APPLICATION OF THE METHOD OF FUZZY CLUSTERING OF K-MEANS IN THE PROBLEMS OF RECOGNITION OF PATIENTS IN MEDICAL MONITORING SYSTEMS

Zaiarna Viktoriia, student of 304dep.*
National Aerospace University – «Kharkov Aviation Institute»

The mathematical apparatus clustering is widely used for diagnostic purposes, solving classification tasks and find new patterns for establishing new scientific hypotheses. In this paper the actual problem of clustering data in medicine.

The analysis of the main methods of clustering, as well as the substantiation of the choice of the k-mean method, was carried out. Its main advantages are the versatility, speed and ease of software implementation. Also, the method of k-means flexible to use a variety of metrics and changes.

The k-means algorithm builds k clusters are located at possibly large distances from each other. The main type of tasks solved by the k-means algorithm is the presence of assumptions (hypotheses) on the number of clusters, while they must be different as far as possible. The choice of quantity k may be based on the results of previous studies, theoretical considerations or intuitions.

In this paper, the method of fuzzy clustering is used to modify the k-means method, which allows each object to belong to different degrees to several or all clusters at the same time. The number of clusters is known in advance.

The initial information is a sample of observations generated from N the n-dimensional vector of signs $X=\{x(1),x(2),\ldots,x(N)\}$ $x(k)\in X$, $k=1,2,\ldots,N$. The result of the method is to split the initial array of data into m classes with some degree $w_j(k)$ of membership of the k-th feature vector of the j-th cluster. The objective function subject to minimize is:

$$E(w_j(k),c_j)=\sum_{k=1}^{N}\sum_{j=1}^{m}w_j^{\beta}(k)d^2(x(k),c_j)\rightarrow\min \tag{1}$$

with restrictions:

$$\sum_{j=1}^{m}w_j(k)=1,\ \ k=1,\ldots,n, \tag{2}$$

$$0<\sum_{k=1}^{N}w_j(k)<N,\ \ j=1,\ldots,m. \tag{3}$$

Here $w_j(k)\in[0,1]$ – is the level of belonging of the vector $x(k)$ to the j-th cluster, c_j – the centroid of the j-th cluster, $d^2(x(k),c_j)$ – the distance between

$x(k)$ and c_j in the accepted metric, β – the non-negative parameter, called the "phasicifier" (if used $d^2(x(k),c_j)$ as an Euclidean distance, is taken as 2).

The work of the algorithm begins with the assignment of the initial random matrix of the fuzzy partition W_0. According to its values, the initial set of prototype centers c_j^0 is calculated according to the formula:

$$c_j = \frac{\sum_{k=1}^{N} w_j^{\beta}(k)x(k)}{\sum_{k=1}^{N} w_j^{\beta}(k)}. \qquad (4)$$

Based on the calculated prototype centers c_j^0 , the matrix is calculated W_1 in accordance with the formula:

$$w_j = \frac{(d^2(x(k),c_j))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}(d^2(x(k),c_l))^{\frac{1}{1-\beta}}}. \qquad (5)$$

After that, in batch mode $c_j^1, W^2, \ldots, W^t, c_j^t, W^{t+1}$ and so on, until the difference between the current and the next values of the matrix W will not be less than the specified threshold of accuracy. Thus, the entire available sampling data is processed repeatedly.

As a result of the algorithm get fuzzy partition matrix in which patients will be divided into clusters (diagnoses). The shape of the clusters can vary from the hypershape to the hyperlipsoid, depending on the form of the source data, that is, from the choice of the distance between $x(k)$ and c_j:

$$d(x(k),c_j) = \sqrt{(x(k)-c_j)^T A_j (x(k)-c_j)}, \qquad (6)$$

where A_j – is a matrix that can be defined as the inverse fuzzy covariance matrix of each cluster.

If we take A_j as a matrix identity matrix, the result is the Euclidean distance $d(x(k),c_j) = \sqrt{(x(k)-c_j)^T (x(k)-c_j)}$, and form clusters will round (hipershary).

To give clusters the form of hyperlipidsomes as a matrix A_j can use a symmetric positive definite matrix, that is, a matrix in which all eigenvalues are real and positive.

As a result of the clustering algorithm, we obtain the division of our data into homogeneous clusters, which can take the form of arbitrarily oriented hyperellipsoid spaces. Also, the degree of belonging of each object to each cluster will be known $w_j(k)$.

**Scientific advisor –*
D.Sc., Professor of department 304 M.L. Ugryumov.