Shafronenko, A. Yu. (2026). Online fuzzy clusterization of distorted Data Streams. *Actual Issues of Modern Science. European Scientific e-Journal, 42*, 390–413. Ostrava.

DOI: 10.47451/esej-tec-78

Article metadata are deposited, archived and/or discoverable through the journal's indexing, repository and metadata dissemination services listed on the Indexing and Abstracting page.

Alina Yu. Shafronenko, Doctor of Technical Sciences, Professor at the Department of Informatics, Kharkiv National University of Radio Electronics. Kharkiv, Ukraine.
E-mail: alina.shafronenko@nure.ua
ORCID 0000-0002-8040-0279

Article history:
Received: April 26, 2026
Revised: May 29, 2026
Accepted: June 15, 2026
Published: July 4, 2026

Online Fuzzy Clusterization of Distorted Data Streams

Abstract: Modern computational intelligence technologies are widely used to solve complex problems that often lack an exact analytical solution. A special place among them is held by artificial neural networks, which are widely used in signal processing, optimization, adaptive control, pattern recognition, object identification, time-series forecasting, and many other areas. On this basis, effective systems for computer vision, technical and medical diagnostics, aerospace object management, analysis of financial processes, energy systems, transport flows, and military equipment have been created. The list of applications for these technologies is continually expanding due to their high efficiency and ability to handle complex data. Currently, a significant amount of information has been accumulated about the activities of enterprises, institutions, organizations, and socio-economic systems. Such data reflect various aspects of their functioning and may contain valuable patterns and relationships. In particular, these may be indicators of inflation, population income, household expenditure structure, the cost of housing and communal services, the state of industrial and agricultural production, employment levels, and the timeliness of wage and pension payments. Analysis of such information is difficult due to its substantial volume, multidimensionality, and nonlinear cause-and-effect relationships. That is why there is a need to use modern methods of intelligent data analysis and machine learning that can automatically detect hidden patterns and generate forecasts. An important problem when analyzing empirical data is the presence of missing values, which may arise from errors in data collection, transmission, or storage. The presence of omissions reduces the quality of subsequent analysis and can undermine the accuracy of the constructed models. Despite significant research in missing data recovery, approaches based on computational intelligence methods, in particular artificial neural networks and fuzzy systems, are of particular interest. Such methods can account for complex nonlinear dependencies between features and achieve high accuracy in reconstructing lost observations, making them an effective tool for processing incomplete data sets. Modern methods for recovering missing data are effective mainly for static data sets whose structure does not change during processing. In many practical tasks, information must arrive in real time, requiring adaptive analysis methods. To solve such tasks, artificial neural networks, fuzzy systems, and neuro-fuzzy systems are widely used and, due to their ability to learn and approximate nonlinear dependencies, effectively work with streaming data. Of particular importance are the processes by which evolving systems form their structure and determine the parameters governing their functioning. However, the latest approaches focus on batch processing, which limits their applicability to online analysis of time series and data streams. Additional problems include gaps and anomalous outputs, which significantly reduce the efficiency of traditional processing and clustering algorithms. Among the methods of data mining, fuzzy clustering plays an important role, allowing you to capture the uncertainty and ambiguity of information. However, most known algorithms are ineffective in the simultaneous presence of missing values, outliers, and streaming data. Therefore, the development of new adaptive methods for clustering and data analysis that operate in real time is relevant. Significant contributions to the development of data recovery methods, fuzzy clustering, and evolving intelligent systems have been made by scientists such as T. Kohonen, J. Bezdek, R. Krishnapuram, L. Rutkowski, and T. Marwala, as well as others whose work laid the theoretical foundations of modern intelligent data analysis systems.

Keywords: Data Streams, distorted data, clustering, online mode, adaptive method, optimal expansion strategy.

Abbreviations:

ANNs are artificial neural networks,
PD is partial distance,
WTM is "winner-takes-more".

Introduction

In recent years, computational intelligence technologies have become one of the most effective tools for solving complex problems that often lack analytical solutions. Among these technologies, ANNs have gained particular importance due to their ability to model nonlinear relationships, learn from data, and adapt to changing environments. Today, ANNs are successfully applied in signal processing, optimization, adaptive control, pattern recognition, identification, time-series forecasting, and many other fields. Their practical applications include image processing and computer vision systems, aerospace control systems, medical and technical diagnostics, economic and financial analysis, energy management, transportation systems, military technologies, forensic science, and various signal analysis tasks. The range of applications continues to expand as new computational methods and intelligent systems are developed.

At present, vast amounts of information describing the activities of enterprises, healthcare institutions, financial organizations, and other complex systems are continuously accumulated. These data contain valuable information about the behavior and performance of the corresponding objects and processes. When properly analyzed, they can reveal objective patterns and hidden relationships that characterize the underlying phenomena. Examples include economic indicators such as inflation rates, household income levels, expenditure structures, utility costs, industrial and agricultural production statistics, employment rates, and social welfare indicators. Due to the large volume of available information and the presence of complex nonlinear dependencies, manual analysis of such data is often impractical or impossible.

As a result, there is a growing demand for advanced data analysis and forecasting techniques capable of automatically discovering hidden patterns in large datasets. Machine learning and computational intelligence methods provide powerful tools for addressing these challenges by extracting useful knowledge from empirical observations and supporting decision-making processes.

One of the major difficulties encountered in real-world data analysis is the presence of missing values. In many data mining and knowledge discovery tasks, empirical observations may be incomplete because of measurement errors, data transmission failures, equipment malfunctions, or human-related factors. Missing information can significantly reduce the accuracy and reliability of analytical models. Consequently, the problem of recovering missing observations has attracted considerable research attention. Among the existing approaches, methods based on computational intelligence, particularly artificial neural networks and fuzzy systems, have demonstrated high effectiveness in reconstructing incomplete data. Their ability to capture complex nonlinear relationships and uncertainty makes them especially suitable for restoring missing observations and improving the quality of subsequent data analysis.

Clustering is one of the most important tasks in data analysis and pattern recognition. Unlike classification, clustering is performed without prior knowledge of the underlying data structure, making it a challenging unsupervised learning problem. The main difficulty lies in the absence of additional information about the data, which requires the algorithm to independently identify hidden patterns and group similar observations into meaningful clusters.

An additional challenge arises in online learning environments, where data are received sequentially and must be processed in real time. In such scenarios, traditional clustering algorithms often require retraining whenever new observations become available. For large-scale datasets, repeated retraining significantly increases computational costs and processing time, limiting the applicability of these methods in real-world systems.

To ensure efficient and accurate clustering, it is necessary to develop methods that automatically select the most appropriate clustering solution based on predefined quality criteria. Therefore, the development of adaptive clustering approaches that can dynamically adjust their parameters and structure during operation is of considerable practical and scientific interest.

Despite the substantial number of studies devoted to clustering and data analysis, several important challenges remain unresolved. In particular, existing methods often demonstrate limited effectiveness when processing distorted or incomplete data containing missing values and outliers. Furthermore, there is a growing need for clustering techniques capable of handling fuzzy, overlapping clusters, in which an object may belong to multiple clusters simultaneously with varying degrees of membership. Equally important is the ability of such methods to operate in real time, adapting to continuously arriving data streams without requiring complete retraining. These requirements motivate the development of new adaptive fuzzy clustering algorithms designed for dynamic and uncertain data environments.

Methods

This study was designed as a theoretical and algorithmic investigation aimed at developing an adaptive method for online fuzzy clusterization of distorted data streams. The methodological focus was placed on data streams represented by object–property tables in which incoming observations may contain both missing values and anomalous outliers. Since conventional clustering algorithms are usually designed for complete static datasets processed in batch mode, the study develops a sequential framework capable of updating cluster prototypes, membership levels and recoverable data components in real time.

The methodological basis of the research combines fuzzy clustering, neuro-fuzzy modelling, self-organising Kohonen-type learning, recurrent optimisation procedures, partial-distance computation, optimal completion of missing observations and robust similarity-based clustering. The general logic of the proposed approach is to transform classical fuzzy clustering algorithms into recurrent online procedures and then adapt them to distorted streams by introducing mechanisms for handling incomplete and abnormal data.

The initial data model assumes that the processed dataset is represented by a sequence of n-dimensional feature vectors. Each observation is denoted as an input vector belonging to an n-dimensional feature space. Before clustering, all incoming variables are centred and standardised so that the values of each feature lie within the interval $[-1, 1]$. This normalisation is necessary to ensure comparability of feature scales, to prevent domination of variables with larger numerical ranges, and to make the subsequent use of distance- and similarity-based measures methodologically consistent.

The clustering task is formulated as an unsupervised partitioning problem. The purpose is to divide the incoming data stream into a predefined or adaptively updated number of clusters, while assigning each observation a degree of membership in every cluster. Unlike crisp clustering, fuzzy clustering allows an observation to belong to several clusters simultaneously with different membership levels. This property is important for real-world data streams, where cluster boundaries may be uncertain, overlapping or dynamically changing.

As a baseline, the study uses the general fuzzy c-means framework. In this framework, each cluster is represented by a prototype or centroid, while each observation is characterised by membership levels that satisfy probabilistic constraints. The classical objective function minimises the weighted sum of distances between observations and cluster prototypes. Membership levels are controlled by a fuzzifier parameter, which determines the degree of blurring of boundaries between clusters. When the fuzzifier is set to the standard value and the Euclidean metric is used, the procedure corresponds to Bezdek's fuzzy c-means algorithm.

To make fuzzy clustering suitable for online processing, the batch optimisation scheme is rewritten in recurrent form. In the recurrent version, observations are processed sequentially as they arrive. At each time step, the algorithm calculates current membership levels for the new observation and updates the corresponding cluster prototypes using a learning-rate parameter. This procedure makes it possible to avoid full retraining of the model after each new observation and therefore substantially reduces computational costs in streaming environments. If the data flow is high-frequency, the method performs a single recurrent update per incoming observation; if the flow is lower-frequency, several accelerated internal iterations can be carried out between consecutive real-time observations.

The recurrent fuzzy clustering procedure is interpreted as a neuro-fuzzy extension of Kohonen-type self-organisation. In this interpretation, cluster prototypes correspond to adaptive weight vectors, and the update mechanism follows a winner-takes-more learning principle. Unlike the winner-takes-all rule, in which only one prototype is substantially updated, the winner-takes-more principle allows several prototypes to be updated according to their fuzzy membership values. This makes the procedure more appropriate for overlapping clusters and uncertain data distributions.

A separate methodological stage is devoted to the processing of data with missing values. To describe incomplete observations, the input data array is divided into available and unavailable components. For each observation, a binary indicator is introduced to distinguish observed feature values from missing ones. Instead of the standard Euclidean metric, the method uses a partial-distance strategy. The partial distance is calculated only over the available components of the observation and is normalised with respect to the number of observed features. This enables the algorithm to estimate the distance between an incomplete observation and a cluster prototype without preliminary imputation.

On the basis of the partial-distance strategy, the standard fuzzy c-means objective function is modified. The modified objective function minimises distances computed only over the available feature components. The corresponding optimisation procedure yields updated membership levels and cluster prototypes that are calculated with regard to the pattern of missingness in each observation. This approach allows incomplete data streams to be processed directly in online mode and avoids the systematic bias that may arise from deleting incomplete observations or replacing missing values by simple statistical estimates.

The partial-distance fuzzy clustering algorithm is then reformulated as a recurrent learning procedure. In this recurrent form, each cluster prototype is updated only through the available components of the incoming observation, while the missing components do not directly contribute to the prototype correction. The resulting algorithm can be implemented within the architecture of a neuro-fuzzy Kohonen network. In this setting, the partial-distance computation defines the current membership levels, while the winner-takes-more update rule provides adaptive correction of the cluster prototypes in online mode.

Since the partial-distance strategy may become insufficient when the number of missing values is high, the study additionally introduces an optimal completion strategy. Under this strategy, missing feature values are treated as additional unknown variables to be estimated simultaneously with clustering. The algorithm alternates between three operations: calculation of fuzzy membership levels, recalculation of cluster prototypes and estimation of missing observations. Missing values are estimated by minimising the clustering objective function with respect to the unknown components of the data vector. Thus, the method performs clustering and data recovery as parts of a single optimisation process.

To organise the optimal completion strategy in online mode, two-time scales are introduced. The first scale is real time, corresponding to the sequential arrival of observations in the data stream. The second scale is accelerated computational time, during which internal iterations are performed between two consecutive real-time observations. Within accelerated time, the algorithm refines membership levels and estimates missing values; within real time, it updates cluster prototypes using the recurrent winner-takes-more learning rule. This two-scale organisation makes it possible to combine real-time operation with a more accurate recovery of missing observations.

The study also considers the limitations of probabilistic fuzzy clustering. In probabilistic clustering, the membership levels of each observation across all clusters must sum to unity. Although this condition is mathematically convenient, it may become restrictive in the presence of outliers and atypical observations. To address this limitation, the methodology includes possibilistic fuzzy clustering, in which membership levels are not constrained by the requirement of summing to one. In the possibilistic formulation, an additional scale parameter is introduced for each cluster, defining the distance at which the membership level equals 0.5. This makes the algorithm less sensitive to the forced redistribution of membership among clusters and more suitable for distorted data.

The possibilistic clustering objective function is minimised with respect to membership levels, cluster prototypes and cluster-specific scale parameters. The resulting expressions are also converted into recurrent online form. As in the probabilistic case, the procedure can be interpreted as a self-learning Kohonen-type rule with Cauchy functions acting as neighbourhood functions. This interpretation provides a unified neuro-fuzzy framework for both probabilistic and possibilistic clustering of streaming data.

To improve robustness in the presence of anomalous outliers, the study replaces purely distance-based criteria with similarity-based measures. A similarity measure is introduced as a function satisfying weaker conditions than a metric, since it does not require the triangle inequality. The proposed similarity measure is based on a Cauchy-type function with a width parameter controlling the influence of distant observations. By selecting this parameter appropriately, observations located at the edges of the normalised interval can be down-weighted, thereby reducing the effect of outliers on prototype updates. This robustification cannot be achieved by the Euclidean metric alone, since large distances generated by outliers may strongly distort the clustering process.

The similarity-based objective function is constructed by replacing distance minimisation with similarity maximisation. Under probabilistic constraints, the corresponding Lagrange optimisation problem is formulated and solved to obtain updated membership levels and prototype correction rules. The resulting robust fuzzy clustering algorithm is then adapted to incomplete data by using a partial similarity measure based on partial distance. This allows the method to handle missing values and outliers simultaneously.

The final methodological framework consists of a group of adaptive online fuzzy clustering algorithms. The first group includes probabilistic fuzzy clustering procedures for complete data streams. The second group extends these procedures to incomplete observations using the partial-distance strategy. The third group applies the optimal completion strategy to estimate missing values during clustering. The fourth group introduces possibilistic clustering to reduce the limitations of probabilistic membership constraints. The fifth group incorporates robust similarity-based measures to suppress the influence of outliers. Together, these procedures form an adaptive neuro-fuzzy approach for online clusterization of distorted data streams.

Algorithmic implementation follows a sequential processing logic. At each real-time step, a new observation is received and normalised if necessary. The algorithm identifies available and missing components, calculates partial distances or partial similarities, determines fuzzy or possibilistic membership levels, updates missing-value estimates where required, and corrects cluster prototypes according to the winner-takes-more rule. When accelerated internal iterations are enabled, membership levels and missing-value estimates are refined before prototype updating. The process continues for the full data stream without requiring complete retraining on the accumulated dataset.

The stopping and control conditions depend on the difference between successive prototype updates, the threshold accuracy parameter and the maximum number of accelerated internal iterations. In batch-derived procedures, the calculation continues until the norm of the difference between successive cluster prototypes becomes smaller than the predefined threshold. In online processing, this condition is adapted to sequential operation by controlling the magnitude of prototype correction and by limiting the number of accelerated iterations between consecutive observations.

The proposed methodology was evaluated at the theoretical and algorithmic level through consistency with known fuzzy clustering principles, compatibility with Kohonen-type self-organisation and the ability to process incomplete and distorted observations without batch recomputation. The key methodological criteria were online applicability, capacity to process missing values, robustness to outliers, reduction of memory requirements due to the absence of summation over the full historical sample in recurrent updates, and suitability for numerical implementation by adaptive linear associators.

The methodological contribution of the study lies in combining several previously separate approaches into one coherent framework: fuzzy c-means clustering, partial-distance processing, optimal estimation of missing values, possibilistic membership modelling, similarity-based robustification and neuro-fuzzy Kohonen learning. This combination enables real-time clusterization of data streams in situations where observations are incomplete, noisy and affected by abnormal outputs. The proposed methodology is therefore suitable for applications in which data are continuously received and must be processed under uncertainty, distortion and changing statistical conditions.

Literature Review

The problem of data sets described by vector-image clustering often arises in many applications associated with Data Mining (*Höppner et al., 1999*, *Bezdek, 2013*), where a processed vector-image with varying levels of probability, possibility, or membership can belong to more than one class.

However, there are situations when the data sets contain missing values. In this situation, it is more effective to use the mathematical apparatus of Computational and, first of all, artificial neural networks (*Marwala, 2009*), which solve the task of restoring the lost observations and modifications of the popular method of fuzzy c-means (*Hathaway & Bezdek, 2001*), which solves the problem of clustering without recovery of data.

Existing approaches for data processing with missing values (*Bodyanskiy et al., 2012*) are efficient when the mass of the original observations is provided in batch form and remains constant during processing. At the same time, there is a wide class of problems in which the data arriving for processing takes the form of a sequence, fed in real time as it occurs, as in the training of Kohonen self-organizing maps (*Kohonen & Maps, 1995*) or their modifications (*Gorshkov et al., 2009*). In this regard, we have introduced (*Bodyanskiy et al., 2012*) the adaptive neuro-fuzzy Kohonen network to address the problem of clustering data with gaps, based on the partial-distance strategy (PDS FCM). However, in situations where the number of such missing values is too big, the strategy of partial distances may not be effective, and therefore, it may be necessary, along with the solution of fuzzy clustering, to simultaneously estimate the missing observations. In this situation, a more efficient approach is based on the optimal expansion strategy (OCS FCM) (*Hathaway & Bezdek, 2001*).

Results

Baseline information for solving clustering tasks in batch mode is a sample of observations consisting of n-dimensional feature vectors $X = \{x_1, x_2, ..., x_N\} \subset R^n, x_k \in X, k = 1, 2, ..., N$. The result of clustering is the partition of original data set into m classes $(1 \le m \le N)$ with some levels of membership $U_q(k)$ of k-th feature vector to the q-th cluster $(1 \le q \le m)$. Incoming data are centered and standardized for all features, so that all observations lie within the hypercube $[-1,1]^n$. Therefore, the data for clustering form array $\tilde{X} = \{\tilde{x}_1, ..., \tilde{x}_k, ..., \tilde{x}_N\} \subset R^n$, $\tilde{x}_k = (\tilde{x}_{k1}, ..., \tilde{x}_{ki}, ..., \tilde{x}_{kn})^T$, $-1 \le \tilde{x}_{ki} \le 1$, $1 < m < N$, $1 \le q \le m$, $1 \le i \le n$, $1 \le k \le N$.

Introducing the objective function of clustering

$$E(U_q(k), w_q) = \sum_{k=1}^{N} \sum_{q=1}^{m} U_q^{\beta}(k) D^2(\tilde{x}_k, w_q)$$

with constraints $\sum_{q=1}^{m} U_q(k) = 1,\ 0 < \sum_{k=1}^{N} U_q(k) < N$ and solving the nonlinear programming problem, we get the probabilistic fuzzy clustering algorithm (*Höppner et al., 1999*, *Bezdek, 2013*)

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{(D^2(\tilde{x}_k, w_q^{(\tau)}))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m} (D^2(\tilde{x}_k, w_l^{(\tau)}))^{\frac{1}{1-\beta}}}, \\ w_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N} (U_q^{(\tau+1)})^{\beta} \tilde{x}_k}{\sum_{k=1}^{N} (U_q^{(\tau+1)}(k))^{\beta}}, \end{cases} \quad (1)$$

where

w_q is prototype (centroid) of q-th cluster,

$\beta > 1$ is parameter that is called fuzzyfier and defines "blurring" the boundaries between classes,

$D^2(\tilde{x}_k, w_q)$ is the distance between $\tilde{x}_k$ and w_q in adopted metric,

$\tau = 0, 1, 2, ...$ is index of epoch information processing which is organized as a sequence of $w_q^{(0)} \to U_q^{(1)} \to w_q^{(1)} \to U_q^{(2)} \to ....$

The calculation process continues until satisfy the condition

$$\left\| w_q^{(\tau+1)} - w_q^{(\tau)} \right\| \le \varepsilon \ \forall\ 1 \le q \le m$$

where

ε is defines threshold of accuracy.

Choosing $\beta = 2$ and taking the Euclidean distance, we get a popular algorithm of Bezdek's fuzzy c-means (FCM)

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{\left\| \tilde{x}_k - w_q^{(\tau)} \right\|^{-2}}{\sum_{l=1}^{m} \left\| \tilde{x}_k - w_l^{(\tau)} \right\|^{-2}}, \\ w_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N} (U_q^{(\tau+1)}(k))^2 \tilde{x}_k}{\sum_{k=1}^{N} (U_q^{(\tau+1)}(k))^2}. \end{cases}$$

The process of fuzzy clustering can be organized in on-line mode as sequentially processing. At this situation batch algorithm (1) can be rewritten in recurrent form (*Bodyanskiy et al., 2005*)

$$\begin{cases} U_q(k+1) = \dfrac{(D^2(\tilde{x}_{k+1}, w_q^{(k)}))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m} (D^2(\tilde{x}_{k+1}, w_l(k)))^{\frac{1}{1-\beta}}}, \\ w_q(k+1) = w_q(k) + \eta(k+1) U_q^{\beta}(k+1)(\tilde{x}_{k+1} - w_q(k)) \end{cases} \tag{2}$$

(here $\eta(k+1)$ is learning rate parameter), which is a generalization of the clustering gradient procedure of Park-Dagher and the learning algorithm of Chung-Lee (*Chung & Lee, 1994*). If the data are fed to the processing with high-frequency, recalculation of epochs is not made, if this frequency is low, between the instants k and $k+1$ it is possible to organize several epochs in an accelerated time.

It should be noted that the first expression in (2) can be rewritten in the form

$$U_q(k+1) = \frac{(D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m} (D^2(\tilde{x}_k, w_l(k)))^{\frac{1}{1-\beta}}} = \frac{(D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{1-\beta}}}{(D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{1-\beta}} + \sum_{\substack{l=1 \\ l \neq q}}^{m} (D^2(\tilde{x}_k, w_l(k)))^{\frac{1}{1-\beta}}} =$$

$$= \frac{1}{1 + (D^2(\tilde{x}_k, w_q(k)))^{\frac{1}{\beta-1}} \sum_{\substack{l=1 \\ l \neq q}}^{m} (D^2(\tilde{x}_k, w_l(k)))^{\frac{1}{1-\beta}}}$$

for the Euclidean metric and $\beta = 2$ taking the form of the Cauchy density distributionfunction with a parameter of width σ^2:

$$U_q(k+1) = \frac{1}{1 + \dfrac{\left\| \tilde{x}_k - w_q(k) \right\|^2}{\sigma^2}},$$

$$\sigma^2 = \Big(\sum_{\substack{l=1 \\ l \neq q}}^{m} \left\| \tilde{x}_k - w_l(k) \right\|^{-2}\Big)^{-1}.$$

This fact allows us to rewrite the second expression in (2) with $\beta = 2$ in the form

$$w_q(k+1) = w_q(k) + \eta(k+1) U_q^2(k+1)(\tilde{x}_{k+1} - w_q(k)) = w_q(k) + \eta(k+1)\varphi_q(k+1)(\tilde{x}_{k+1} - w_q(k))$$

where $U_q^2(k+1)=\varphi_q(k+1)$ is the bell-shaped neighborhood function of neuro-fuzzy Kohonen network (*Gorshkov et al., 2009*) designed for solving the fuzzy clustering task (*Shafronenko, et al., 2018*) using the principle WTM.

In the situation if the data in the array $\tilde{X}$ contain gaps, the approach discussed above should be modified accordingly. For example, in (*Hathaway & Bezdek, 2001*) it was proposed the modification of the FCM-procedure based on partial distance strategy (PDS FCM). Thus introducing, additional arrays:

$X_F=\{\tilde{x}_k\in\tilde{X}\,|\,\tilde{x}_k\text{ - } vector\ containing\ all\ components\}$;

$X_P=\{\tilde{x}_{ki},1\le i\le n,1\le k\le N\,|\ values\ \tilde{x}_k,\ available\ in\ \tilde{X}\}$;

$X_G=\{\tilde{x}_{ki}=?,1\le i\le n,1\le k\le N\,|\ values\ \tilde{x}_k, absent\ in\ \tilde{X}\}$

and using instead of the traditional Euclidean metric PD:

$$D_P^2(\tilde{x}_k,w_q)=\frac{n}{\delta_{k\Sigma}}\sum_{i=1}^{n}(\tilde{x}_{ki}-w_{qi})^2\delta_{ki},$$

the objective function of clustering

$$E(U_q(k),w_q)=\sum_{k=1}^{N}\sum_{q=1}^{m}U_q^{\beta}(k)\frac{n}{\delta_{k\Sigma}}\sum_{i=1}^{n}(\tilde{x}_{ki}-w_{qi})^2\delta_{ki}$$

(here $\delta_{ki}=\begin{cases}0\,|\,\tilde{x}_{ki}\in X_G,\\ 1\,|\,\tilde{x}_{ki}\in X_F,\end{cases}$ $\delta_{k\Sigma}=\sum_{i=1}^{n}\delta_{ki}$)

and solving nonlinear programming problem, we obtain the algorithm

$$\begin{cases}U_q^{(\tau+1)}=\dfrac{(D_P^2(\tilde{x}_k,w_q^{(\tau)}))^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}(D_P^2(\tilde{x}_k,w_q^{(\tau)}))^{\frac{1}{1-\beta}}},\\ w_{qi}^{(\tau+1)}=\dfrac{\sum\limits_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}\delta_{ki}\tilde{x}_{ki}}{\sum\limits_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}\delta_{ki}}\end{cases}\tag{3}$$

which is a generalization of the standard FCM-procedure (1).

Algorithm (3) can be rewritten in recurrent form

$$\begin{cases}U_q(k+1)=\dfrac{(D_P^2(\tilde{x}_{k+1},w_q(k)))^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}(D_P^2(\tilde{x}_{k+1},w_q(k)))^{\frac{1}{1-\beta}}},\\ w_{qi}(k+1)=w_{qi}(k)+\eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1,i}-w_{qi}(k))\delta_{ki}\end{cases}\tag{4}$$

with the second relation (4) that can be represented as learning algorithm for neuro-fuzzy Kohonen network:

$$w_q(k+1)=w_q(k)+\eta(k+1)\phi_q(k+1)(\tilde{x}_{k+1}-w_q(k))\square\ \delta_k\tag{5}$$

where $\phi_q(k+1) = U_q^\beta(k+1)$ - bell-shaped neighborhood function, $\delta_k = (\delta_{k1},...,\delta_{kn})^T$, □ -symbol of direct product.

Thus, using a standard Kohonen network architecture and algorithm of its tuning (5) in online mode, it is possible to solve the problem of fuzzy clustering data with gaps.

The main disadvantage of probabilistic algorithms is connected with the constraints on membership levels whose sum has to be equal unity. This reason has led to the creation of possibilistic fuzzy clustering algorithms (*Bezdek et al., 1999*).

In possibilistic clustering algorithms the objective function has the form

$$E(U_q(k), w_q, \mu_q) = \sum_{k=1}^{N}\sum_{q=1}^{m} U_q^\beta(k) D^2(\tilde{x}_k, w_q) + \sum_{q=1}^{m}\mu_q \sum_{k=1}^{N}(1-U_q(k))^\beta \tag{6}$$

where the scalar parameter $\mu \geq 0$ determines the distance at which level of membership equals to 0.5, i.e. if $D^2(\tilde{x}_k, w_q) = \mu_q$, then $w_q(k) = 0.5$.

Minimizing (6) relatively $U_q(k)$, w_q and μ_q we get the solution

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{1}{1+\left(\dfrac{D^2(\tilde{x}_k, w_q^{(\tau)})}{\mu_q^{(\tau)}}\right)^{\frac{1}{\beta-1}}}, \\ w_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^\beta \tilde{x}_k}{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^\beta}, \\ \mu_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^\beta D^2(\tilde{x}_k, w_q^{(\tau+1)})}{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^\beta} \end{cases} \tag{7}$$

which with $\beta = 2$ and Euclidean metric has the form

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{1}{1+\dfrac{\left\|\tilde{x}_k - w_q^{(\tau)}\right\|^2}{\mu_q^{(\tau)}}}, \\ w_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^2 \tilde{x}_k}{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^2}, \\ \mu_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^2 \left\|\tilde{x}_k - w_q^{(\tau+1)}\right\|^2}{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^2}. \end{cases} \tag{8}$$

Information processing in the on-line mode (7), (8) can be written as (Bodyanskiy et al., 2012; Shafronenko et al., 2018)

$$\begin{cases} U_q(k+1) = \dfrac{1}{1+(\dfrac{D^2(\tilde{x}_{k+1}, w_q(k))}{\mu_q(k)})^{\frac{1}{\beta-1}}}, \\ w_q(k+1) = w_q(k) + \eta(k+1)U_q^\beta(k+1)(\tilde{x}_{k+1} - w_q(k)), \\ \mu_q(k+1) = \dfrac{\sum_{p=1}^{k+1} U_q^\beta(p) D^2(\tilde{x}_p, w_q(k+1))}{\sum_{p=1}^{k+1} U_q^\beta(p)} \end{cases} \tag{9}$$

and

$$\begin{cases} U_q(k+1) = \dfrac{1}{1+\dfrac{\left\|\tilde{x}_k - w_q(k)\right\|^2}{\mu_q(k)}}, \\ w_q(k+1) = w_q(k) + \eta(k+1)U_q^2(k+1)(\tilde{x}_{k+1} - w_q(k)), \\ \mu_q(k+1) = \dfrac{\sum_{p=1}^{k+1} U_q^2(p) \left\|\tilde{x}_p - w_q(k+1)\right\|^2}{\sum_{p=1}^{k} U_q^2(p)}. \end{cases} \tag{10}$$

It's easily to see that relations (9), (10) are the Kohonen's self-learning WTM-rule with Cauchy functions as a neighborhood ones.

Adopting instead of Euclidean metric partial distance (PD), we can write the goal function of the type (6) as

$$E(U_q(k), w_q, \mu_q) = \sum_{k=1}^{N}\sum_{q=1}^{m} U_q^\beta(k) \frac{n}{\delta_{k\Sigma}} \sum_{i=1}^{n} (\tilde{x}_{ki} - w_{qi})^2 \delta_{ki} + \sum_{q=1}^{m} \mu_q \sum_{k=1}^{N} (1 - U_q(k))^\beta$$

and then solving the equations system

$$\begin{cases} \dfrac{\partial E(U_q(k), w_q, \mu_q)}{\partial U_q(k)} = 0, \\ \dfrac{\partial E(U_q(k), w_q, \mu_q)}{\partial \mu_q} = 0, \\ \nabla_{w_q} E(U_q(k), w_q, \mu_q) = \vec{0}, \end{cases}$$

get the procedure of type (9), which can be rewritten in the recurrent form

$$\begin{cases} U_q(k) = \dfrac{1}{1 + (\dfrac{D_P^2(\tilde{x}_{k+1}, w_q(k))}{\mu_q(k)})^{\frac{1}{\beta-1}}}, \\ w_{qi}(k+1) = w_{qi}(k) + \eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1,i} - w_{qi}(k))\delta_{ki}, \\ \mu_q(k+1) = \dfrac{\sum_{p=1}^{k+1} U_q^{\beta}(p) D_P^2(\tilde{x}_p, w_q(k+1))}{\sum_{p=1}^{k+1} U_q^{\beta}(p)}. \end{cases}$$

The second relation can be rewritten too as

$$w_q(k+1) = w_q(k) + \eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1} - w_q(k)) \odot \delta_k$$

coinciding with the learning procedure (7).

Thus, the process of fuzzy possibilistic clustering of data with gaps can also be implemented using a neuro-fuzzy Kohonen network.

An optimal expansion strategy is to treat the elements of the sub-array as additional variables, which are estimated by minimizing the goal function E. Thus, in parallel with clustering (optimization E by $U_q(k)$ and w_q), estimation of missing observations is made (optimization E by $\tilde{x}_{ki} \in X_G$). In this case, the algorithm of fuzzy c-means based on the optimal expansion strategy can be written as the following sequence of steps (*Hathaway & Bezdek, 2001*):

1. Setting the initial conditions for the algorithm: $\beta > 0$; $1 < m < N$; $\varepsilon > 0$; $w_q^{(0)}$; $1 \le q \le m$; $\tau = 0,1,2,...,Q$; $X_G^{(0)} = \{-1 \le \hat{x}_{ki}^{(0)} \le 1\}$, where $X_G^{(0)} - N_G (1 \le N_G \le (n-1)N)$ arbitrary initial estimates $\hat{x}_{ki}^{(0)}$ of missing values $\tilde{x}_{ki} \in X_G$;
2. Calculation of membership levels by solving the optimization problem:

$$U_q^{(\tau+1)}(k) = \arg\min_{U_q(k)} E(U_q(k), w_q^{(\tau)}, X_G^{(\tau)}) = \frac{(D^2(\hat{x}_k^{(\tau)}, w_q^{(\tau)}))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}(D^2(\hat{x}_k^{(\tau)}, w_l^{(\tau)}))^{\frac{1}{1-\beta}}} = \frac{(\left\|\hat{x}_k^{(\tau)} - w_q^{(\tau)}\right\|^2)^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}(\left\|\hat{x}_k^{(\tau)} - w_l^{(\tau)}\right\|^2)^{\frac{1}{1-\beta}}}$$

(here vector $\hat{x}_k^{(\tau)}$ differs from $\tilde{x}_k$ by replacing missing values $\tilde{x}_{ki} \in X_G$ by estimates $\hat{x}_{ki}^{(\tau)}$ that are calculated for the τ-th epoch of data processing);

3. Calculate the centroids of clusters:

$$w_q^{(\tau+1)} = \arg\min_{w_q} E(U_q^{(\tau+1)}(k), w_q, X_G^{(\tau)}) = \frac{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta} \hat{x}_k^{(\tau)}}{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}};$$

4. Checking the stop conditions:

if $\left\|w_q^{(\tau+1)} - w_q^{(\tau)}\right\| < \varepsilon \ \forall \ 1 \le q \le m$ or $\tau = Q$, then the algorithm terminates, otherwise go to step 5;

5. Estimation of missing observations by solving the optimization problem:

$$X_G^{(\tau+1)} = \arg\min_{X_G} E(U_q^{(\tau+1)}(k), w_q^{(\tau+1)}, X_G)$$

or, equivalently

$$\frac{\partial E(U_q^{(\tau+1)}(k), w_q^{(\tau+1)}, X_G)}{\partial \hat{x}_{ki}} = 0.$$

That leads to the final expression

$$\hat{x}_{ki}^{(\tau+1)} = \frac{\sum_{q=1}^{m} (U_q^{(\tau+1)}(k))^{\beta} w_{qi}^{(\tau+1)}}{\sum_{q=1}^{m} (U_q^{(\tau+1)}(k))^{\beta}}.$$

Information processing with this algorithm is organized as a sequence

$$w_q^{(0)} \to U_q^{(1)} \to \hat{x}_{ki}^{(1)} \to w_q^{(1)} \to U_q^{(2)} \to \ldots \to w_q^{(\tau)} \to U_q^{(\tau+1)} \to \hat{x}_{ki}^{(\tau+1)} \to w_q^{(\tau+1)} \to \ldots \to w_q^{(Q)}$$

thus, it is possible to organize online clustering by type of procedure (2). For this purpose, we introduce two-time scales: real time $k = 1, 2, \ldots, N, \ldots$, and accelerated computing time $\tau = 0, 1, 2, \ldots, Q$. Here, we assume that between two instants of real time k and $k+1$ implemented Q iterations of accelerated time.

Then we can write a procedure

$$\begin{cases} U_q^{(\tau+1)}(k+1) = \dfrac{(\|\hat{x}_{k+1}^{(\tau)} - w_q(k)\|^2)^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m} (\|\hat{x}_{k+1}^{(\tau)} - w_l(k)\|^2)^{\frac{1}{1-\beta}}}, \\ \hat{x}_{k+1,i}^{(\tau+1)} = \dfrac{\sum_{q=1}^{m} (U_q^{(\tau+1)}(k+1))^{\beta} w_{qi}(k)}{\sum_{q=1}^{m} (U_q^{(\tau+1)}(k+1))^{\beta}}, \\ w_q(k+1) = w_q(k) + \eta(k+1)(U_q^{(Q)}(k+1))^{\beta} (\hat{x}_{k+1}^{(Q)} - w_q(k)), \end{cases} \tag{11}$$

which shows that the memberships and missing observations are calculated in accelerated time, and centroids are calculated in real time by the WTM self-learning rule.

Of course, centroids can be recalculated in accelerated time too:

$$\begin{cases} U_q^{(\tau+1)}(k+1) = \dfrac{(\|\hat{x}_{k+1}^{(\tau)} - w_q^{(\tau)}(k+1)\|^2)^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}(\|\hat{x}_{k+1}^{(\tau)} - w_l^{(\tau)}(k)\|^2)^{\frac{1}{1-\beta}}}, \\ w_q^{(0)}(k+1) = w_q^{(Q)}(k), \\ w_q^{(\tau+1)}(k+1) = w_q^{(\tau)}(k+1) + \eta(k+1)(U_q^{(\tau+1)}(k+1))^{\beta}(\hat{x}_{k+1}^{(\tau)} - w_q^{(\tau)}(k+1)), \\ \hat{x}_{k+1,i}^{(\tau+1)} = \dfrac{\sum\limits_{q=1}^{m}(U_q^{(\tau+1)}(k+1))^{\beta} w_{qi}^{(\tau+1)}(k+1)}{\sum\limits_{q=1}^{m}(U_q^{(\tau+1)}(k+1))^{\beta}}, \end{cases} \quad (12)$$

in this case anyway, both in (11) and (12) operation of summation about k is absent, that for large N can involve a lot of memory.

As already mentioned, to address the problem of fuzzy clustering of data containing outliers, special objective functions of the form (*Bezdek, 2013*; *Yang & Tseng, 1996*) can be used; however, these anomalies are overwhelming, and the problem itself is associated with minimizing these functions. From a practical point of view, it is more convenient to use, instead of the objective functions, based on the metrics, the so-called measures of similarity (SM) (*Hathaway et al., 2002*), which are subject to milder conditions than metrics:

$$\begin{cases} S(\tilde{x}_k, \tilde{x}_p) \ge 0, \\ S(\tilde{x}_k, \tilde{x}_p) = S(\tilde{x}_p, \tilde{x}_k), \\ S(\tilde{x}_k, \tilde{x}_k) = 1 \ge S(\tilde{x}_k, \tilde{x}_p) \end{cases}$$

(no triangle inequality), and clustering problem can be “tied” to maximize these measures.

If the data are transformed so that $-1 \le \tilde{x}_{ki} \le 1$ the measure of similarity can be structured so as to suppress data lying at the edges of interval $[-1,1]$.

Figure 1 illustrates the use of similarity measure as the Cauchy density distribution function with different parameter widths $\sigma^2 < 1$

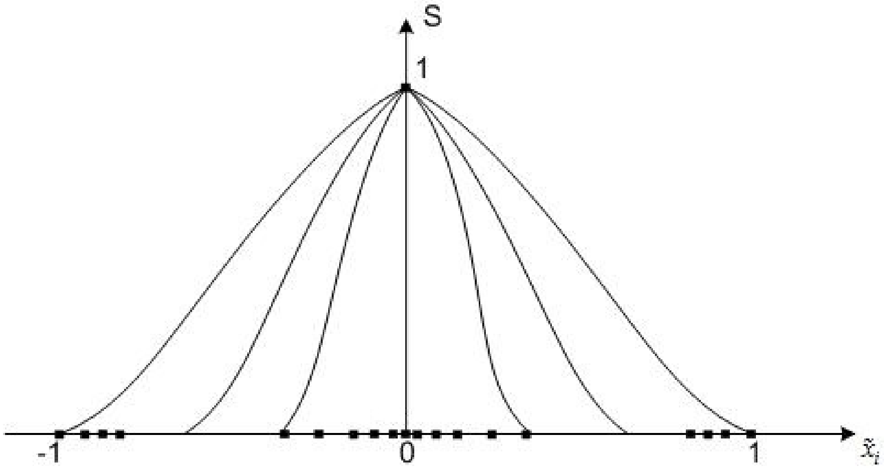

Figure 1 - Similarity measure based on the Cauchy probability distribution

Picking up the width parameter σ^2 functions

$$S(\tilde{x}_k, w_q) = \frac{1}{1+\frac{\left\|\tilde{x}_k - w_q\right\|^2}{\sigma^2}} = \frac{\sigma^2}{\sigma^2 + \left\|\tilde{x}_k - w_q\right\|^2} = \frac{\sigma^2}{\sigma^2 + D^2(\tilde{x}_k, w_q)} \qquad (13)$$

it's possible to exclude the effect outliers, which in principle can not be done using the Euclidean metric

$$D^2(\tilde{x}_k, w_q) = \left\|\tilde{x}_k - w_q\right\|^2. \qquad (14)$$

Further, by introducing the objective function based on similarity measure (13)

$$E_S(U_q(k), w_q) = \sum_{k=1}^{N}\sum_{q=1}^{m} U_q^{\beta}(k) S(\tilde{x}_k, w_q) = \sum_{k=1}^{N}\sum_{q=1}^{m} \frac{U_q^{\beta}(k)\sigma^2}{\sigma^2 + \left\|\tilde{x}_k - w_q\right\|^2},$$

probabilistic constraints

$$\sum_{q=1}^{m} U_q(k) = 1,$$

Lagrange function

$$L_S(U_q(k), w_q, \lambda(k)) = \sum_{k=1}^{N}\sum_{q=1}^{m} \frac{U_q^{\beta}(k)\sigma^2}{\sigma^2 + \left\|\tilde{x}_k - w_q\right\|^2} + \sum_{k=1}^{N} \lambda(k)\Big(\sum_{q=1}^{m} U_q(k) - 1\Big) \qquad (15)$$

(here $\lambda(k)$ is indefinite Lagrange multipliers) and solving the system of Karush-Kuhn-Tucker equations, we arrive at the solution

$$\begin{cases} U_q(k) = \dfrac{(S(\tilde{x}_k, w_q))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m} (S(\tilde{x}_k, w_l))^{\frac{1}{1-\beta}}}, \\ \lambda(k) = -\Big(\sum_{l=1}^{m} (\beta S(\tilde{x}_k, w_l))^{\frac{1}{1-\beta}}\Big)^{1-\beta}, \\ \nabla_{w_q} L_S(U_q(k), w_q, \lambda(k)) = \sum_{k=1}^{N} U_q^{\beta}(k) * \\ \qquad * \dfrac{\tilde{x}_k - w_q}{(\sigma^2 + \left\|\tilde{x}_k - w_q\right\|^2)^2} = \vec{0}. \end{cases} \qquad (16)$$

The last equation (16) has no analytic solution, so to find a saddle point of the Lagrangian (15), we can use the procedure of Arrow-Hurwitz-Uzawa, as a result of which we obtain the algorithm

$$\begin{cases} U_q(k+1) = \dfrac{(S(\tilde{x}_{k+1}, w_q))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}(S(\tilde{x}_{k+1}, w_l))^{\frac{1}{1-\beta}}}, \\ w_q(k+1) = w_q(k) + \eta(k+1)U_q^{\beta}(k+1) * \\ \quad * \dfrac{\tilde{x}_{k+1} - w_q}{(\sigma^2 + \|\tilde{x}_{k+1} - w_q\|^2)^2} = \\ \quad = w_q(k) + \eta(k+1)\varphi_q(k+1)(\tilde{x}_{k+1} - w_q) \end{cases} \quad (17)$$

where

$$\varphi_q(k+1) = \frac{\tilde{x}_{k+1} - w_q}{(\sigma^2 + \|\tilde{x}_{k+1} - w_q\|^2)^2}$$

neighbourhood robust functions of WTM-self-learning rule.

Assuming the fuzzifier value as $\beta = 2$, we arrive at a robust variant of FCM:

$$\begin{cases} U_q(k+1) = \dfrac{(S(\tilde{x}_{k+1}, w_q))}{\sum_{l=1}^{m}(S(\tilde{x}_{k+1}, w_l))}, \\ w_q(k+1) = w_q(k) + \eta(k+1)\dfrac{U_q^2(k+1)}{(\sigma^2 + \|\tilde{x}_{k+1} - w_q\|^2)^2}. \end{cases}$$

Further, using the concept of accelerated time, it is possible to introduce robust adaptive probabilistic fuzzy clustering procedure in the form

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{(S(\tilde{x}_k, w_q^{(\tau)}(k)))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}(S(\tilde{x}_k, w_l^{(\tau)}))^{\frac{1}{1-\beta}}}, \\ w_q^{(Q)}(k) = w_q^{(0)}(k+1), \\ w_q^{(\tau+1)}(k+1) = w_q^{(\tau)}(k+1) + \eta(k+1) * \\ \quad * \dfrac{(U_q^{(Q)}(k))^{\beta}}{(\sigma^2 + \|\tilde{x}_{k+1} - w_q^{(\tau)}(k+1)\|^2)^2} * \\ \quad * (\tilde{x}_{k+1} - w_q^{(\tau)}(k+1)). \end{cases} \quad (18)$$

Similarly, it is possible to synthesize a robust adaptive algorithm for possibilistic (*Hathaway et al., 2002*) fuzzy clustering using criterion

$$E_S(U_q(k), w_q, \mu_q) = \sum_{k=1}^{N}\sum_{q=1}^{m} U_q^{\beta}(k) S(\tilde{x}_k, w_q) + \sum_{q=1}^{m}\mu_q \sum_{k=1}^{N}(1 - U_q(k))^{\beta}.$$

Solving the task of optimization, we obtain the solution:

$$\begin{cases} U_q(k+1) = \left(1 + \left(\dfrac{S(\tilde{x}_{k+1}, w_q(k))}{\mu_q(k)}\right)\right)^{-1}, \\ w_q(k+1) = w_q(k) + \eta(k+1) U_q^{\beta}(k+1) \dfrac{\tilde{x}_{k+1} - w_q(k)}{(\sigma^2 + \|\tilde{x}_{k+1} - w_q(k)\|^2)^2}, \\ \mu_q(k+1) = \dfrac{\sum_{p=1}^{k+1} U_q^{\beta}(p) S(\tilde{x}_p, w_q(k+1))}{\sum_{p=1}^{k+1} U_q^{\beta}(p)}, \end{cases} \tag{19}$$

receiving at $\beta = 2$ the form

$$\begin{cases} U_q(k+1) = \dfrac{1}{1 + \dfrac{S(\tilde{x}_{k+1}, w_q(k))}{\mu_q(k)}}, \\ w_q(k+1) = w_q(k) + \eta(k+1) \dfrac{U_q^2(k+1)}{(\sigma^2 + \|\tilde{x}_{k+1} - w_q(k)\|^2)^2} (\tilde{x}_{k+1} - w_q(k)), \\ \mu_q(k+1) = \dfrac{\sum_{p=1}^{k+1} U_q^2(p) S(\tilde{x}_p, w_q(k+1))}{\sum_{p=1}^{k+1} U_q^2(p)}. \end{cases}$$

And, finally, introducing the accelerated time we obtain the procedure

$$U_q^{(\tau+1)}(k) = \frac{1}{1 + \left(\dfrac{S(\tilde{x}_k, w_q^{(\tau)}(k))}{\mu_q^{(\tau)}(k)}\right)^{\frac{1}{\beta-1}}}, \tag{20a}$$

$$w_q^{(Q)}(k) = w_q^{(0)}(k+1), \tag{20b}$$

$$w_q^{(\tau+1)}(k+1) = w_q^{(\tau)}(k+1) + \eta(k+1) \frac{(U_q^{(Q)}(k))^{\beta}}{(\sigma^2 + \|\tilde{x}_{k+1} - w_q^{(\tau)}(k+1)\|^2)^2} (\tilde{x}_{k+1} - w_q^{(\tau)}(k+1)), \tag{20c}$$

$$\mu_q^{(\tau+1)}(k) = \frac{\sum_{p=1}^{k} (U_q^{(\tau+1)}(p))^{\beta} S(\tilde{x}_p, w_q^{(\tau+1)}(k))}{\sum_{p=1}^{k} (U_q^{(\tau+1)}(p))^{\beta}}. \tag{20d}$$

Nearest prototype strategy (NFS), proposed in (*Hathaway & Bezdek, 2001*), is a modification of FCM-algorithm and leads to the replacement of missing components of the vector observations $\tilde{x}_{ki} \in X_G$ by estimates of the corresponding component prototypes (centroids) of the clusters computed using FCM. Thus, for each $\tilde{x}_{ki} \in X_G$ it's possible to find the prototype $w_q = (w_{q1,\dots,} w_{qi,\dots,} w_{qn})^T$ nearest to $\tilde{x}_k$ in the sense of the partial distance (PD)

$$D_P^2(\tilde{x}_k, w_q) = \frac{n}{\delta_{k\Sigma}} \sum_{i=1}^{n} (\tilde{x}_{ki} - w_{qi})^2 \delta_{ki} \tag{21}$$

where

$$\delta_{ki} = \begin{cases} 0 \mid \tilde{x}_{ki} \in X_G, \\ 1 \mid \tilde{x}_{ki} \in X_F, \end{cases}$$

$$\delta_{k\Sigma} = \sum_{i=1}^{n} \delta_{ki},$$

$w_q^{(\tau+1)} = \arg\min_q \{D_P^2(\tilde{x}_k, w_1^{(\tau+1)}), \dots, D_P^2(\tilde{x}_k, w_m^{(\tau+1)})\}$, then instead $\tilde{x}_{ki} \in X_G$ input estimate $\hat{x}_{ki} \in w_{qi}$ used in place of the missing components.

In (*Shafronenko et al, 2018*) adaptive fuzzy clustering procedure based on the NPS was introduced:

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{\left(\left\| \hat{x}_k^{(\tau)} - w_q^{(\tau)}(k) \right\|^2\right)^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m} \left(\left\| \hat{x}_k^{(\tau)} - w_l(k) \right\|^2\right)^{\frac{1}{1-\beta}}}, \\ \hat{x}_{ki}^{(\tau)} = w_{qi}^{(\tau)}(k), \quad w_q^{(\tau)}(k) = \arg\min_q \{D_P^2(\tilde{x}_k, w_1^{(\tau)}(k)), \dots, D_P^2(\tilde{x}_k, w_m^{(\tau)}(k))\}, \\ w_q^{(Q)}(k) = w_q^{(0)}(k+1), \\ w_q^{(\tau+1)}(k+1) = w_q^{(\tau)}(k+1) + \eta(k+1)(U_q^{(Q)}(k))^{\beta} (\hat{x}_k^{(\tau)} - w_q^{(\tau)}(k+1)) . \end{cases} \tag{22}$$

To address the problem of robust data clustering with missing values, let's introduce the partial similarity measure (PCM), a hybrid of the partial distance (PD) (21) and a similarity measure (SM). It is easy to see that such a PSM has the form

$$S_P(\tilde{x}_k, w_q) = \frac{\sigma^2}{\sigma^2 + D_P^2(\tilde{x}_k, w_q)}, \tag{23}$$

that allows to obtain the desired properties of algorithms based on procedures described above.

So, on the basis of the procedures (23) and (20) we can introduce the robust adaptive probabilistic fuzzy clustering algorithm for data with missing values:

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{(S_P(\hat{x}_k^{(\tau)}, w_q^{(\tau)}(k)))^{\frac{1}{\beta-1}}}{\sum_{l=1}^{m} (S_P(\hat{x}_k^{(\tau)}, w_l^{(\tau)}))^{\frac{1}{\beta-1}}}, \\ \hat{x}_{ki}^{(\tau)} = w_{qi}^{(\tau)}, \; w_q^{(\tau)}(k) = \arg\max_q \{S_P(\tilde{x}_k^{(\tau)}, w_1^{(\tau)}(k)), \dots, S_P(\tilde{x}_k^{(\tau)}, w_m^{(\tau)}(k))\}, \\ w_q^{(Q)}(k) = w_q^{(0)}(k+1), \\ w_q^{(\tau+1)}(k+1) = w_q^{(\tau)}(k+1) + \eta(k+1) \dfrac{(U_q^{(Q)}(k))^{\beta}}{(\sigma^2 + \left\| \hat{x}_{k+1}^{(\tau)} - w_q^{(\tau)}(k+1) \right\|^2)^2} (\hat{x}_{k+1}^{(\tau)} - w_q^{(\tau)}(k+1)), \end{cases} \tag{24}$$

based on procedures (24) and (19), we can also write a robust adaptive algorithm for possibilistic fuzzy clustering of data with missing values:

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{1}{1+\left(\dfrac{S^{-1}(\hat{x}_k, w_q^{(\tau)}(k))}{\mu_q^{(\tau)}(k)}\right)^{\frac{1}{\beta-1}}}, \\ \hat{x}_{ki}^{(\tau)} = w_{qi}^{(\tau)},\ w_q^{(\tau)}(k) = \arg\max_q \{S_P(\tilde{x}_k^{(\tau)}, w_1^{(\tau)}(k)), ..., S_P(\tilde{x}_k^{(\tau)}, w_m^{(\tau)}(k))\}, \\ w_q^{(Q)}(k) = w_q^{(0)}(k+1), \\ w_q^{(\tau+1)}(k+1) = w_q^{(\tau)}(k+1) + \eta(k+1)\dfrac{(U_q^{(Q)}(k))^\beta}{(\sigma^2 + \left\|\hat{x}_{k+1} - w_q^{(\tau)}(k+1)\right\|^2)^2}(\hat{x}_{k+1}^{(\tau)} - w_q^{(\tau)}(k+1)), \\ \mu_q^{(\tau+1)}(k) = \dfrac{\sum_{p=1}^{k}(U_q^{(\tau+1)}(p))^\beta S_P^{-1}(\hat{x}_p, w_q^{(\tau+1)}(k))}{\sum_{p=1}^{k}(U_q^{(\tau)}(p))^\beta}. \end{cases} \tag{25}$$

Thus, the use of a partial similarity measure based on partial distance (21) allows us to solve the problem of fuzzy clustering of data containing both missing values and outliers.

Discussion

The results of the study demonstrate that online fuzzy clusterization of distorted data streams should be considered not merely as a modification of classical clustering procedures, but as a separate methodological problem requiring the simultaneous treatment of three complicating factors: sequential data arrival, missing feature values and anomalous observations. Classical fuzzy clustering methods were originally developed for complete datasets processed in batch mode. In contrast, many applied systems in economics, engineering diagnostics, technical monitoring, transport, energy management, medicine and socio-economic analysis operate with continuously arriving observations whose structure may be incomplete, noisy and dynamically changing. This makes the transition from batch clustering to adaptive online neuro-fuzzy procedures theoretically and practically significant.

The proposed approach extends the conventional fuzzy c-means framework by transforming it into recurrent online learning procedures. This transformation is important because repeated recalculation of the entire clustering model after each new observation becomes computationally inefficient for large data streams. By updating membership levels and cluster prototypes sequentially, the method reduces memory requirements and enables real-time processing. The use of a Kohonen-type neuro-fuzzy architecture provides an additional advantage because cluster prototypes can be interpreted as adaptive weight vectors, while the update mechanism follows a self-learning logic. This interpretation connects fuzzy clustering with neural network learning and makes the proposed approach suitable for evolving intelligent systems.

A significant methodological contribution of the study lies in the use of the "winner-takes-more" principle instead of the stricter "winner-takes-all" rule. In distorted and uncertain data environments, an incoming observation cannot always be unambiguously assigned to a single cluster. The fuzzy membership mechanism allows several prototypes to be updated with different intensities, reflecting the degree to which the observation is associated with each cluster. This is especially important for overlapping clusters and transitional states in multidimensional data streams. Therefore, the proposed recurrent procedure preserves the flexibility of fuzzy partitioning while ensuring the adaptivity required for online processing.

The treatment of missing values is one of the central issues addressed in the study. In many practical datasets, omissions arise from measurement errors, transmission failures, incomplete registration, sensor malfunction or human-related factors. Traditional strategies such as deleting incomplete observations or replacing missing values with mean or median estimates may distort the cluster structure and reduce analytical reliability. The partial-distance strategy used in the study is more appropriate for online clustering because it allows incomplete observations to be processed directly, using only the available components of each vector. This makes it possible to avoid preliminary imputation and maintain the continuity of data stream processing.

At the same time, the study correctly recognises that the partial-distance strategy has limitations when the proportion of missing components becomes too large. In such cases, the available information may be insufficient for reliable membership estimation, and therefore the optimal expansion strategy becomes necessary. Treating missing values as additional variables estimated during clustering is a stronger methodological solution because data recovery and cluster formation are performed within a single optimisation process. This makes the clustering procedure more internally coherent: missing values are not restored independently from the clustering model, but are inferred from the same prototype and membership structure that defines the data partition.

The introduction of two-time scales—real time and accelerated computational time—is particularly important for online implementation. Real time corresponds to the arrival of new observations, while accelerated time allows several internal iterations to be performed between two incoming data points. This construction gives the method a flexible computational organisation. If the stream arrives rapidly, the algorithm can perform minimal updates; if time permits, it can refine memberships and missing-value estimates before updating prototypes. Such a design is highly relevant for real-world systems where data rates may vary and where the balance between computational accuracy and processing speed must be dynamically controlled.

The study also addresses the limitations of probabilistic fuzzy clustering. In probabilistic algorithms, the sum of membership levels for each observation must equal one. This assumption is useful for mathematically well-structured cluster partitions, but it may be restrictive when data contain outliers or observations that do not naturally belong to any existing cluster. The incorporation of possibilistic fuzzy clustering weakens this restriction and makes the model more robust to atypical observations. Possibilistic membership reflects the absolute compatibility of an observation with a cluster rather than its relative distribution among all clusters. This is especially valuable in distorted data streams, where abnormal outputs may otherwise force artificial redistribution of membership levels and distort cluster prototypes.

The use of similarity measures based on the Cauchy distribution further strengthens the robustness of the proposed approach. Distance-based objective functions are sensitive to outliers because large deviations may dominate the optimisation process and shift cluster centroids in undesirable directions. Similarity-based measures operate differently: they can suppress the influence of remote or edge observations by assigning them lower similarity values. The width parameter of the Cauchy-type function becomes a control mechanism regulating the sensitivity of the algorithm to deviations. This makes the clustering procedure more stable in the presence of abnormal outputs and provides a practical alternative to purely metric-based clustering.

An important advantage of the proposed methodology is that missing values and outliers are not treated as separate problems. In many applied systems, incomplete observations and anomalous values occur simultaneously. Methods designed only for omissions may be unstable in the presence of outliers, while robust methods designed for anomalies may not handle missing feature values. The partial similarity measure based on partial distance creates a unified mechanism for processing both types of distortion. This makes the proposed framework more applicable to real data streams, where data quality problems are usually mixed rather than isolated.

From a theoretical perspective, the study contributes to the development of evolving neuro-fuzzy systems by combining several methodological lines: fuzzy c-means clustering, partial-distance strategies, optimal completion of missing observations, possibilistic membership modelling, similarity-based robustification and Kohonen-type recurrent learning. The strength of this contribution lies not in proposing a single isolated algorithm, but in constructing a family of related adaptive procedures that can be selected depending on the type and degree of data distortion. This modularity increases the methodological flexibility of the approach and allows it to be adapted to different application contexts.

From a computational perspective, the proposed algorithms are promising because they avoid repeated summation over the entire historical dataset during recurrent updates. This is particularly relevant for large-scale data streams, where storing and reprocessing all previous observations may be impractical. The representation of the neuro-system as a set of adaptive linear associators also supports fast numerical implementation. Such computational simplicity is important for embedded systems, monitoring platforms and online decision-support tools, where available memory and processing time may be limited.

The practical significance of the study is broad. Online fuzzy clusterization of distorted data streams can be used in technical diagnostics, where sensor data may contain missing values and abnormal measurements; in financial and socio-economic monitoring, where indicators may arrive irregularly and contain reporting gaps; in medical data analysis, where incomplete patient records and atypical observations are common; and in transport, energy and military systems, where real-time adaptation is essential. The ability to process incomplete and anomalous observations without interrupting the clustering procedure makes the proposed approach suitable for operational environments where delayed batch preprocessing is impossible.

Nevertheless, several limitations should be acknowledged. The study is primarily theoretical and algorithmic. Although the proposed procedures are mathematically grounded, their empirical performance should be further evaluated on benchmark datasets and real data streams. Future validation should include comparisons with classical fuzzy c-means, possibilistic fuzzy c-means, Kohonen self-organising maps, robust clustering algorithms and modern online clustering methods. Such comparison should be based on clustering accuracy, stability of prototype updates, reconstruction quality for missing values, robustness to outliers, computational complexity and memory consumption.

Another limitation concerns the selection of algorithmic parameters. The fuzzifier, learning rate, accuracy threshold, number of clusters, number of accelerated internal iterations, scale parameters in possibilistic clustering and width parameter of the Cauchy similarity function may significantly influence the quality of clustering. The article provides the mathematical framework for these procedures, but further research should develop adaptive parameter-selection strategies. In particular, it would be useful to investigate mechanisms for automatic adjustment of the learning rate and similarity-function width depending on the observed level of noise, missingness and outlier intensity in the data stream.

The issue of cluster number selection also remains important. Many practical data streams are non-stationary: new clusters may emerge, old clusters may disappear, and existing clusters may change their position or shape over time. The present framework provides mechanisms for updating prototypes and membership levels, but further development may include evolving structure adaptation, where the number of clusters is not fixed in advance. Such an extension would bring the proposed method closer to fully evolving clustering systems capable of adapting not only parameters, but also model structure.

Future studies should also examine the behaviour of the proposed algorithms under different missingness mechanisms. Missing data may be missing completely at random, missing at random, or missing not at random. The effectiveness of partial-distance and optimal expansion strategies may differ depending on the mechanism that produces omissions. Similarly, the type of outliers should be considered: isolated point anomalies, contextual anomalies and collective anomalies may affect cluster prototypes in different ways. A systematic simulation study could clarify the robustness boundaries of the proposed methods.

Another promising direction is the integration of the proposed online fuzzy clustering procedures with forecasting, classification and anomaly detection systems. Since clustering can reveal hidden structure in streaming data, the resulting membership levels and prototype dynamics may serve as additional features for downstream predictive models. In diagnostic systems, changes in membership patterns may indicate early-stage abnormal behaviour. In socio-economic monitoring, evolving clusters may reflect changes in the structure of regional or institutional indicators. Therefore, the proposed methods may become part of broader intelligent decision-support systems.

Overall, the study provides a coherent algorithmic basis for online fuzzy clusterization of distorted data streams. Its main value lies in the simultaneous consideration of online operation, fuzzy uncertainty, missing observations and outlier robustness. By combining partial-distance processing, optimal expansion, possibilistic membership and similarity-based robustification within a neuro-fuzzy Kohonen framework, the proposed approach addresses several limitations of classical clustering algorithms. The method therefore has strong potential for applications in dynamic environments where data arrive sequentially, are incomplete or distorted, and must be analysed without full batch retraining.

Conclusion

The problem of clustering data in object-property tables, with gaps in the mode of sequential receipt of this data for processing.

An adaptive neural network method is proposed that enables online clustering of damaged data with constant correction of recoverable table elements and cluster centroids. The introduced neuro-system consists of a set of adaptive linear associators, characterized by high speed and ease of numerical implementation.

Methods of probabilistic and probabilistic fuzzy online data clustering with gaps based on the partial distance strategy are proposed, and it is shown that it can be solved on the basis of the Kohonen self-organizing neuro-fuzzy network and the proposed self-learning algorithm, which is a hybrid of the "winner gets more" rule and recurrent fuzzy clustering algorithms.

A group of adaptive robust fuzzy methods is proposed for clustering, allowing online processing of distorted data containing both omissions and abnormal outliers. The proposed methods are based on the classical procedures of fuzzy c-means J. Bezdek, self-training map T. Kohonen, as well as a specially introduced measure of similarity, allowing you to work with distorted information. The considered methods are simple to implement numerically, being essentially gradient-based procedures for optimizing objective functions of a special type.

Data clustering methods based on recurrent optimization of a special type of objective functions, while the missing observations are replaced by estimates also obtained in the process of solving the optimization problem.

Methods of probabilistic and probabilistic adaptive fuzzy clustering of distorted data streams are proposed, based on the strategy of the closest prototype centroid. The developed approach is computationally simple, and the processing can be organized around a self-organizing Kohonen map.

The methods of probabilistic and probabilistic fuzzy adaptive clustering for data containing an unknown number of a priori gaps, based on the optimal expansion strategy, are proposed.

Funding

No external funding was received.

Conflict of Interest

The author declares that there is no conflict of interest.

Acknowledgements:

Not applicable.

References:

Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.

Bezdek, J. C., Keller, J., Krisnapuram, R., & Pal, N. (1999). *Fuzzy models and algorithms for pattern recognition and image processing* (Vol. 4). Springer Science & Business Media.

Bodyanskiy, Y., Gorshkov, Y., Kokshenev, I., & Kolodyazhniy, V. (2005). Robust recursive fuzzy clustering algorithms. *Proceedings of East West Fuzzy Colloquim*, 301–308.

Bodyanskiy, Y., Shafronenko, A., & Volkova, V. (2012). Adaptive clustering of incomplete data using neuro-fuzzy Kohonen network. *Artificial Intelligence Methods and Techniques for Business and Engineering Applications"–Rzeszow-Sofia: ITHEA*, 287–296.

Bodyanskiy, Y., Shafronenko, A., & Volkova, V. (2012). Adaptive fuzzy probabilistic clustering of incomplete data. Information Models *& Analyses*, *112*.

Chung, F. L., & Lee, T. (1994). Fuzzy competitive learning. *Neural Networks*, *7*(3), 539–551.
Gorshkov, Y., Kolodyazhniy, V., & Bodyanskiy, Y. (2009, June). New recursive learning algorithms for fuzzy Kohonen clustering network. In *Proc. 17th Int. Workshop on Nonlinear Dynamics of Electronic Systems* (pp. 58–61).
Hathaway, R. J., & Bezdek, J. C. (2001). Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *31*(5), 735–744.
Hathaway, R. J., Bezdek, J. C., & Hu, Y. (2002). Generalized fuzzy c-means clustering strategies using L/sub p/norm distances. *IEEE transactions on Fuzzy Systems*, *8*(5), 576–582.
Höppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*. John Wiley & Sons.
Kohonen, T., & Maps, S. O. (1995). Springer series in information sciences. *Self-organizing maps*, *30*.
Marwala, T. (Ed.). (2009). *Computational intelligence for missing data imputation, estimation, and management: Knowledge optimization techniques: Knowledge optimization techniques*. IGI global.
Shafronenko, A., Dolotov, A., Bodyanskiy, Y., & Setlak, G. (2018, August). Fuzzy clustering of distorted observations based on optimal expansion using partial distances. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)* (pp. 327–330). IEEE.
Yang, S. S., & Tseng, C. S. (1996). An orthogonal neural network for function approximation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *26*(5), 779–785.