

Online Hybrid Probabilistic-Fuzzy Clustering in Medical Data Mining Tasks

Yevgeniy Bodyanskiy¹[0000-0001-5418-2143], Anastasiia Deineko¹[0000-0002-3279-3135],

Iryna Pliss¹[0000-0001-7918-7362] and Olha Chala¹[0000-0002-7603-1247]

¹ Control systems research laboratory, Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine

yevgeniy.bodyanskiy@nure.ua

anastasiya.deineko@gmail.com

iryna.pliss@nure.ua

olha.chala@nure.ua

Abstract. In the paper the online fuzzy clustering recurrent procedure has been introduced that allows the forming of hyperellipsoidal clusters with an arbitrary orientation of the axes is investigated. The proposed clustering system is the generalization of a number of known algorithms, it is intended to solve tasks within the general problems of Medical Data Mining, when information is sequentially fed to processing in online mode.

Keywords: Medical Data Mining; Big Data; Computational Intelligence; Fuzzy Clustering; EM-Algorithm; Kohonen's Self-Learning; Soft Clustering.

1 Introduction

Clustering task has a special place in the general problem of Data Mining [1,2] since it's solution implements in self-learning mode (unsupervised learning) when the researcher doesn't have a marked-up training dataset in advance. It is clear that here the level of a priori uncertainty is much higher comparatively to other problems of data analysis, which led to the emergence of a variety of approaches, methods and algorithms for this problem solving [3-6], differing both in initial premises and in mathematical procedures, which often leads to different results in the end. This is distinguishes the task of clustering from other traditional Data Mining tasks such as classification, forecasting, identification of hidden dependencies contained in data, etc.

Clustering task is complicated if the data for processing are received sequentially in online mode, forming a data stream [7], in doing so the frequency of data income is such that it's impossible to process an accumulated information in time between two neighboring observations. The situation is even more complex if the amount of data is so large that it's processing is impossible as a single array (Big Data concept [8]). Hence, the idea of online clustering of data flow seems to be very attractive, especially in Medical Data Mining task, connected with mass examination of patients.

Artificial neural networks, fuzzy reasoning systems, and hybrid neuro-fuzzy systems [3, 9, 10] can be successfully used to solve the problems of processing data in online mode, while performance issues come to the fore, especially learning processes. Obviously, that multi-epoch learning in this situation is ineffective. The incremental learning is more promising when the parameters of the clustering system are refined sequentially synchronously with data arrival. Here, first of all, it is necessary to note the clustering Kohonen's neural networks [9] – self-organizing maps (SOM), which have shown their effectiveness in solving many real-world problems. It should be remembered that SOMs solve clustering problems under convex (linearly separable) nonoverlapped classes.

In real-world problems, data usually form overlapping classes, wherein each observation simultaneously belonging to two or more classes. It's clear, in this case, in the process of clustering, it is necessary to calculate both the possible classes and the probability membership levels of each vector-pattern to each of the possible classes. Obviously, that in this situation, the traditional Kohonen's neural network is ineffective. In this case, the so-called soft computing methods come to the fore, among them the most popular is the Expectation-Maximization approach (EM-algorithm) [9-15], which is based on probabilistic assumptions. It should also be noted a powerful method of Fuzzy-C-Means (FCM) [10-12] proposed by J.C.Bezdek. Both of these methods were combined in the hybrid approach proposed in [15].

However, it must be noted that EM and FCM are algorithms that process information in batch form in multi-epoch mode, which makes it impossible to use them in Data Stream Mining tasks. In this regard, recurrent adaptive FCM versions operating in the online sequential mode were introduced in [16, 17]. The disadvantages of this approach include the use of conventional Euclidean metrics, which allows the formation of clusters of only a spherical shape. Obviously that in the class conditions of a complex nonconvex form, there may be too many such spheroid clusters, which reduces the speed of the processing.

In this regard, it is advisable to develop recurrent online algorithms of sequential fuzzy clustering which allow forming data clusters of a more complex form than the hyperspheres and, particularly, hyprellipsoidal one arbitrarily oriented in the feature space of overlapping classes.

2 Batch Clustering Using Probabilistic And Fuzzy Approaches

Let the initial data array be given in the dataset form of n-dimensional observation vectors such that $x(k) = (x_1(k), \dots, x_n(k))^T \in R^n$, where $k = 1, 2, \dots, N$ describes the number of the observation in this dataset. As a result of clustering, this dataset should be divided into m ($1 < m < N$) overlapping ellipsoidal classes.

In the framework of the statistical approach (EM-algorithm), it is assumed that the data are of a random nature and have a Gaussian density distribution:

$$p_j(x) = \left((2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma_j} \right)^{-1} \exp \left(-\frac{1}{2} (x - w_j)^T \Sigma_j^{-1} (x - w_j) \right) \quad (1)$$

where w_j - n -dimensional vector centroid of the j -th cluster, Σ_j - the correlation matrix of the j -th cluster of dimension $(n \times n)$:

$$\Sigma_j = \frac{1}{N} \sum_{k=1}^N (x(k) - w_j)(x(k) - w_j)^T. \quad (2)$$

Basing on (1), it is easy to write down the joint distribution density of observations of an initial dataset in the form

$$\begin{aligned} p(x) &= \sum_{k=1}^N p_j p_j(x) = \sum_{k=1}^N p_j \left((2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma_j} \right)^{-1} \exp \left(-\frac{1}{2} (x - w_j)^T \Sigma_j^{-1} (x - w_j) \right) = \\ &= \sum_{k=1}^N p_j \left((2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma_j} \right)^{-1} \exp \left(-\frac{1}{2} d_M^2(x, w_j) \right) \end{aligned} \quad (3)$$

where p_j - a priori probabilities that satisfy standard conditions

$$\sum_{j=1}^m p_j = 1. \quad (4)$$

It is easy to see that equation (4) coincides with the constraint on the sum of the memberships in the Fuzzy C-Means algorithm

$$\sum_{j=1}^m \mu_j = 1 \quad (5)$$

where $0 < \mu_j(k) < 1$ - membership level of k -th observation to j -th cluster.

Due to the constraints of (5) procedures of FCM-type are called fuzzy probabilistic algorithms [10].

Let's introduce Mahalanobis distance between centroids w_j and vectors-pattern $x(k)$ in the form of

$$d_m^2(x(k), w_j) = (x(k) - w_j)^T \Sigma_j^{-1} (x(k) - w_j). \quad (6)$$

It's obviously that using of this metrics instead of the traditional Euclidean one in the FCM algorithm allows to restore the classes of the hyperellipsoidal form with an arbitrary orientation of the axes in the initial space of features.

In the process of clustering using the EM-approach, the maximization of the function of log likelihood is implemented:

$$E(x(k), w_j, \Sigma_j, p_j) = \sum_{k=1}^N \log \left(\sum_{j=1}^m p_j p_j(x(k)) \right) \quad (7)$$

wherein, the final result can be written in the form [13]

$$\begin{cases} p_j(x(k)) = \exp\left(-\frac{1}{2}d_M^2(x(k), w_j)\right) / \sum_{i=1}^m \exp\left(-\frac{1}{2}d_M^2(x(k), w_i)\right), \\ w_j = \sum_{k=1}^N p_j(x(k))x(k) / \sum_{k=1}^N p_j(x(k)). \end{cases} \quad (8)$$

It should be noted that when $p_j = m^{-1}$ and the unity matrix Σ^{-1} , the EM algorithm coincides with the standard K-means clustering procedure.

As it is known [4,5] K-means clustering procedure is related with minimization of the goal function

$$E(x(k), w_j) = \sum_{k=1}^N \sum_{j=1}^m \mu_j(k) \|x(k) - w_j\|^2 = \sum_{k=1}^N \sum_{j=1}^m \mu_j(k) d_E^2(x(k), w_j) \quad (9)$$

where

$$\mu_j(k) = \begin{cases} 1, & \text{if } x(k) \text{ belongs to } j\text{-th cluster,} \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

In doing so, the use of Euclidean metrics leads to the fact that the emerging clusters have a spherical shape. A modification of the standard K-means is the procedure [6] of Mahalanobis K-means, associated with the goal function

$$\begin{aligned} E(x(k), w_j) &= \sum_{k=1}^N \sum_{j=1}^m \mu_j(k) (x(k) - w_j)^T \Sigma_j^{-1} (x(k) - w_j) = \\ &= \sum_{k=1}^N \sum_{j=1}^m \mu_j(k) d_M^2(x(k), w_j) \end{aligned} \quad (11)$$

the minimization of which leads to an assessment of the centroids' position of the form:

$$w_j = \sum_{k=1}^N \mu_j(k) x(k) / \sum_{k=1}^N \mu_j(k) = \frac{1}{N_j} \sum_{x(k) \in Cl_j} x(k) \quad (12)$$

where N_j – is the number of observations of the initial dataset associated with the j -th cluster.

Crisp goal functions (9), (11) are the partial case of fuzzy clustering criterion [18]

$$E(x(k), w_j, \mu_j) = \sum_{k=1}^N \sum_{j=1}^m \mu_j^\beta(k) d^2(x(k), w_j) \quad (13)$$

where a positive fuzzifier $\beta > 0$ describes blurring of the clusters, wherein most often (FCM) the value of this parameter is equal to two. At the same time, as distances $d^2(x(k), w_j)$ – Euclidean metrics are taken.

The solution of the minimization of the goal function problem (13) with the constraints (5) leads to fuzzy probabilistic clustering algorithm [3]:

$$\begin{cases} \mu_j(k) = d^{1-\beta}(x(k), w_j) / \sum_{l=1}^m d^{1-\beta}(x(k), w_l), \\ w_j = \sum_{k=1}^N \mu_j^\beta(x(k), w_j) / \sum_{k=1}^N \mu_j^\beta(k) \end{cases} \quad (14)$$

which when $\beta = 2$ turns into a standard FCM procedure:

$$\begin{cases} \mu_j(k) = \frac{d_E^{-2}(x(k), w_j)}{\sum_{l=1}^m d_E^{-2}(x(k), w_l)} = \frac{\|x(k) - w_j\|^2}{\sum_{l=1}^m \|x(k) - w_l\|^2}, \\ w_j = \sum_{k=1}^N \mu_j^2(k) x(k) / \sum_{k=1}^N \mu_j^2(k). \end{cases} \quad (15)$$

The first relation (14) can be rewritten in the form

$$\begin{cases} \mu_j(k) = \left(1 + \frac{d^{\frac{2}{\beta-1}}(x(k), w_j)}{\gamma_j} \right)^{-1}, \\ \gamma_j = \left(\sum_{\substack{l=1, \\ l \neq j}}^m d^{\frac{2}{\beta-1}}(x(k), w_l) \right)^{-1}, \end{cases} \quad (16)$$

the corresponding description of generalized Gaussian [19], which when $\beta = 2$ turns into Cauchy probabilities density function, which leads to the expression

$$\mu_j(k) = \frac{1}{1 + \frac{d^2(x(k), w_j)}{\gamma_j}}, \quad \gamma_j = \left(\sum_{\substack{l=1, \\ l \neq j}}^m d^2(x(k), w_l) \right)^{-1}. \quad (17)$$

It is interesting to note here that if the EM algorithm is based on Gaussian distribution, then fuzzy procedures are connected with the distribution of Cauchy.

It is also interesting to note that the popular clustering algorithm of Gath-Geva [20], which minimizes the goal function (13), occupies an intersection between the EM and FCM approaches since the estimate uses as the distance

$$\begin{aligned} d_{GG}^2(x(k), w_j) &= q_j \left(\sqrt{\det \Sigma_j} \right)^{-1} \exp \left(-\frac{1}{2} (x - w_j)^T \Sigma_j^{-1} (x - w_j) \right) = \\ &= q_j \left(\sqrt{\det \Sigma_j} \right)^{-1} \exp \left(-\frac{1}{2} d_M^2(x, w_j) \right) \end{aligned} \quad (18)$$

where

$$q_j = \sum_{k=1}^N \mu_j^\beta(k) / \sum_{k=1}^N \sum_{l=1}^m \mu_l^\beta(k). \quad (19)$$

As a result of minimization (18) with the constraints (5), (19), we came to the algorithm

$$\begin{cases} \mu_j(k) = d_{GG}^{\frac{2}{1-\beta}}(x(k), w_j) / \sum_{l=1}^m d_{GG}^{\frac{2}{1-\beta}}(x(k), w_l), \\ w_j = \sum_{k=1}^N \mu_j^\beta(k) x(k) / \sum_{k=1}^N \mu_j^\beta(k), \\ \Sigma_j = \sum_{k=1}^N \mu_j^\beta(k) (x(k) - w_j)(x(k) - w_j)^T / \sum_{k=1}^N \mu_j^\beta(k) \end{cases} \quad (20)$$

which when $\beta = 2$ takes the form

$$\begin{cases} \mu_j(k) = d_{GG}^{=2}(x(k), w_j) / \sum_{l=1}^m d_{GG}^{=2}(x(k), w_l), \\ w_j = \sum_{k=1}^N \mu_j^2(k) x(k) / \sum_{k=1}^N \mu_j^2(k), \\ \Sigma_j = \sum_{k=1}^N \mu_j^2(k) (x(k) - w_j)(x(k) - w_j)^T / \sum_{k=1}^N \mu_j^2(k), \end{cases} \quad (21)$$

which is the FCM modification for the clusters-hyperellipsoids case.

3 Online Fuzzy Probabilistic Clustering In The Case Of Hyperellipsoidal Classes

The procedures discussed above assume that the initial dataset is specified in the batch of data form, which is processed several times in multi-epoch learning mode. It is clear that if the information is fed for processing in the form of a data stream $x(1), x(2), \dots, x(k), x(k+1), \dots$, (here k – index of the current discrete time), the clustering methods discussed above are ineffective.

As is known, self-organizing T. Kohonen's map solves the clustering problem in sequential mode by minimizing the goal function (9), i.e., in fact, implements the K-means method for data flow. Popular WTA self-learning rule that looks like [21]

$$w_j(k+1) = \begin{cases} w_j(k) + \eta(k+1)(x(k+1) - w_j(k)), \\ \text{if } w_j(k) - \text{"winner"}, \\ w_j(k) - \text{otherwise} \end{cases} \quad (22)$$

(here $0 < \eta(k+1) < 1$ – the learning rate parameter is chosen according to stochastic approximation conditions) actually computes the usual arithmetic mean in recurrent form.

The self-learning rule (22) is closely related to the EM algorithm since the E-step of expectation implements the process of Kohonen's competition, and the M-step of maximization implements the synaptic adaptation process. With this distance:

$$d_E^2(x(k+1), w_j(k)) = \|x(k+1) - w_j(k)\|^2 \quad (23)$$

minimizing by the gradient procedure (24), in fact, coinciding with (22)

$$w_j(k+1) = \begin{cases} w_j(k) - \eta(k+1) \nabla_{c_j} d_E^2(x(k+1), w_j(k)), \\ \text{if } w_j(k) - \text{"winner"}, \\ w_j(k) - \text{otherwise}. \end{cases} \quad (24)$$

Similarly (24) Mahalanobis metrics can be minimized (6) using a recurrent procedure [22]:

$$w_j(k+1) = \begin{cases} w_j(k) - \eta(k+1) \nabla_{c_j} d_M^2(x(k+1), w_j(k)), \\ \text{if } w_j(k) - \text{"winner"}, \\ w_j(k) - \text{otherwise} \end{cases} \quad (25)$$

or, which is the same:

$$w_j(k+1) = \begin{cases} w_j(k) + \eta(k+1) \Sigma_j^{-1}(k) (x(k+1) - w_j(k)), & \text{if } w_j(k) \text{ - "winner",} \\ \Sigma_j(k) = \frac{1}{k_j} \sum_{\tau=1}^k (x(\tau) - w_j(k)) (x(\tau) - w_j(k))^T = \\ = \Sigma_j(k-1) + \frac{1}{k_j} \left((x(k) - w_j(k)) (x(k) - w_j(k))^T - \right. \\ \left. - \Sigma_j(k-1) \right), w_j(k) \text{ - otherwise} \end{cases} \quad (26)$$

where k_j – the number of “wins” of the j -th neuron of Kohonen’s map.

For fuzzy situations where the clusters being formed are mutually overlapped, in addition to the procedure (26), an assessment of the membership level can be declared similar to (15):

$$\begin{aligned} \mu_j(k) &= d_M^{-2}(x(k), w_j(k)) / \sum_{l=1}^m d_M^{-2}(x(k), w_l(k)) = \\ &= \frac{\left((x(k) - w_j(k))^T \Sigma_j^{-1}(k) (x(k) - w_j(k)) \right)^{-1}}{\sum_{l=1}^m \left((x(k) - w_l(k))^T \Sigma_l^{-1}(k) (x(k) - w_l(k)) \right)^{-1}}. \end{aligned} \quad (27)$$

Next, solving the nonlinear programming problem (goal function (13) with constraints (5)), using the Arrow-Hurwitz-Uzawa algorithm, it is easy to write down relations:

$$\begin{cases} w_j(k+1) = w_j(k) + \eta(k+1) \mu_j^\beta(k+1) (x(k+1) - w_j(k)), \\ \mu_j(k+1) = d_E^{1-\beta}(x(k+1), w_j(k)) / \sum_{l=1}^m d_E^{1-\beta}(x(k+1), w_l(k)), \end{cases} \quad (28)$$

which when $\beta = 2$ taken the simple form [16]:

$$\begin{cases} w_j(k+1) = w_j(k) + \eta(k+1) \mu_j^2(k+1) (x(k+1) - w_j(k)), \\ \mu_j(k+1) = \|x(k+1) - w_j(k)\|^{-2} / \sum_{l=1}^m \|x(k+1) - w_l(k)\|^{-2}. \end{cases} \quad (29)$$

Here, the multipliers $\mu_j^\beta(k+1)$, $\mu_j^2(k+1)$ in fact are neighbourhood functions in Kohonen’s self-learning rule “Winner Takes More”. Wherein, instead of the usual Gaussian, generalized Gaussian and Cauchian are used. It is interesting to note that the receptive fields parameters of these functions are automatically evaluated here.

It should be noted that the recurrent modification of the Gath-Geva algorithm [20] was introduced in [23], but this modification is not related to optimization procedures. Further modification can be written in the form of recurrence relations

$$\begin{cases} w_j(k+1) = w_j(k) + \eta(k+1)(x(k+1) - w_j(k)), \text{ if } w_j(k) - \text{"winner"}, \\ \Sigma_j(k+1) = (1 - \eta(k+1))\Sigma_j(k) + \eta(k)(x(k+1) - w_j(k))(x(k+1) - w_j(k))^T. \end{cases} \quad (30)$$

It can be noticed that relations (30) are the WTA self-learning rule of T.Kohonen and the procedure for correcting the correlation matrix. It is interesting to note that this matrix doesn't fluent the process of the centroids tuning.

A more effective is algorithm also proposed in [23], where a correction is made to the membership levels of the form:

$$U_j(k+1) = \sum_{\tau=1}^{k+1} \mu_j^\beta(\tau) = U_j(k) + \mu_j^\beta(k+1). \quad (31)$$

In this case, the algorithm has the form:

$$\begin{cases} w_j(k+1) = w_j(k) + \frac{\mu_j^\beta(k+1)}{U_j(k+1)}(x(k+1) - w_j(k)), \\ \Sigma_j(k+1) = U_j(k)/U_j(k+1) \left(\Sigma_j(k) + \frac{\mu_j^\beta(k+1)}{U_j(k+1)}(x(k+1) - w_j(k)) \cdot \right. \\ \left. \cdot (x(k+1) - w_j(k))^T \right), \\ \mu_j(k+1) = d_{GG}^{\frac{2}{1-\beta}}(x(k+1), w_j(k)) / \sum_{l=1}^m d_{GG}^{\frac{2}{1-\beta}}(x(k+1), w_j(k)). \end{cases} \quad (32)$$

Algorithm (32) is close to procedure (28) and coincides with it when $\eta(k+1) = U_j^{-1}(k+1)$, however, the metric $d_{GG}^2(x(k+1), w_j(k))$ used is significantly different from the previously used one $d_E^2(x(k+1), w_j(k))$. Once again, in this algorithm, the correlation matrix $\Sigma_j(k)$ does not affect the process of centroids tuning.

Combining procedure (25) using the gradient of Mahalanobis 'metrics and the standard Gath-Geva algorithm, we will get a relation describing the fuzzy clustering method with the hyperellipsoidal classes:

$$\begin{cases}
w_j(k+1) = w_j(k) + \frac{\mu_j^\beta(k+1)}{U_j(k+1)} \sum_j^{-1} (x(k+1) - w_j(k)), \\
\sum_j(k+1) = \frac{U_j(k)}{U_j(k+1)} \left(\sum_j(k) + \frac{\mu_j^\beta(k+1)}{U_j(k+1)} (x(k+1) - w_j(k)) \cdot \right. \\
\left. \cdot (x(k+1) - w_j(k))^T \right), \\
\mu_j(k+1) = d_{GG}^{\frac{2}{1-\beta}}(x(k+1), w_j(k)) \Big/ \sum_{l=1}^m d_{GG}^{\frac{2}{1-\beta}}(x(k+1), w_j(k)).
\end{cases} \quad (33)$$

Using the fuzzifier $\beta = 2$, we get adaptive recurrent procedure [24]:

$$\begin{cases}
U_j(k+1) = U_j(k) + \mu_j^2(k+1), \\
w_j(k+1) = w_j(k) + \frac{\mu_j^2(k+1)}{U_j(k+1)} \sum_j^{-1} (x(k+1) - w_j(k)), \\
\sum_j(k+1) = \frac{U_j(k)}{U_j(k+1)} \left(\sum_j(k) + \frac{\mu_j^2(k+1)}{U_j(k+1)} (x(k+1) - w_j(k)) \cdot \right. \\
\left. \cdot (x(k+1) - w_j(k))^T \right), \\
\mu_j(k+1) = d_{GG}^{-2}(x(k+1), w_j(k)) \Big/ \sum_{l=1}^m d_{GG}^{-2}(x(k+1), w_j(k)).
\end{cases} \quad (34)$$

4 Computer Experiments With Medical Data

As a demonstration of the developed method Online Fuzzy Probabilistic Clustering, two medical samples from the UCI repository were used. The first series of experiments was performed on a dataset of Breast Cancer Wisconsin, which related to the applied problem in the medical field. The data describe two types of tumors – malignant or benign. Clustering results are represented as a 3D model.

The structure of Breast Cancer Wisconsin dataset is complex and non-linear. The maximum classification accuracy reached 97.5 percent. Fig. 1, 2, and 3 show the visualization of clustering results in three different projections. For visualization of our dataset, that was previously compressed with Principal Component Analysis method (PCA-method) into three principal components.

The received results we compared with standard EM-algorithm and Fuzzy C-means method of J. Bezdek. The results of the experiment can be seen in Table 1. In the first row of the table, the outputs of EM-procedure are represented, in the second row – FCM one and in the third – the proposed approach. It can certainly be seen, that considered procedures surpass the quality of results both EM and FCM algorithms.

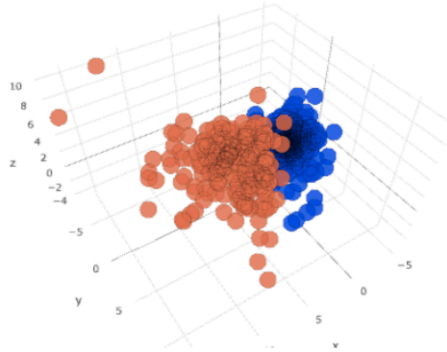


Fig. 1. The visualization of Breast Cancer side view.

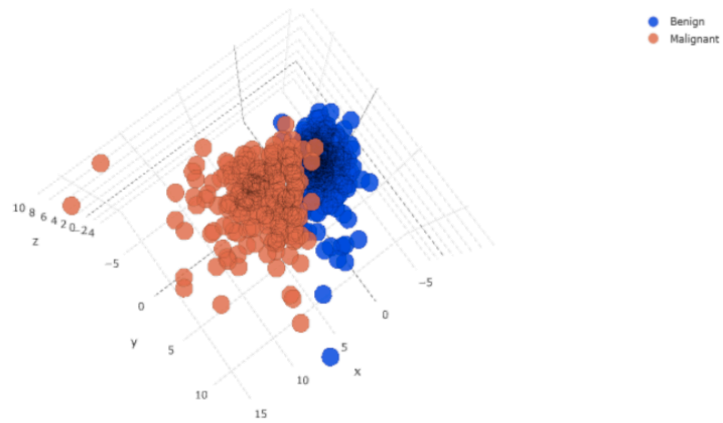


Fig. 2. The visualization of Breast Cancer top view.

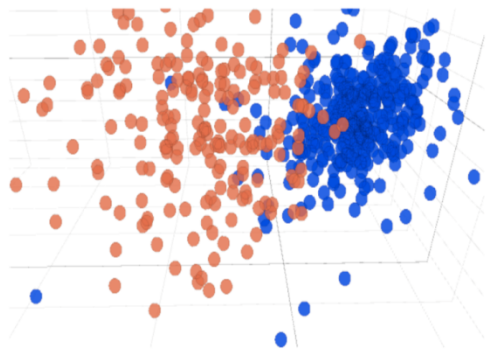


Fig. 3. The visualization of Breast Cancer front view.

Table 1. The results of the experiment.

The results of experiment	The results of experiment
	Breast Cancer Wisconsin
Expectation-Maximization algorithm	81.3 %
Fuzzy C-means method	79.4 %
Online Fuzzy Probabilistic Clustering system	97.5 %

For each cell nucleus, ten real signs are calculated that are columns in the sample: radius (average distance from the center to the points along the perimeter), texture (standard deviation of gray values); perimeter; square; smoothness; compactness; concavity (the severity of the concave parts of the contour); the number of firing points (the number of concave parts on the contour of the tumor); symmetry; fractal size.

The obtained accuracy ranged from 89% to 97.5%, which in general was from 569 examples correctly classified from 490 to 547 examples.

The second series of experiments was carried out on a sample of Dermatology which is the data obtained as a result of a study of the histopathological features of patients with dermatological diseases. The goal is to determine the type of dermatological disease.

The dataset has 33 attributes, of which: one target attribute, which shows the presence of a specific dermatological disease, there are 6 classes. Classification: 1st class (psoriasis) – 112 examples, 2nd class (seborrheic dermatitis) – 61 examples, 3rd grade (lichen) – 72 examples, 4th grade (pink lichen) – 49 examples, 5th grade (chronic dermatitis) – 52 examples, 6th grade (lichen planus) – 20 examples.

This sample is not quite popular for testing various algorithms for medical data analytics since it has several linearly inseparable clusters that are difficult to process with standard clustering algorithms.

The final results of the clustering of the developed method are presented in Fig. 4 and 5. As can be seen in Fig. 4 and 5 some examples of one cluster are very close to another cluster, this suggests that some clusters are not linearly separable.

A series of experiments indicate that the proposed Online Fuzzy Probabilistic Clustering method works quite quickly and shows the high quality of clustering of data arrays in conditions of non-linear data. And also the experiment proves that the proposed method can solve practical problems in the field of medical data processing.

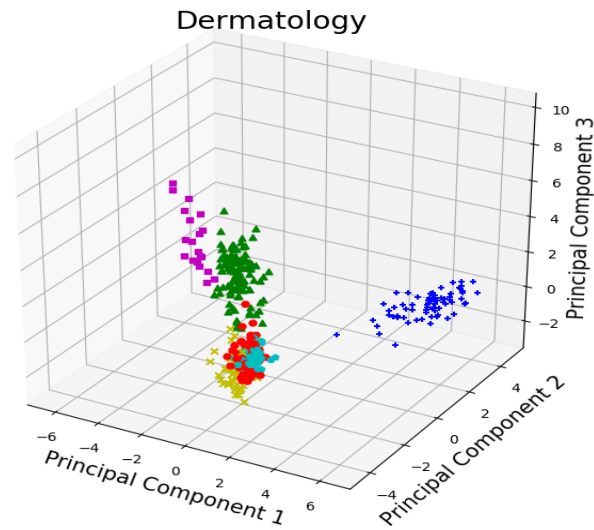


Fig. 4. The visualization of Dermatology dataset clusterisation.

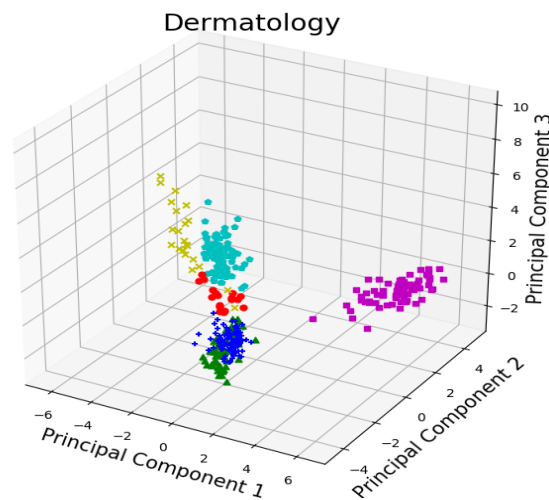


Fig. 5. The result of work of the system on the Dermatology dataset.

5 Conclusion

The problem of online fuzzy clustering of data that are fed to processing sequentially in the form of data stream was considered. A feature of approach under consideration is that the formed classes have the hyperellipsoidal shape with an arbitrary orientation of the axes in feature space. The proposed procedure uses Mahalanobis' metrics and also it is the generalization of a number of well-known clustering algorithms, has high

speed and simple, in computational implementation. The obtained results allow solving a number of problems arising in Data Mining and especially Medical Data Mining ones connected with mass examination of patients.

References

1. Aggarwal, I. C. C.: Data Mining. Cham: Springer, Int. Publ., Switzerland (2015).
2. Bramer, M.: Principles of Data Mining. Springer-Verlag London (2016).
3. Höppner, F., Klawonn, F., Kruse, R., Runkler, T.: Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. John Wiley & Sons. Chichester (1999).
4. Gan, G., Ma, Ch., Wu, J.: Data Clustering: Theory, Algorithms and Applications, Philadelphia: SIAM (2007).
5. Xu, R., Wunsch, D. C.: Clustering. IEEE Press Series on Computational Intelligence. Hoboken, NJ: John Wiley & Sons, Inc. (2009).
6. Aggarwal, C. C., Reddy, C. K.: Data Clustering. Algorithms and Application. Boca Raton: CRC Press (2014).
7. Bifet, A.: Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams, IOS Press (2010).
8. Kacprzyk, J., Pedrycz, W.: Springer Handbook of Computational Intelligence, Berlin Heidelberg: Springer, Verlag (2015).
9. Du, K.-L., Swamy, M. N. S.: Neural Networks and Statistical Learning. London: Springer-Verlag (2014).
10. Bezdek, J.-C.: Pattern Recognition with Fuzzy Objective Function Algorithms, N.Y.: Plenum Press (1981).
11. Bodyanskiy, Ye. V., Deineko, A. O., Kutsenko, Y. V.: On-line kernel clustering based on the general regression neural network and T. Kohonen's self-organizing map, Automatic Control and Computer Sciences 51(1), 55-62 (2017).
12. Bezdek, J. C., Keller, J., Krishnapuram, R., Pal, N.: Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. The Handbook of Fuzzy Sets. Kluwer, Dordrecht, Netherlands: Springer, vol. 4 (1999).
13. Dempster, A. P., Laird, N. M., Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm, J. of the Royal Statistical Society, Ser.B 39(1), pp. 1-38 (1977).
14. Hathaway, R.: Another interpretation of the EM algorithm for mixture distributions, J. of Statistics & Probability Letters, vol. 4, pp. 53-56 (1986).
15. Meng, X. L., Rubin, D. B.: Maximum likelihood estimation via the ECM algorithm: a general framework, Biometrika, vol. 80, pp. 267-278 (1993).
16. Bodyanskiy, Ye.: Computational intelligence techniques for data analysis, Lecture Notes in Informatics, Bonn: GI, pp. 15 – 36 (2005).
17. Gorshkov, Ye., Kolodyazhnyy, V., Bodyanskiy, Ye.: New recursive learning algorithms for fuzzy Kohonen clustering network, Proc. 17th Int. Workshop on Nonlinear Dynamics of Electronic Systems, Rapperswil, Switzerland, pp. 58-61 (2009).
18. Mumford, C., Jain, L.: Computational Intelligence, Collaboration, Fuzzy and Emergence, Berlin: Springer, Verlag (2009).
19. Osowski, S.: Sieci neuronowe do przetwarzania informacji, Warszawa: Oficjalna Wydawnicza Politechniki Warszawskiej (2006).

20. Gath, I., Geva, A. B.: Unsupervised optimal fuzzy clustering, *Pattern Analysis and Machine Intelligence* 2(7), pp. 773-787 (1989).
21. Kohonen, T.: *Self-Organizing Maps*. Berlin: Springer-Verlag (1995).
22. Bodyanskiy, Ye., Deineko, A., Kutsenko, Y., Zayika, O.: Data streams fast EM-fuzzy clustering based on Kohonen's self-learning, *The 1st IEEE International Conference on Data Stream Mining & Processing (DSMP 2016): Proc. of Int. Conf., Lviv*, pp. 309-313 (2016).
23. Geva, A.B.: Clustering as a basis for evolving neuro-fuzzy modeling, *Evolving Systems*, pp. 59-71 (2010).
24. Deineko, A., Zhernova, P., Gordon, B., Zayika, O., Pliss, I., Pabyrivska, N.: Data stream online clustering based on fuzzy expectation-maximization approach. *The 2nd IEEE International Conference on Data Stream Mining and Processing (DSMP 2018): Proc. of Int. Conf., Lviv*, pp 171-176 (2018).