. Lokal'no-adaptivnaya fil'tratsiya nestatsionarnogo shuma v dlitel'nykh elektrokardiograficheskikh signalakh [Locally adaptive filtering of non-stationary noise in long-term electrocardiographic signals]. *Radioelektronni i komp'yuterni sistemi – Radioelectronic and computer systems*, 2020, no. 4(96), pp. 16-33. DOI: 10.32620/reks.2020.4.02.
16. MacKay, David. Chapter 20. An Example Inference Task: Clustering. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003, pp. 284-292.
17. Umargono, E., Suseno, J. E., S.K., V. K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median. *Proceedings of the International Conferences on Information System and Technology*, 2018, pp. 234-240. DOI: 10.5220/0009908402340240.
18. Kinlaw, W. B., Kritzman, M. P., Turkington, D. A New Index of the Business Cycle. *Macroeconomics: Prices*, 2020. 30 p. DOI: 10.2139/ssrn.3521300.
19. Gaidyshev, I. G. *Analiz i obrabotka dannykh. Spetsialnyi spravochnik* [Analysis and data processing: a special reference book]. Sankt Peterburg, Peter, 2001. 752 p.
20. Finn, V. K., Shesternikova O. P. The Heuristics of Detection of Empirical Regularities by JSM Reasoning. *Automatic Documentation and Mathematical Linguistics*, 2018, vol. 52, Issue 5, pp. 215-247. DOI: 10.3103/S0005105518050023.
21. Bakumenko, N., Strilets, V., Ugryumov, M. Application of the C-Means Fuzzy Clustering Method for the Patient's State Recognition Problems in the Medicine Monitoring Systems. *CEUR Workshop Proceedings of 3rd International Conference on Computational Linguistics and Intelligent Systems, COLINS 2019*, 2019, vol. I, pp. 218-227. Available at: <https://www.researchgate.net/publication/338819685> (accessed 15.02.2021).
22. Winkler, R., Klawonn, F., Kruse, R. Problems of Fuzzy c-Means Clustering and Similar Algorithms with High Dimensional Data Sets. *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, 2012, pp. 79-87. DOI: 10.1007/978-3-642-24466-7_9.
23. Amruthnath, N., Gupta, T. Fault Class Prediction in Unsupervised Learning using Model-Based Clustering Approach. *2018 International Conference on Information and Computer Technologies (ICICT)*, 2018, pp. 5-12. DOI: 10.13140/RG.2.2.22085.14563.
24. Askari, S. Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*, 2020, vol. 165, article no. 113856. DOI: 10.1016/j.eswa.2020.113856.
25. Zarinbala, M., Zarandia, Fazel M. H., Turksen, I.B. Relative entropy fuzzy c-means clustering. *Information Sciences*, 2014, vol. 260, pp. 74-97. DOI: 10.1016/j.ins.2013.11.004.
26. Menga, Yinfeng., Liangb, Jiye., Caob, Fuyuan., He, Yijun. A new distance with derivative information for functional k-means clustering algorithm. *Information*

Sciences, 2018, vol. 463–464, pp. 166–185. DOI: 10.1016/j.ins.2018.06.035.

27. Siddique, A. B., Arif, R. B., Khan, M. M. R., Ashrafi, Z. Implementation of Fuzzy C-Means and Possibilistic C-Means Clustering Algorithms, Cluster Tendency Analysis and Cluster Validation. *ArXiv e-Journal*, 2019. 8 p. Available at: <https://arxiv.org/abs/1809.08417v3> (accessed 27.02.2021).

28. Nielsen, F. On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means. *Entropy*, 2019, vol. 21, iss. 5, article no. 485. DOI: 10.3390/e21050485.

29. Möllersen, K., Dhar, S., Godtliebsen, F. On Data-Independent Properties for Density-Based Dissimilarity Measures in Hybrid Clustering. *Applied Mathematics*, 2016, vol. 7, no. 15, pp. 1674–1706. DOI: 10.4236/am.2016.715143.

Надійшла до редакції 07.07.2021, розглянута на редколегії 16.02.2021

АГЕНТНО-ОРИЄНТОВА КЛАСТЕРИЗАЦІЯ ДАНИХ ДЛЯ МЕДИЧНОГО МОНІТОРИНГУ

**В. Є. Стрілець, В. В. Донець, М. Л. Угрюмов, С. В. Артюх,
Р. О. Зеленський, Т. А. Гончарова**

Обробка медичних даних є одним з пріоритетних напрямків машинного навчання. Розв'язання задач кластеризації, класифікації або прогнозування даних, які отримані у процесі медичного моніторингу станів пацієнтів, потребує створення нових або удосконалення існуючих методів для підвищення точності і ефективності результатів. У роботі були проаналізовані класичний метод кластеризації k-середніх та метод нечіткої кластеризації c-середніх. На основі теорії мультиагентних систем запропоновано використовувати у методі c-середніх роздільні правила елітного відбору при формуванні кластерів та відбору найкращих з них у процесі селекції, відповідно до вибраних заходів внутрішньокластерної відстані. Результатом розв'язання такої задачі є число кластерів, а також кількість елементів у них.

Якість роботи методу перевірена на наборі даних по ірисам Фішера із використанням трьох мір внутрішньокластерної відстані: відстань Махаланобіса, відстань Махаланобіса з урахуванням функції приналежності та ентропія Кульбака–Лейблера. Найвища точність у 98% була отримана для міри відстані за ентропією Кульбака–Лейблера. Тому саме ця міра була обрана для розв'язання задачі кластеризації даних медичного моніторингу за захворюваннями передміхурової залози. Дані моніторингу були розподілені за чотирма класами за станами пацієнтів. Точність кластеризації за медичними даними склала 95,6%. Крім точності для оцінки якості роботи методу кластеризації були використані матриця похибок, ROC- та LF-криві. Мінімальне значення за ROC-кривою отримане 0.96 для ірисів Фішера та 0.95 для даних медичного моніторингу за захворюваннями передміхурової залози, що характеризує високу якість запропонованого методу. Значення функції втрат досягнуте достатньо мале (-0.056 і -0.0176 для кожного з розглядуваних наборів даних), тобто отримані оптимальна кількість кластерів і ефективне розподілення даних за ними.

На основі аналізу отриманих результатів можна рекомендувати пропонований метод для використання в медичних інформаційно-діагностичних системах підтримки прийняття рішень.

Ключові слова: кластеризація; нечітка кластеризація; агентно-орієнтований підхід; внутрішньокласова відстань; медична діагностика.

АГЕНТНО-ОРИЕНТИРОВАННАЯ КЛАСТЕРИЗАЦИЯ ДАННЫХ ДЛЯ МЕДИЦИНСКОГО МОНИТОРИНГА

**В. Е. Стрелец, В. В. Донец, М. Л. Угрюмов, С. В. Артюх,
Р. А. Зеленский, Т. А. Гончарова**

Обработка медицинских данных является одним из приоритетных направлений машинного обучения. Решение задач кластеризации, классификации или прогнозирования данных, полученных в процессе медицинского мониторинга состояний пациентов, требует создания новых или усовершенствования существующих методов для повышения точности и эффективности результатов. В работе были проанализированы классический метод кластеризации k-средних и метод нечеткой кластеризации c-средних. На основе теории мультиагентных систем предложено использовать в методе c-средних отдельные правила элитного отбора при формировании кластеров и отбора лучших из них в процессе селекции, в соответствии с выбранными мерами внутрикластерного расстояния. Результатом решения такой задачи является число кластеров, а также количество элементов в них.

Качество работы метода проверено на наборе данных по ирисам Фишера с использованием трех мер внутрикластерного расстояния: расстояние Махаланобиса, расстояние Махаланобиса с учетом функции принадлежности и энтропия Кульбака–Лейблера. Самая высокая точность в 98% была получена для меры расстояния по энтропии Кульбака–Лейблера. Поэтому именно эта мера была выбрана для решения задачи кластеризации данных медицинского мониторинга по заболеваниям предстательной железы. Данные мониторинга были распределены по четырем классам. Точность кластеризации по медицинским данным составила 95,6%.

Кроме точности для оценки качества работы метода были использованы матрица ошибок, ROC- и LF-кривые. Минимальное значение по ROC-кривой получено 0.96 для ирисов Фишера и 0.95 для данных медицинского мониторинга, что характеризует высокое качество предложенного метода кластеризации. Значение функции потерь также достигнуто достаточно малое (-0.056 и -0.0176 для каждого из рассматриваемых наборов данных), т.е. получено оптимальное число кластеров и распределение данных по ним.

На основе анализа полученных результатов можно рекомендовать разработанный метод для использования в медицинских информационно-диагностических системах поддержки принятия решений.

Ключевые слова: кластеризация; нечеткая кластеризация; агентно-ориентированный подход; внутри-классовое расстояние; медицинская диагностика.

Стрілець Вікторія Євгенівна – канд. техн. наук, доц. каф. теоретичної та прикладної системотехніки, Харківський національний університет імені В. Н. Каразіна, Харків, Україна.

Донець Володимир Віталійович – асп. каф. теоретичної та прикладної системотехніки, Харківський національний університет імені В.Н. Каразіна, Харків, Україна.

Угрюмов Михайло Леонідович – д-р техн. наук, проф., проф. каф. теоретичної та прикладної системотехніки, Харківський національний університет імені В. Н. Каразіна, Харків, Україна.

Артюх Сергій Вікторович – канд. мед. наук, старш. наук. співроб. групи променевої терапії відділу радіології ДУ «Інститут медичної радіології та онкології ім. С. П. Григор'єва НАМН України, асист. каф. радіології та радіаційної медицини Харківського національного медичного університету, Харків, Україна.

Зеленський Роман Олександрович – асист. каф. урології та дитячої урології, Харківська медична академія післядипломної освіти, Харків, Україна.

Гончарова Тамара Анатоліївна – викл. каф. управління та організації діяльності у сфері цивільного захисту, Національний університет цивільного захисту України, Харків, Україна.

Viktoriia Strilets – Candidate of Technical Sciences, Associate Professor of the Theoretical and Applied system engineering Department, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine, e-mail: viktoriia.strilets@karazin.ua, ORCID: 0000-0002-2475-1496, Scopus Author ID: 57208899640.

Volodymyr Donets – PhD student of the Theoretical and Applied system engineering Department, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine, e-mail: vol.donets@gmail.com, ORCID: 0000-0002-5963-9998, Scopus Author ID: 57219929489.

Mykhaylo Ugryumov – Doctor of Technical Sciences, Professor, Professor of the Theoretical and Applied system engineering Department, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine, e-mail: m.ugryumov@karazin.ua, ORCID: 0000-0003-0902-2735, Scopus Author ID: 55819006000.

Sergii Artiuch – Candidate of Technical Sciences, senior researcher of the radiotherapy group of the radiology department State Organization “Grigoriev Institute for Medical Radiology and Oncology of the National Academy of Medical Sciences of Ukraine”, Assistant Professor of Radiology and Radiation Medicine, Kharkiv National Medical University, Kharkiv, Ukraine, e-mail: artiukhsrgii@ukr.net, ORCID: 0000-0002-7189-3614.

Roman Zelenskyi – Assistant of the department of Urology and Pediatric Urology, Kharkiv medical academy of postgraduate education, Kharkiv, Ukraine, e-mail: le4impo4ki@gmail.com, ORCID: 0000-0001-7002-3388.

Tamara Goncharova – teacher of the Department of Management and organization in the field of civil protection, National University of Civil Defence of Ukraine, Kharkiv, Ukraine, e-mail: super-gusenichka@ukr.net, ORCID: 0000-0003-3210-3867, Scopus Author ID: 57219923486.