Neural Process Lett (2016) 43:231–253
DOI 10.1007/s11063-015-9415-8

Asymmetric k-Means Clustering of the Asymmetric Self-Organizing Map

Dominik Olszewski[1]

Published online: 19 March 2015

Abstract An asymmetric approach to clustering of the asymmetric self-organizing map is proposed. The clustering is performed using an improved asymmetric version of the well-known k-means algorithm. The improved asymmetric k-means algorithm is the second proposal of this paper. As a result, we obtain a two-stage fully asymmetric data analysis technique. In this way, we maintain the methodological consistency of the both utilized methods, because they are both formulated in asymmetric versions, and consequently, they both properly adjust to asymmetric relationships in analyzed data. The results of our experiments on real data confirm the effectiveness of the proposed approach.

Keywords Self-organizing map · Asymmetric self-organizing map · Clustering · k-means algorithm · Asymmetric k-means algorithm

1 Introduction

The self-organizing map (SOM) [1–7] is an example of the artificial neural network architecture. It was introduced by T. Kohonen in [8] as a generalization and extension of the concepts proposed in [9]. This approach can be also interpreted as a visualization technique, since the algorithm may perform a projection from multidimensional space to 2-dimensional space, this way creating a map structure. The location of points in 2-dimensional grid aims to reflect the similarities between the corresponding objects in multidimensional space. Therefore, the SOM algorithm allows for visualization of relationships between objects in multidimensional space. The asymmetric version of the SOM algorithm was introduced in [10], and the justification of the asymmetric approach was extended in [11].

The k-means clustering algorithm [12–17] is a well-known statistical data analysis tool used in order to form arbitrary settled number of clusters in an analyzed dataset. The algorithm

✉ Dominik Olszewski
dominik.olszewski@ee.pw.edu.pl

[1] Faculty of Electrical Engineering, Warsaw University of Technology, Warsaw, Poland

aims to separate clusters of possibly most similar objects. An object represented as a vector of d features can be interpreted as a point in d-dimensional space. Hence, the k-means algorithm can be formulated as follows: given n points in d-dimensional space, and the number k of desired clusters, the algorithm seeks a set of k clusters so as to minimize the sum of squared dissimilarities between each point and its cluster centroid. The name "k-means" was introduced in [15], however, the algorithm, itself, was formulated by H. Steinhaus in [16].

An asymmetric version of the k-means clustering algorithm was introduced in [18]. However, the asymmetry in the algorithm from [18] arises caused by usage of dissimilarities, which are asymmetric by definition (for example, the Kullback–Leibler divergence). On the other hand, the paper [19] proposes an asymmetric k-means algorithm using symmetric similarities, which are asymmetrized by employing the asymmetric coefficients. This kind of approach provides a proper adjustment to asymmetric relationships in analyzed data (explained in detail in Sect. 3). Therefore, in this paper, we utilize the asymmetric version of the k-means algorithm introduced in [19], we improve it, and employ it for cluster analysis on the asymmetric SOM.

1.1 Our Proposal

The improvement of the asymmetric k-means algorithm, introduced in this paper, consists in utilizing the current number of objects of clusters, when computing the asymmetric similarities in each cycle of the k-means clustering process. In this way, the algorithm can successfully handle even the datasets containing clusters of considerably different number of objects. The goal is achieved by incorporating a mechanism of different treating of different clusters in a dataset, according to theoretical fundamentals of the hierarchy-based asymmetric approach in data analysis (explained in detail in Sects. 3, 4, and 6). In order to accomplish the aforementioned purpose, we introduce the cluster coefficients, which convey the information about the current number of objects in clusters. The novel improved version of the asymmetric k-means algorithm uses both coefficients—the asymmetric coefficients, like it was done in [19], and cluster coefficients, which are the proposal of this paper.

Finally, we combine the asymmetric SOM visualization technique and the improved asymmetric k-means algorithm in order to perform the two-stage asymmetric cluster analysis.

The general order of data analysis in our work is the following: First, the asymmetric SOM is generated, and then, the neurons (processing units) in the grid of the asymmetric SOM are clustered using the proposed asymmetric k-means algorithm. In other words, the clustering process is carried out in the output space of the asymmetric SOM, i.e., in 2-dimensional space.

In this way, we maintain the methodological consistency between the asymmetric SOM and the asymmetric k-means, i.e., both employed methods are asymmetry-sensitive, and therefore, both can effectively operate on asymmetric data. As a result, we obtain a fully asymmetric two-stage data analysis approach.

Recapitulating, this paper proposes:

- the improvement of the asymmetric k-means algorithm,
- the asymmetric k-means clustering of the asymmetric SOM.

It is worthy of mentioning that the asymmetric version of the k-means clustering algorithm can be recognized as a generalization of this method, which makes it capable to handle data regardless whether it is symmetric or asymmetric.

A complete theoretical justification of both of the proposals of the present paper is provided in Sect. 4.

1.2 Remainder of this Paper

The rest of this paper is organized as follows: In Sect. 2, the related work is discussed as a background for our study; in Sect. 3, the phenomenon of asymmetry in data analysis is discussed, and the cluster coefficients are introduced; in Sect. 4, a theoretical justification of the two methods proposed in the present paper is provided; in Sect. 5, the asymmetric version of the SOM technique is presented; in Sect. 6, the second proposal of the paper (i.e., the asymmetric k-means clustering of the asymmetric SOM) is described; in Sect. 7, the results of the experimental study on real data in four different research fields are reported; while in Sect. 9, the whole paper is summarized, and certain directions for future research are given.

2 Related Work

This section presents the state-of-the-art in the field of asymmetric data analysis. However, we claim that the problem of asymmetry in data analysis has not gained the deserved attention, and it has been relatively rarely studied in the literature.

One of the first researchers dealing with the asymmetric approach in data analysis was A. Tversky, who questioned the geometric representation of similarity [20]. He argued that the notion of similarity had been dominated by geometric models, which represent objects as points in some coordinate space and that dissimilarities between objects simply correspond to the metric distances between the points. He argues that a similarity statement, such as "a is like b", is directional. It has a subject and a referent, and is not equivalent to the statement "b is like a". His well-known example states that "North Korea is more similar to China than China to North Korea", since China is a larger and a more general entity. Or, we say "the son resembles the father" rather that "the father resembles the son", since the father is the more prominent entity. His claims were validated in his numerous psychological experiments [21], and his idea was undoubtedly an inspiration for many later works concerning the asymmetric dissimilarities and the general problem of asymmetry in data analysis.

Another similar concept closely related to the idea of Tversky appears in the work of M. Martín-Merino and A. Muñoz [10], where the asymmetric version of the SOM was proposed. The authors notice the same asymmetric directional relationships between objects of different level of generality (or prominence). Their example from the field of textual data analysis concerns the dissimilarity between the two words—"mathematics" and "Bayes". The former is a more general entity, which makes the relationship between those words strongly asymmetric. Symmetric dissimilarities produce large values for most pairs of objects, and consequently, they do not reflect properly the associations between objects of different level of generality. As it is stated in [10], asymmetry can be interpreted as a particular type of hierarchy. In [22], Martín-Merino and Muñoz also find the cause of asymmetric nature of data in hierarchical relationships between objects. Diego, Muñoz and Moguerza [23] combine several similarity matrices into one kernel, and train a Support Vector Machine.

The asymmetric SOM from [10] was later extended in [11], where the justification of the method was proved to be applicable in a wider research spectrum than only textual data analysis.

The problem of asymmetric proximities has been also discussed by E. Holman [24]. He has proposed a series of monotonic models for asymmetric proximities. Each of these models represents an asymmetric square data matrix as a monotonic combination of a symmetric

function and a bias function. These models generalize his several previous models for proximity and dominance data. His publication [24] became an inspiration to the later work of D. Weeks and P. M. Bentler [25], and also to the work of B. Zielman and W. Heiser [26,27]. Zielman and Heiser in [27] consider the models for asymmetric proximities as a combination of a symmetric similarity component and an asymmetric dominance component. They apply a certain decomposition to the model parameters, clearly separating the dominance and symmetric similarity components. This decomposition allowed them to classify the models discussed into two categories: one that assumes that the asymmetric relationships are transitive, and the other one that consists of models that can also represent circular asymmetric relationships. A very clear reference to this concept can be found in the work of G. Bove [28], where the possibilities of incorporating external information to the models for asymmetric proximities are shown. Such information can help the analysis of proximities between rows and columns of data matrices.

The research of A. Okada and T. Imaizumi [29–32] is focused on using the dominance point governing asymmetry in the proximity relationships among objects, represented as points in the multidimensional Euclidean space. They have concentrated in their work on the multidimensional scaling for analyzing one-mode two-way (object $\times$ object) and two-mode three-way (object $\times$ object $\times$ source) asymmetric proximities. They claim that ignoring or neglecting the asymmetry in proximity analysis discards potentially valuable information.

In the paper [33], the usage of an asymmetric dissimilarity measure in centroid-based clustering is proposed. The dissimilarity employed is the Alpha–Beta divergence, which is asymmetrized using its parameters, and the degree of asymmetry of the Alpha–Beta-divergence is computed on the basis of the within-cluster variances.

In general, the asymmetric dissimilarities were found to be useful in a variety of research fields, and their property of asymmetry was not acknowledged as an inhibiting constraint. For example, the well-known Kullback–Leibler divergence attracted the attention of scientists from different areas [34–36].

Finally, the paper [18] introduces an asymmetric version of the k-means clustering algorithm using the dissimilarities, which are asymmetric by definition (for example, the Kullback–Leibler divergence), and the paper [19] proposes an asymmetric k-means algorithm using the asymmetric coefficients.

3 Asymmetry in Data

When an analyzed dataset appears to have asymmetric properties, the symmetric measures of similarity or dissimilarity (for example, the most popular Euclidean distance) does not apply properly to this phenomenon, and for most pairs of data points, they produce small values (similarities) or big values (dissimilarities). Consequently, they do not reflect accurately the relationships between objects. The asymmetry in dataset arises, for example, in case, when the data associations have a hierarchical nature. The hierarchical connections in data are closely related to the asymmetry. This relation has been noticed in [37]. In case of a dissimilarity, when it is computed in the direction—from a more general entity to a more specific one—it should be greater than in the opposite direction. As stated in [10], asymmetry can be interpreted as a particular type of hierarchy.

An idea to overcome this problem is to employ the asymmetric similarities and dissimilarities. They should be applied in algorithms in such a way, so that they would properly reflect the hierarchical asymmetric relationships between objects in analyzed dataset. Therefore, it

should be guaranteed that their application is consistent with the hierarchical associations in data. This can be achieved by use of the asymmetric coefficients and cluster coefficients, inserted in the formulae of symmetric measures. In this way, we can obtain the asymmetric measures on the basis of the symmetric ones. The asymmetric coefficients and cluster coefficients should assure the consistency with the hierarchy. Hence, in case of the similarities, they should assure greater values in the direction—from more specific concept to more general one.

3.1 Asymmetric Coefficients

Asymmetric coefficients convey the information provided by asymmetry. Two coefficients were introduced in [22]. The first one is derived from fuzzy-logic-based index, and the second one is formulated on the basis of the Kullback–Leibler divergence. Both of these quantities are widely used in statistics and probability theory. In our experimental study, we have used an asymmetric coefficient based on the fuzzy-logic-based index.

Hence, the asymmetric coefficient employed in our research is formulated as follows:

$$a_i = \frac{\varrho(x_i)}{\max_j\left(\varrho\left(x_j\right)\right)}, \tag{1}$$

where x_i, $i = 1, \ldots, n$, is the ith object in analyzed dataset; x_j, $j = 1, \ldots, n$, is the jth object in analyzed dataset; n is the total number of objects; and $\varrho(\cdot)$ is a function defined in the following way:

$$\varrho(x_i) = |\{x \in X \mid d_{\text{Euc}}(x_i, x) \leq \tau\}|, \tag{2}$$

where x_i, $i = 1, \ldots, n$, is the ith object in analyzed dataset; n is the total number of objects; X is the full dataset in the Euclidean space $\mathbb{R}^d$; d is the number of dimensions in the dataset X (i.e., the number of features of the objects x_i); $\tau > 0$ is the arbitrarily settled tolerance threshold; $|\cdot|$ is the cardinality of a set; and $d_{\text{Euc}}(\cdot, \cdot)$ is the Euclidean distance.

The function $\varrho(\cdot)$ allows for establishing micro-regions (determined using the parameter τ), in which a highly similar data objects are located. Assuming a sufficiently small value of τ, the objects in those micro-regions are so similar that from the point of view of the asymmetric data analysis, they may be treated as the same. Consequently, one can talk about multiple occurrences of objects in an analyzed dataset.

In case of certain datasets, the data objects may indeed occur repeatedly—like, in case of textual data, words can (and usually do) occur a number of times. However, in case of different datasets—like in case of datasets consisting of time series, there may be no exact repetitions of data objects, whereas the asymmetry phenomenon may still exist and affect the performance of data analysis. In this scenario, the function $\varrho(\cdot)$ is a solution to grasp and take into account the asymmetric data properties, which do exist because of the hierarchical data relationships, even though, the data objects do not occur multiple times as entire feature vectors.

The asymmetric coefficient defined in this way measures the normalized frequencies of occurrences of the objects pointed out by the function $\varrho(\cdot)$ (the same objects or the very similar ones) in the high-dimensional input space. And this is a way to express numerically the level of generality of an object in the input space. The objects occurring frequently in the input space may be recognized as the general ones. On the other hand, the objects occurring rarely in the input space are the specific entities. The phenomenon is particularly apparent and noticeable in case of the textual data, where the general words (objects) occur definitely more frequently than the specific words.

The asymmetric coefficient takes values in the $\langle 0, 1\rangle$ interval. Intuitively speaking, it will become large for general (broad) concepts.

The asymmetric coefficient is assigned to each object in an analyzed dataset.

In our experimental study, we have investigated textual and time series datasets in order to verify the aforementioned theoretical claims, and in order to provide a full and thorough evaluation of the proposed approach.

The tuning of the value of the parameter τ is described in Sect. 7.2.

3.2 Cluster Coefficients

Cluster coefficients allow to utilize the information about the cluster memberships. In other words, they convey the information about the cardinality of clusters. Cluster centroids are computed on the basis of objects belonging to a given cluster. Consequently, a centroid of a cluster reflects the properties of all objects in that cluster. Therefore, cluster centroids are the entities of a very high level of generality, and consequently, they strongly generate the hierarchy in data analysis. Considering that the hierarchical associations result in asymmetric character of data, the cluster centroids essentially affect the asymmetric relationships between objects in an analyzed dataset, and this fact should be taken into account, when the similarities are computed.

In this paper, we introduce the following cluster coefficient:

$$\eta_j = \begin{cases} \frac{n_j}{\max_i \varrho(x_i)} & \text{for the direction from object to centroid} \\ \frac{1}{\max_i \varrho(x_i)} & \text{for the direction from centroid to object} \end{cases}, \tag{3}$$

where n_j, $j = 1, \ldots, k$, is the number of objects in the jth cluster; k is the number of clusters; $i = 1, \ldots, n$; n is the total number of objects; and $\varrho(\cdot)$ is the function defined in (2).

This coefficient takes values in the $\langle 0, n_j \rangle$ interval. It becomes larger for clusters with a larger number of objects (when the direction from object to centroid is considered).

The cluster coefficient is assigned to each cluster in an analyzed dataset. The value of the cluster coefficient is constant within a cluster in case of both directions—from object to centroid and from centroid to object, whereas the values in case of the direction from centroid to object are also constant for the entire analyzed dataset. The reason of this is that the cluster coefficient aims to distinguish clusters only on the basis of the computation corresponding to the direction from object to centroid.

The values of the cluster coefficients need to be updated each time, when the cardinality of a cluster changes. Hence, in case of the k-means clustering algorithm, the values of the cluster coefficients should be updated each time a new object is assigned to a cluster.

4 Justification of the Two Proposed Methods

The motivation behind the proposed enhanced asymmetric k-means clustering algorithm is that clusters of different number of objects (in other words, of different size measured on the basis of the current number of objects in clusters) should be treated differently in the clustering process. The justification of this modification is that according to the hierarchy-based asymmetry theory in data analysis (described in Sect. 3), the more general entities should be treated in a privileged manner, i.e., the dissimilarities should produce greater values, when computed from these general entities to those more specific ones, than, when computed in the opposite direction. Following this idea, the clusters of greater number of

objects are the more general entities than the clusters with lower number of objects, and most of all, than single objects in a dataset. Consequently, the "big" clusters should be treated in a privileged way, and the dissimilarities from the centroids representing these clusters to single objects in a dataset should be greater than in the opposite direction, and the degree of this asymmetry should be proportionate to the current sizes of the clusters (current numbers of objects in the clusters, in the current k-means cycle—Steps 1 and 2 described in Sect. 6). This mechanism assures that clusters of different size are treated in a different way by the asymmetric clustering algorithm according to the theoretical principles of the asymmetric approach in data analysis.

The described enhancement to the asymmetric k-means clustering algorithm is achieved by incorporating the cluster coefficients (introduced in Sect. 3.2) in the existing version of the asymmetric k-means algorithm operating on the asymmetric coefficients (described in Sect. 3.1). The usage of the cluster coefficients in the algorithm is shown in Sect. 6. The proposed technique satisfies the theoretical claims and requirements described before in this section, and simultaneously, it is mathematically simple.

The second proposal of the present paper is motivated by an endeavor of establishing a combination of the two asymmetric data analysis methods, i.e., the visualization by means of the asymmetric SOM technique and the asymmetric k-means clustering algorithm. The aim is to maintain the methodological consistency of the entire data analysis approach. As a result, one obtains a two-stage fully asymmetric data visualization and clustering technique.

Both of the methods, i.e., SOM and k-means, are examples of (dis)similarity-based data analysis approaches. Furthermore, it is a well-known fact that SOMs with low number of neurons operate in a way similar to the k-means algorithm, which confirms the close relationship between these two methods, especially in terms of the role and significance of the (dis)similarities. This observation is a point of departure of our second proposal. Establishing a well-grounded and theoretically justified combination of SOM and k-means requires utilizing the same (dis)similarity measure. And since it has been proven in [10] and in [11] that the asymmetric SOM is superior over its symmetric counterpart, we claim that the marriage of asymmetric SOM and asymmetric k-means (especially in our enhanced version) is a promising and justified idea.

5 Asymmetric Self-Organizing Map

The traditional symmetric SOM algorithm provides a non-linear mapping between a high-dimensional original data space and a 2-dimensional map of neurons. The neurons are arranged according to a regular grid, in such a way that the similar vectors in input space are represented by the neurons close in the grid. Therefore, the SOM technique visualize the data associations in the input high-dimensional space.

According to [10], the results obtained by the SOM method are equivalent to the results obtained by minimizing the following error function with respect to the prototypes w_r and w_s:

$$e = \sum_r \sum_{x_i \in V_r} \sum_s h_{rs} d^2_{\text{Euc}}(x_i, w_s) \tag{4}$$

$$\approx \sum_r \sum_{x_i \in V_r} d^2_{\text{Euc}}(x_i, w_r) \; + \; K \sum_r \sum_{s \neq r} h_{rs} d^2_{\text{Euc}}(w_r, w_s), \tag{5}$$

where x_i, $i = 1, \ldots, n$, is the ith object in high-dimensional space; n is the total number of objects; w_r, $r = 1, \ldots, m$, and w_s, $s = 1, \ldots, m$, are the prototypes of objects in the grid;

m is the total number of prototypes/neurons in the grid; h_{rs} is a neighborhood function (for example, the Gaussian kernel) that transforms non-linearly the neuron distances (see [1] for other choices of neighborhood functions); K is the number of neighbors of neurons in the SOM grid (4 or 8 in case of the rectangular grid, and 6 in case of the hexagonal grid); $d_{\text{Euc}}(\cdot, \cdot)$ is the Euclidean distance; and V_r is the Voronoi region corresponding to prototype w_r. The Voronoi region is defined as the set of neurons closest to a given neuron (its neighbors) in the SOM lattice. The number of prototypes is assumed to be sufficiently large so that $d^2_{\text{Euc}}(x_i, w_s) \approx d^2_{\text{Euc}}(x_i, w_r) + d^2_{\text{Euc}}(w_r, w_s)$.

In order to formulate the asymmetric version of the SOM algorithm, we will refer to the error function (4).

The asymmetric SOM algorithm is derived in three steps:

Step 1. Transform the Euclidean distance into a similarity:

$$s_{is}^{\text{SYM}} = C - d^2_{\text{Euc}}(x_i, w_s), \tag{6}$$

where C is a constant being the upper boundary of the squared Euclidean distance over the entire dataset, and the rest of notation is described in (4).

Step 2. Transform the symmetric similarity into the asymmetric similarity:

$$s_{is}^{\text{ASYM}} = a_i \left(C - d^2_{\text{Euc}}(x_i, w_s)\right), \tag{7}$$

where a_i is the asymmetric coefficient defined in Sect. 3.1, in (1), and the rest of notation is described in (6). The asymmetric similarity defined in this way, using the asymmetric coefficient, guarantees the consistency with the asymmetric hierarchical associations among objects in the dataset.

Step 3. Insert the asymmetric similarity in the error function (4), in order to obtain the energy function, which needs to maximized:

$$E = \sum_r \sum_{x_i \in V_r} \sum_s h_{rs} a_i \left(C - d^2_{\text{Euc}}(x_i, w_s)\right), \tag{8}$$

where the notation is explained in (4), (6), and (7). The energy function (8) can be optimized in the similar way as the error function (4). For detailed information, see [38] or [11].

6 Asymmetric Clustering of Asymmetric Self-Organizing Map

In order to obtain a methodological consistency in data analysis, the asymmetric SOM, discussed in Sect. 5, is clustered by means of the asymmetric clustering approach. We have chosen the improved version of the asymmetric k-means algorithm for this purpose. This choice was motivated by the fact that the traditional version of the k-means method is well-known as a very effective and efficient clustering technique, and the asymmetric form of this algorithm assures the consistency with the asymmetric version of the SOM.

The neurons in the grid of the asymmetric SOM are used as input for the asymmetric k-means algorithm. In other words, the clustering process is carried out in the output space of the asymmetric SOM, i.e., in 2-dimensional space.

The improved asymmetric k-means clustering algorithm consists of two alternating steps:

Step 1. Forming of the clusters: The algorithm iterates over the entire set of neurons (each neuron is a point in 2-dimensional output SOM space), and allocates each neuron to the cluster represented by the centroid—nearest to this neuron. The nearest centroid

is determined using a chosen similarity measure. Hence, for each neuron v_r, $r = 1, \ldots, m$ in an analyzed SOM, the following maximal squared similarity has to be found:

$$\max_j \ \eta_j s^2 \left(v_r, c_j\right), \tag{9}$$

where $s(\cdot, \cdot)$ is a chosen similarity measure; η_j, $j = 1, \ldots, k$, is the cluster coefficient defined in Sect. 3.2, in (3); c_j, $j = 1, \ldots, k$, is the centroid of the jth cluster; m is the total number of neurons in the grid; and k is the number of clusters.

Step 2. Finding centroids for the clusters: For each cluster, a centroid is determined on the basis of neurons belonging to this cluster. The algorithm calculates centroids of the clusters so as to maximize the given energy function:

$$E\left(\Upsilon_j\right) = \sum_{z=1}^{n_j} a_z \eta_j s^2 \left(v_z, c_j\right), \tag{10}$$

where Υ_j, $j = 1, \ldots, k$, is the jth cluster; n_j, $j = 1, \ldots, k$, is the number of neurons in the jth cluster; and the rest of notation is described in (9).

Both these steps must be carried out with the same similarity measure, in order to guarantee the monotone property of the k-means algorithm.

Steps 1 and 2 have to be repeated until the termination condition is met. The termination condition might be either reaching convergence of the iterative application of the function (11), or reaching the pre-defined number of cycles.

After each cycle (Steps 1 and 2, the value of the energy function (11) needs to be computed for the entire analyzed dataset, in order to track the convergence of the whole clustering process:

$$E(\Upsilon) = \sum_{j=1}^{k} \sum_{z=1}^{n_j} a_z \eta_j s^2 \left(v_z, c_j\right), \tag{11}$$

where Υ is the whole set of SOM neurons, and the rest of the notation is described in (9).

Unfortunately, also the well-known drawback of the standard symmetric k-means algorithm regarding the uncertainty of its convergence process still holds. Likewise the traditional technique, the proposed method does not assure the convergence to globally optimal solution. Random initialization of the considered method is another major issue. The algorithm needs to be multistarted with random starts in order to return a reasonable solution.

6.1 Computational Complexity

The additional computational demand associated with the approach proposed in this paper comes down to the computational cost of the function $\varrho(\cdot)$ in (2). Taking into consideration that pairwise Euclidean distances between all data objects need to be computed, the resulting estimated computational complexity of the function $\varrho(\cdot)$ is $\mathcal{O}\left(n^2\right)$, where n is the total number of objects in an analyzed dataset. Consequently, considering the complexity of the traditional SOM is $\mathcal{O}(n)$, and the complexity of the traditional symmetric k-means algorithm is also $\mathcal{O}(n)$, the entire approach proposed in this paper is characterized by the quadratic estimated computational complexity $\mathcal{O}\left(n^2\right)$.

7 Experiments

Our experimental study aims to confirm that clustering of asymmetric SOM by means of the improved asymmetric k-means algorithm is superior over seven reference methods, i.e., traditional symmetric k-means clustering algorithm, Gaussian-Mixture-Model-based (GMM-based) clustering method, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) approach, clustering of the traditional symmetric SOM using the traditional symmetric k-means method, clustering of asymmetric SOM using the classical symmetric k-means algorithm, clustering of asymmetric SOM using the GMM-based clustering technique, and over clustering of asymmetric SOM using the DBSCAN approach.

The first three reference methods are well-known state-of-the-art techniques, whereas the remaining four comparison methods are combinations of visualization & clustering. In this way, our experimental research evaluates the proposed approach competing with pure clustering methods and two-stage visualization and clustering techniques.

The experiments have been carried out on real data in the four different research fields: in the field of words clustering, in the field of individual household electric power consumption data clustering, in the field of sound signals clustering, and in the field of human heart rhythm signals clustering. The first two parts of the experimental study were conducted on the large dataset (Sects. 7.5, 7.6), while the remaining two experimental parts were carried out on smaller datasets (Sects. 7.7, 7.8). In this way, one can assess the performance of the investigated methods operating on datasets of different size and nature, and consequently, one can better evaluate the effectiveness of the proposed approach.

The words clustering experiment was conducted on the "Bag of Words" dataset, while the electric power consumption data clustering was carried out on the "Individual Household Electric Power Consumption" dataset, both from the UCI Machine Learning Repository [39].

The sound signals clustering was carried out on the piano music recordings, and the human heart rhythm signals clustering was conducted using the ECG recordings derived from the MIT-BIH ECG Databases [40].

In case of the piano music dataset and the ECG recordings dataset, a graphical illustration of the U-matrices generated by SOM is provided, while in case of the "Bag of Words" dataset and "Individual Household Electric Power Consumption" dataset, no such illustration is given, because of the high number of instances in those datasets, which would make such images unreadable.

Each of the examined approaches was investigated in 10-fold cross-validation setup, in order to avoid the problem of overfitting during the training phase of the SOM technique (in both: symmetric and asymmetric versions).

7.1 Evaluation Criteria

In case of all four parts of our experiments, we have compared the clustering results obtained with use of the investigated methods. As the basis of the comparisons, i.e., as the clustering evaluation criteria, we have used the accuracy degree [11,18], and the uncertainty degree [10, 11].

In general, the results of clustering can be assessed with use of two groups of evaluation criteria [41] (also called as the validity indeces). First group are the external criteria, which are computed using the ground knowledge about the clustered data, i.e., what should be the correct result of clustering. These criteria are much easier to formulate, and they allow for the precise assessment of the clustering results, however, they are useless in real-world clustering problems, where no additional information about the analyzed data is available. Second group

are the internal criteria, which are computed without using the ground knowledge about the clustered data. Formulation of these criteria is, naturally, more difficult, however, they can be employed to assess the results of clustering in real-life problems. Therefore, their usefulness in data analysis is of much value. More information on the issue of clustering output assessment can be found, for example, in [41] and in [42].

In case of all four parts of our empirical study, the ground knowledge about the data was known. Therefore, an application of external evaluation criterion (the accuracy degree) was possible. In order to provide more reliable assessment of the clustering results, we have used also the second internal evaluation criterion (the uncertainty degree).

Hence, the following two evaluation criteria have been used:

7.1.1 Accuracy Degree

This evaluation criterion determines the number of correctly assigned objects divided by the total number of objects.

Hence, for the ith cluster, the accuracy degree is determined as follows:

$$q_i = \frac{c_i}{n_i}, \tag{12}$$

where c_i, $i = 1, \ldots, k$, is the number of objects correctly assigned to the ith cluster; n_i, $i = 1, \ldots, k$, is the number of objects in the ith cluster ("gold standard"); and k is the number of clusters.

And, for the entire dataset, the total accuracy degree is determined as follows:

$$q_{\text{total}} = \frac{c}{n}, \tag{13}$$

where c is the total number of correctly assigned objects, and n is the total number of objects in the entire dataset.

The accuracy degrees q_i and the total accuracy degree q_{total} assume values in the interval $\langle 0, 1 \rangle$, and naturally, greater values are preferred.

The total accuracy degree q_{total} was used in our experimental study as the main basis of the clustering accuracy comparison of the eight investigated approaches.

7.1.2 Uncertainty Degree

This evaluation criterion determines the number of overlapping objects divided by the total number of objects in the dataset. This means, the number of objects, which are in the overlapping area between clusters, divided by the total number of objects. The objects belonging to the overlapping area are determined on the basis of the ratio of dissimilarities between them and the two nearest clusters centroids. If this ratio is in the interval $\langle \beta_1, \beta_2 \rangle$, then the corresponding object is said to be in the overlapping area. In other words, the objects in the overlapping area are more likely to be assigned to incorrect clusters than other objects, and therefore, their number should be minimized. One can also say that the uncertainty degree determines the uncertainty of the assignments of objects to correct clusters. The values of the boundary parameters β_1 and β_2 have been tuned as it is described in Sect. 7.2.

The uncertainty degree is determined as follows:

$$U_{\text{d}} = \frac{\mu}{n}, \tag{14}$$

where μ is the number of overlapping objects in the dataset, and n is the total number of objects in the dataset.

The uncertainty degree assumes values in the interval $\langle 0, 1 \rangle$, and smaller values are desired.

Because of the tenfold cross-validation setup of our experiments, the average accuracies and uncertainty degrees were calculated in order to obtain reliable results.

7.2 Parameters Tuning

In the experimental study, the values of the parameters τ, β_1, and β_2 have been set empirically on the basis of the preliminary set of experiments regarding all the investigated methods in the 10-fold cross-validation setup (on the "Bag of Words" dataset) so as to satisfy the following requirements. In case of the value of the parameter τ, our goal was to maximize the performance of the proposed method, and the resulting value of τ was set as 0.45. On the other hand, in case of the values of the parameters β_1 and β_2 of the uncertainty degree defined in (14), we aimed to detect the area between clusters, in which the likelihood of assigning objects to incorrect clusters is dangerously high. Precisely, after setting $\beta_1 = 0.86$ and $\beta_2 = 1.14$ (chosen empirically after a sequence of attempts), in each of the cross-validation experimental repetitions, not less than 95 % of the objects in the resulting overlapping area have been assigned to incorrect clusters. Consequently, the values of the boundary parameters β_1 and β_2 were set to: $\beta_1 = 0.86$ and $\beta_2 = 1.14$, and the obtained interval was $\langle 0.86, 1.14 \rangle$.

As the neighborhood function h_{rs}, we have chosen the Gaussian kernel of the width determined in the standard way, as it is described, for example, in [11].

The value of the parameter K has been set to 6, because the hexagonal SOM lattice has been used in all experiments, we have carried out.

7.3 Statistical Significance

In case of all comparisons between the proposed approach and the reference methods, the statistical significance has been verified on the basis of the statistical Student's t-test, this way confirming that the difference in the results produced by a pair of evaluated approaches is statistically significant. The p values calculated in case of each comparison are reported in Tables 1, 5, and 7. Each p value corresponding to a given comparison method should be referenced to the proposed method. The p values computed in case of each comparison are lower than the significance level $\alpha = 0.001$, which indicates the high statistical significance of the obtained empirical results. Therefore, the Student's t-test confirmed the reliability and high statistical significance of the conducted experimental research.

7.4 Text Feature Extraction

Feature extraction of the textual data investigated in this part of our experimental study was carried out using the term frequency—inverse document frequency (*tf-idf*) approach. The vector space model (VSM) constructed this way is particularly useful in our research, because it implicitly captures the terms frequency (both: local—document-dependent and global—collection-dependent), which are the source of the hierarchy-based asymmetric relationships in analyzed data (i.e., in this case, between words).

The dimensionality of the analyzed VSM model (i.e., the number of features) was chosen as the minimal length of the vocabularies in the five considered text collections. Consequently, the number of features utilized in this part of our experimental study was 6906. It was necessary to truncate the longer vocabularies in order to build the data matrix comprising the

analyzed VSM model. As a result, not all of the words in the remaining four text collections have been taken into account. Nevertheless, the considered experimental problem remains a high-dimensionality issue, and the number and variety of the words in the analyzed vocabularies makes the problem complex and challenging. Of course, also the highly asymmetric nature of the investigated dataset is preserved.

7.5 Words Visualization and Clustering

In the first part of our experimental study, we have utilized the "Bag of Words" dataset from the UCI Machine Learning Repository [39].

It is a high-dimensional dataset of strongly asymmetric nature, especially useful in case of the asymmetric data relationships analysis. It is so, because of the significant differences in frequencies of occurrences of different words in the entire dataset. Therefore, the experimental investigation on the "Bag of Words" dataset clearly shows the superiority of the proposed fully asymmetric approach over the other examined techniques.

7.5.1 Dataset Description

The "Bag of Words" dataset consists of five text collections:

- *Enron E-mail Collection* This collection was prepared by the A Cognitive Assistant that Learns and Organizes Project (CALO). It contains data from about 150 users, mostly senior management of Enron, organized into folders. The number of documents in this collection is 39,861, the number of words in the vocabulary is 28,102, and the total number of words is approximately 6,400,000.
- *Neural Information Processing Systems (NIPS) full papers* The number of documents in this collection is 1500, the number of words in the vocabulary is 12,419, and the total number of words is approximately 1,900,000.
- *Daily KOS Blog Entries* The number of documents in this collection is 3430, the number of words in the vocabulary is 6906, and the total number of words is approximately 468,000.
- *New York Times News Articles* The number of documents in this collection is 300,000, the number of words in the vocabulary is 102,660, and the total number of words is approximately 100,000,000. We have utilized only an excerpt from this collection, i.e., 3000 documents, 11,203 words in the vocabulary, and approximately 2,000,000 of words in total.
- *PubMed Abstracts* This is the collections of abstracts of the U.S. National Library of Medicine, National Institute of Health. The number of documents in this collection is 8,200,000, the number of words in the vocabulary is 141,043, and the total number of words is approximately 730,000,000. We have utilized only an excerpt from this collection, i.e., 1000 documents, 8520 words in the vocabulary, and approximately 100,000 of words in total.

The total number of analyzed words was approximately 10,868,000.

The investigated methods were forming five clusters representing those five text collections in the "Bag of Words" dataset.

7.5.2 Experimental Results

The results of this part of our experiments are reported in Table 1 and in Table 2. Table 1 presents the accuracy degrees, standard deviations, and p values (of the statistical t-test),

Table 1 Accuracy degrees, standard deviations, and p values for the words visualization and clustering

Investigated approach	q_{total}	s (%)	p value
Symmetric k-means	0.7454	0.0015	$<10^{-4}$
GMM-based clustering	0.7497	0.0024	$<10^{-4}$
DBSCAN	0.7581	0.0012	$<10^{-4}$
Symmetric SOM and symmetric k-means	0.7712	0.0005	$<10^{-4}$
Asymmetric SOM and symmetric k-means	0.8234	0.0021	$<10^{-4}$
Asymmetric SOM and GMM-based clustering	0.8381	0.0019	$<10^{-4}$
Asymmetric SOM and DBSCAN	0.8525	0.0018	$<10^{-4}$
Asymmetric SOM and asymmetric k-means	0.9035	0.0014	

Table 2 Uncertainty degrees for the words visualization and clustering

Investigated approach	U_{d}
Symmetric k-means	0.2587
GMM-based clustering	0.2538
DBSCAN	0.2473
Symmetric SOM and symmetric k-means	0.2119
Asymmetric SOM and symmetric k-means	0.1801
Asymmetric SOM and GMM-based clustering	0.1782
Asymmetric SOM and DBSCAN	0.1721
Asymmetric SOM and asymmetric k-means	0.1073

while Table 2 reports the uncertainty degrees corresponding to each of the investigated approaches.

The results of this part of our experimental study show that clustering of the asymmetric SOM using the asymmetric k-means algorithm outperforms the seven comparison methods. The proposed approach leads to the higher clustering accuracy measured on the basis of the total accuracy degree, and also to the lower clustering uncertainty measured on the basis of the uncertainty degree.

7.6 Household Electric Power Consumption Data Visualization and Clustering

In the first part of our experimental study, we have utilized the "Individual Household Electric Power Consumption" dataset from the UCI Machine Learning Repository [39].

7.6.1 Dataset Description

The dataset consists of measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Precisely, the data collection was gathered between December, 2006 and November, 2010 (47 months). Different electrical quantities and some sub-metering values were utilized.

The number of instances (i.e., electrical measurements) in this dataset is 2,075,259.

The following features were used: date; time; global active power (household global minute-averaged active power in kilowatt); global reactive power (household global minute-averaged reactive power in kilowatt); voltage (minute-averaged voltage in volt); global intensity (household global minute-averaged current intensity in ampere); sub-metering 1 (energy sub-metering No. 1 in watt-hour of active energy corresponding to the kitchen, containing mainly a dishwasher, an oven, and a microwave); sub-metering 2 (energy sub-metering No. 2 in watt-hour of active energy corresponding to the laundry room, containing a washing-machine, a tumble-drier, a refrigerator, and a light); sub-metering 3 (energy sub-metering No. 3 in watt-hour of active energy corresponding to an electric water-heater and an air-conditioner).

The total number of features was nine, and therefore, no feature extraction was necessary.

The investigated methods were forming four clusters representing each of the 4 years in the dataset, because according to our experimental study, the electrical measurements indicate certain differences in the electrical power consumption between each of the four years.

7.6.2 Experimental Results

The results of this part of our experiments are reported in Tables 3 and 4. Table 3 presents the accuracy degrees, standard deviations, and p values (of the statistical t-test), whereas Table 4 reports the uncertainty degrees corresponding to each of the investigated approaches.

Table 3 Accuracy degrees, standard deviations, and p values for the household power consumption data visualization and clustering

Investigated approach	q_{total}	s (%)	p value
Symmetric k-means	0.7093	0.0473	$<10^{-4}$
GMM-based clustering	0.7121	0.0120	$<10^{-4}$
DBSCAN	0.7196	0.0959	$<10^{-4}$
Symmetric SOM & symmetric k-means	0.7805	0.0758	$<10^{-4}$
Asymmetric SOM & symmetric k-means	0.8474	0.0042	$<10^{-4}$
Asymmetric SOM & GMM-based clustering	0.8668	0.0553	$<10^{-4}$
Asymmetric SOM & DBSCAN	0.8859	0.0759	$<10^{-4}$
Asymmetric SOM & asymmetric k-means	0.9211	0.0149	

Table 4 Uncertainty degrees for the household power consumption data visualization and clustering

Investigated approach	U_{d}
Symmetric k-means	0.3095
GMM-based clustering	0.3071
DBSCAN	0.2857
Symmetric SOM & symmetric k-means	0.2519
Asymmetric SOM & symmetric k-means	0.1602
Asymmetric SOM & GMM-based clustering	0.1598
Asymmetric SOM & DBSCAN	0.1127
Asymmetric SOM & asymmetric k-means	0.0807

In case of this dataset, the proposed approach also defeated all the reference methods, regardless if it was pure clustering or the combination of visualization and clustering. The observation was based on the obtained values of the two considered evaluation criteria, i.e., the total accuracy degree and the uncertainty degree.

7.7 Piano Music Composers Visualization and Clustering

In this part of our experiments, the investigated methods were forming three clusters representing three piano music composers: Johann Sebastian Bach, Ludwig van Beethoven, and Fryderyk Chopin.

7.7.1 Dataset Description

Each music piece was represented by a 30-s sound signal sampled with the 44,100 Hz frequency. The entire dataset consisted of 32 sound signals. Feature extraction process was carried out according to the Discrete-Fourier-Transform-based (DFT-based) method described in Appendix.

7.7.2 Experimental Results

The results of this part of our experiments are demonstrated in Fig. 1, and in Tables 5 and 6. Figure 1 presents the maps (U-matrices) generated by the symmetric (Fig. 1a) and asymmetric (Fig. 1b) SOM techniques. The U-matrix is a graphical presentation of SOM. Each entry of the U-matrix corresponds to a neuron in the SOM grid, while value of that entry is the average dissimilarity between the weight vector of the neuron (prototype in the SOM grid) and the weight vectors of its neighbors. Table 5, in turn, presents the accuracy degrees, standard deviations, and p values (of the statistical t-test), while Table 6 reports the uncertainty degrees corresponding to each of the examined approaches.

The issue of comparison between the symmetric and the asymmetric SOM and their mutual relationships has been elaborated in detail in [10] and in [11], where the asymmetric SOM

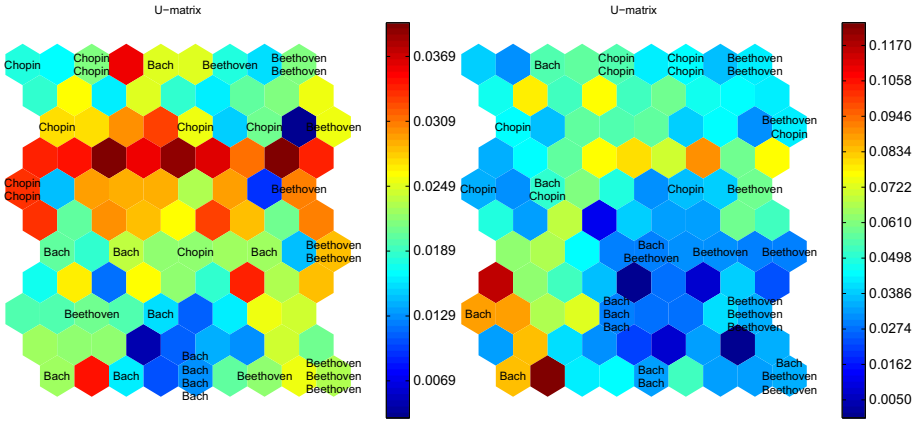

Fig. 1 Piano music composers maps (U-matrices). **a** Symmetric SOM **b** asymmetric SOM

Table 5 Accuracy degrees, standard deviations, and p values for the piano music composers visualization and clustering

Investigated approach	q_{total}	s (%)	p value
Symmetric k-means	23/32 = 0.7188	1.5417	$<10^{-4}$
GMM-based clustering	24/32 = 0.7500	2.8610	$<10^{-4}$
DBSCAN	24/32 = 0.7500	2.3743	$<10^{-4}$
Symmetric SOM & symmetric k-means	26/32 = 0.8125	2.6557	$<10^{-4}$
Asymmetric SOM & symmetric k-means	29/32 = 0.9063	1.6470	$<10^{-4}$
Asymmetric SOM & GMM-based clustering	29/32 = 0.9063	1.6137	$<10^{-4}$
Asymmetric SOM & DBSCAN	30/32 = 0.9375	1.5095	$<10^{-4}$
Asymmetric SOM & asymmetric k-means	32/32 = 1.0000	2.2097	

Table 6 Uncertainty degrees for the piano music composers visualization and clustering

Investigated approach	U_{d}
Symmetric k-means	8/32 = 0.2500
GMM-based clustering	7/32 = 0.2188
DBSCAN	7/32 = 0.2188
Symmetric SOM & symmetric k-means	8/32 = 0.2500
Asymmetric SOM & symmetric k-means	6/32 = 0.1875
Asymmetric SOM & GMM-based clustering	6/32 = 0.1875
Asymmetric SOM & DBSCAN	4/32 = 0.1250
Asymmetric SOM & asymmetric k-means	1/32 = 0.0313

version has been justified, analyzed, and experimentally verified in a variety of different scenarios using different datasets. Now, in the present paper, we focus on the idea of cooperation between the asymmetric SOM and the improved asymmetric k-means algorithm, introduced in this paper.

Based only on images presented in Fig. 1, it would be difficult to assert the superiority either of the symmetric or the asymmetric SOM, because the images present the data visualizations, which need to be further evaluated (or subsequently analyzed) in order to formulate any judgments regarding the performance of the symmetric and the asymmetric SOM. In our research, we conduct the evaluation on the basis of the clustering results assessment using the evaluation criteria described in Sect. 7.1.

The size of the constructed SOM was 11×9 neurons. The number of clusters in the k-means clustering was set to 3.

Also in this part of our experiments, the proposed combination of the asymmetric SOM and the asymmetric k-means clustering appeared to be superior over the other seven investigated data analysis approaches.

7.8 Human Heart Rhythms Visualization and Clustering

The human heart rhythm signals clustering experiment was carried out on the dataset of ECG recordings derived from the MIT-BIH ECG Databases [40].

In this part of our experiments, the investigated methods were forming three clusters representing three types of human heart rhythms: normal sinus rhythm, atrial arrhythmia, and ventricular arrhythmia. This kind of clustering can be interpreted as the cardiac arrhythmia detection and recognition based on the ECG recordings.

7.8.1 Dataset Description

In general, the cardiac arrhythmia disease may be classified either by rate (tachycardias—the heart beat is too fast, and bradycardias—the heart beat is too slow) or by site of origin (atrial arrhythmias—they begin in the atria, and ventricular arrhythmias—they begin in the ventricles). Our clustering recognizes the normal rhythm, and also, recognizes arrhythmias originating in the atria, and in the ventricles.

We analyzed 20-min ECG holter recordings sampled with the 250 Hz frequency. The entire dataset consisted of 63 ECG signals. Feature extraction was carried out according to the DFT-based method described in Appendix.

7.8.2 Experimental Results

The results of this part of our experiments are presented in Fig. 2, and in Tables 7 and 8, which are constructed in the same way as in Sect. 7.7.

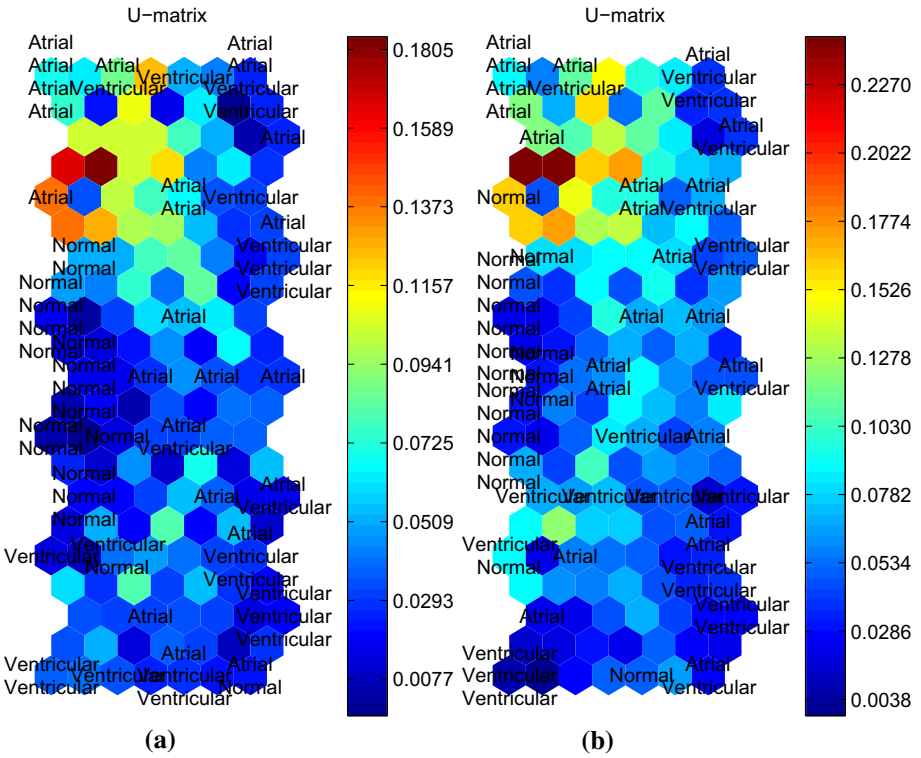

Fig. 2 Human heart rhythms maps (U-matrices). **a** Symmetric SOM **b** asymmetric SOM

Table 7 Accuracy degrees, standard deviations, and p values for the human heart rhythms visualization and clustering

Investigated approach	q_{total}	s (%)	p value
Symmetric k-means	40/63 = 0.6349	1.0775	$<10^{-4}$
GMM-based clustering	41/63 = 0.6508	1.9253	$<10^{-4}$
DBSCAN	41/63 = 0.6508	2.8459	$<10^{-4}$
Symmetric SOM & symmetric k-means	44/63 = 0.6984	1.7468	$<10^{-4}$
Asymmetric SOM & symmetric k-means	49/63 = 0.7778	2.0898	$<10^{-4}$
Asymmetric SOM & GMM-based clustering	49/63 = 0.7778	2.3002	$<10^{-4}$
Asymmetric SOM & DBSCAN	52/63 = 0.8254	2.2757	$<10^{-4}$
Asymmetric SOM & asymmetric k-means	58/63 = 0.9206	1.8329	

Table 8 Uncertainty degrees for the human heart rhythms visualization and clustering

Investigated approach	U_{d}
Symmetric k-means	21/63 = 0.3333
GMM-based clustering	21/63 = 0.3333
DBSCAN	23/63 = 0.3651
Symmetric SOM & symmetric k-means	20/63 = 0.3175
Asymmetric SOM & symmetric k-means	16/63 = 0.2540
Asymmetric SOM & GMM-based clustering	15/63 = 0.2381
Asymmetric SOM & DBSCAN	12/63 = 0.1905
Asymmetric SOM & asymmetric k-means	10/63 = 0.1587

The size of the constructed SOM was 21×7 neurons. The number of clusters in the k-means clustering was set to 3.

Finally, in the last part of our empirical study, the proposed marriage of the asymmetric SOM and the asymmetric k-means clustering produced results superior over the results returned by the seven reference methods, confirming the usefulness and effectiveness of the proposed solution.

8 Discussion on the Experimental Results

The experimental evaluation of the investigated methods concerned four different datasets containing real data. The datasets were of different size and nature in order to provide an extensive and reliable verification of the usefulness of the approach proposed in this paper on the basis of the comparison with the seven employed reference techniques.

In case of all four datasets, our method produced the results superior with respect to the results returned by the comparison methods. Such an observation can lead to a conclusion that the asymmetric relationships in data occurred noticeably in all analyzed datasets. Furthermore, the enhancement to the asymmetric k-means clustering algorithm introduced in this paper appeared successful, because it properly grasps one of the main properties of the algorithm (i.e., representing clusters by their centroids) by distinguishing the cluster centroids and normal data objects.

Taking into account the fact that the "Bag of Words" dataset was substantially larger than the remaining two datasets, and that the variances in that dataset were very low (significantly lower than in the remaining two datasets), the experimental results corresponding to that dataset can be considered as especially important and meaningful. The absolute values of the numbers of words assigned to correct clusters are in this case definitely different for all the examined methods, which vouches for the superiority of the proposed approach over the other investigated techniques. And this result can be interpreted as fact confirming the strongly asymmetric nature of the "Bag of Words" dataset, and generally, highly asymmetric relationships in any textual datasets. The explanation of this observation is that in case of textual data of big size (consisting of high number of words), the differences in frequencies of occurrences of different words are especially significant, and the resulting asymmetry phenomenon is especially noticeable. The choice of the VSM model representing the textual datasets is highly recommended, when it comes to the asymmetric data analysis approach, because it implicitly captures the frequency information regarding the words in the datasets, which is the source of the asymmetric data relationships. In this way, the asymmetric properties of data are preserved, and fully taken into account during any subsequent analysis, like data visualization and clustering in case of our research.

9 Summary and Future Research

In this paper, the two-stage data analysis approach was proposed. The first step consisted in data visualization by means of the asymmetric SOM, while in the second step, the asymmetric SOM was clustered using the asymmetric k-means algorithm. This kind of combination assures that in both these steps, the asymmetric relationships in data will be taken into account and properly handled by both methods. In this way, the introduced approach maintains the methodological consistency of the entire analysis.

Our experiments were carried out on the four datasets: "Bag of Words" dataset, "Individual Household Electric Power Consumption" dataset, piano music dataset, and human heart rhythms dataset. The results of the conducted empirical research confirmed the superiority of the proposed fully asymmetric approach over the three well-known state-of-the-art clustering methods, i.e., the traditional k-means clustering algorithm, the GMM-based clustering method, and the DBSCAN technique; and the following two-stage combinations: symmetric SOM & symmetric k-means, asymmetric SOM & symmetric k-means, asymmetric SOM & GMM-based clustering, and asymmetric SOM & DBSCAN.

Possible directions of future research may concern formulating the asymmetric versions of some different clustering techniques, which can be used to perform clustering on the asymmetric SOM. One can consider either the group of the hierarchical clustering methods or the density-based clustering techniques or, finally, the model-based clustering approaches. That will also result in obtaining a fully asymmetric combination of data analysis methods. Especially interesting may be designing the asymmetric versions of the hierarchical clustering techniques, since the asymmetry in our work originated from the hierarchical relationships in data. From this point of view, the analysis of the hierarchical clustering methods is highly recommended and potentially beneficial.

Another idea may be a more general extension of this paper's proposal leading to utilizing different than SOM visualization techniques, and designing their combinations with various asymmetric clustering algorithms, like, for example, those mentioned beforehand.

Appendix: Time Series Feature Extraction

In this appendix, we describe the feature extraction procedure of the time series employed in our research.

Feature discovery process preceding the actual visualization and clustering is an important stage of data pre-processing. It has a strong impact on the final accuracy of visualization and clustering, and consequently, on the performance of the whole analysis. Feature discovery aims to form possibly smallest set of most relevant, informative, and discriminative features. A proper choice of the feature set results in higher visualization and clustering quality.

Features of the time series considered in Sects. 7.7 and 7.8 have been extracted using a method based on the discrete Fourier transform (DFT), which proceeds according to Procedure 1. In general, the DFT-based feature extraction is described, for example, in [43].

Procedure 1 *Features of analyzed signals are retrieved according to the following procedure:*

Step 1. *Separate N time intervals from the discrete-time signal representation. All intervals are of the same length, hence all have the same number of samples. As a result, one obtains N functions in the discrete-time domain: $f_i\left(t_j\right)$, $i = 1, \ldots, N$, $j = 1, \ldots, \widetilde{K}$, where $\widetilde{K}$ is the number of samples in each separated time interval.*

Step 2. *Perform the DFT on each of the N discrete-time functions obtained in the Step 1, considering the absolute values of the complex DFT-vectors entries. As a result, one obtains N functions in the discrete-frequency domain: $\left|\widetilde{f}_i\left(\omega_l\right)\right| = \left|\mathcal{F}\left(f_i\right)\left(\omega_l\right)\right|$, $i = 1, \ldots, N$, $l = 1, \ldots, \text{floor}\left(\frac{\widetilde{K}}{2}\right)$ (the rest of the DFT result is mirrored—it does not contain any new information), where* floor $(\cdot)$ *returns the largest integer that is less than or equal to the argument.*

Step 3. *Calculate the average DFT result on the basis of the results of DFT for each of the intervals. The average DFT result denotes the DFT-vector, which entries are the arithmetic averages—calculated on the basis of the corresponding entries of the partial DFT-vectors.*

$$\left|\widetilde{f}_{\text{AVG}}\left(\omega_l\right)\right| = \frac{1}{N}\sum_{i=1}^{N}\left|\widetilde{f}_i\left(\omega_l\right)\right| \tag{15}$$

where $i = 1, \ldots, N$, $l = 1, \ldots, \text{floor}\left(\frac{\widetilde{K}}{2}\right)$.

Step 4. *Normalize the function obtained in Step 3.*

$$f_{\text{FE}}\left(\omega_l\right) = \frac{1}{\sum_{p=1}^{r}\left|\widetilde{f}_{\text{AVG}}\left(\omega_p\right)\right|}\left|\widetilde{f}_{\text{AVG}}\left(\omega_l\right)\right|, \tag{16}$$

where $f_{\text{FE}}\left(\omega_l\right)$ is the function representing the set of the obtained features, $l = 1, \ldots, r$, and $r = \text{floor}\left(\frac{\widetilde{K}}{2}\right)$.

As the final result of Procedure 1, one obtains the discrete function f_{FE} representing the retrieved vector of features of a single signal.

The number N of the time intervals is set arbitrarily. We provide no principled way to determine it, which can be regarded as a drawback of this feature extraction method.

Normalization in Step 4 of Procedure 1 is a common practice in pattern recognition. The normalized features are of benefit in many contexts of multivariate analysis, not only in clustering, but also, for example, in discriminant analysis. Normalization of features especially accounts in the field of sound recognition, which was one of the areas of our experiments. It filters out the irrelevant feature of loudness (feature of loudness should not affect the results of recognition—a music piece played softly remains the same piece if it is played loudly, but without normalization, the features would change). In other words, Step 4 determines the relative intensities as the characteristic feature set, and not the absolute values, which would be largely influenced by the irrelevant features, like levels of loudness in case of sound recognition.

References

1. Kohonen T (2001) Self-organizing maps, 3rd edn. Springer, New York
2. Kohonen T (2013) Essentials of the self-organizing map. Neural Netw 37:52–65
3. Olszewski D (2014) Fraud detection using self-organizing map visualizing the user profiles. Knowl-Based Syst 70:324–334
4. Olszewski D (2014) An improved adaptive self-organizing map. In: Rutkowski L, Korytkowski M, Scherer R, Tadeusiewicz R, Zadeh L, Zurada J (eds) Artificial intelligence and soft computing (Lecture notes in computer science), vol 8467. Springer, New York, pp 109–120
5. Piastra M (2013) Self-organizing adaptive map: autonomous learning of curves and surfaces from point samples. Neural Netw 41:96–112
6. Płński P, Zaremba K (2014) Visualizing random forest with self-organising map. In: Rutkowski L, Korytkowski M, Scherer R, Tadeusiewicz R, Zadeh L, Zurada J (eds) Artificial intelligence and soft computing (Lecture notes in computer science), vol. 8468, pp 63–71
7. Segev A, Kantola J (2012) Identification of trends from patents using self-organizing maps. Expert Syst Appl 39(18):13235–13242
8. Kohonen T (1990) The self-organizing map. In: Proceedings of the IEEE, vol. 28. pp 1464–1480
9. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43(1):59–69
10. Martín-Merino M, Muñoz A (2005) Visualizing asymmetric proximities with SOM and MDS models. Neurocomputing 63:171–192
11. Olszewski D (2011) An experimental study on asymmetric self-organizing map. In: Yin H, Wang W, Rayward-Smith V (eds) Intelligent data engineering and automated learning—IDEAL 2011 (Lecture notes in computer science), vol. 6936, pp 42–49
12. Biau G, Devroye L, Lugosi G (2008) On the performance of clustering in Hilbert spaces. IEEE Trans Inf Theory 54(2):781–790
13. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY (2002) An efficient k-means clustering algorithm: analysis and implemetation. IEEE Trans Pattern Anal Mach Intell 24(7):881–892
14. Laszlo M, Mukherjee S (2006) A genetic algorithm using hyper-quadtrees for low-dimensional K-means clustering. IEEE Trans Pattern Anal Mach Intell 28(4):533–543
15. MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, pp 281–297
16. Steinhaus H (1956) Sur la Division des Corp Matériels en Parties. Bulletin de l'Académie Polonaise des Sciences, C1. III 4(12):801–804
17. Xiong H, Wu J, Chen J (2009) K-means clustering versus validation measure: a data-distribution perspective. IEEE Trans Syst Man Cybern-Part B 39(2):318–331
18. Olszewski D (2011) Asymmetric k-means algorithm. In: Dobnikar A, Lotrič U, Šter B (eds) Adaptive and natural computing algorithms (Lecture notes in computer science), vol. 6594, pp 1–10
19. Olszewski D (2012) k-means clustering of asymmetric data. In: Corchado E, Snášel V, Abraham A, Woźniak M, Grana M, Cho SB (eds) Hybrid artificial intelligent systems (Lecture notes in computer science), vol. 7208, pp 243–254
20. Tversky A (1977) Features of similarity. Psychol Rev 84(4):327–352

21. Tversky A (2004) Preference, belief, and similarity (selected writings). A Bradford book. The MIT Press, Cambridge
22. Muñoz A, Martín-Merino M (2002) New asymmetric iterative scaling models for the generation of textual word maps. In: Proceedings of the international conference on textual data statistical analysis JADT'02, pp 593–603
23. de Diego IM, Muñoz A, Moguerza JM (2010) Methods for the combination of kernel matrices within a support vector framework. Mach Learn 78:137–174
24. Holman EW (1979) Monotonic models for asymmetric proximities. J Math Psychol 20(1):1–15
25. Weeks DG, Bentler PM (1982) Restricted multidimensional scaling models for asymmetric proximities. Psychometrika 47(2):201–208
26. Zielman B (1991) Three-way scaling of asymmetric proximities. Tech. Rep. Research Report RR91-01, Department of Data Theory, University of Leiden
27. Zielman B, Heiser WJ (1996) Models for asymmetric proximities. Br J Math Stat Psychol 49:127–146
28. Bove G (2010) Models for asymmetry in proximity data. Data analysis and classification, studies in classification. In: Data analysis, and knowledge organization, Springer, Berlin, pp 79–84
29. Okada A (2000) An asymmetric cluster analysis study of car switching data. In: Data analysis, studies in classification, data analysis, and knowledge organization, Springer, Berlin
30. Okada A, Imaizumi T (1997) Asymmetric multidimensional scaling of two-mode three-way proximities. J Classif 14(2):195–224
31. Okada A, Imaizumi T (2003) Joint space model for multidimensional scaling of two-mode three-way asymmetric proximities. In: Innovations in classification, data science, and information systems, studies in classification, data analysis, and knowledge organization. Springer, Berlin, pp 371–378
32. Okada A, Imaizumi T (2007) Multidimensional scaling of asymmetric proximities with a dominance point. In: Advances in data analysis, studies in classification, data analysis, and knowledge organization. Springer, Berlin, pp 307–318
33. Olszewski D, Šter B (2014) Asymmetric clustering using the alpha–beta divergence. Pattern Recognit 47(5):2031–2041
34. Martín F, Moreno L, Blanco D, Muñoz ML (2014) Kullback–Leibler divergence-based global localization for mobile robots. Robot Auton Syst 62(2):120–130
35. Olszewski D (2011) Fraud detection in telecommunications using Kullback–Leibler divergence and latent Dirichlet allocation. In: Dobnikar A, Lotrič U, Šter B (eds) Adaptive and natural computing algorithms (Lecture notes in computer science), vol. 6594, pp 71–80
36. Zeng J, Kruger U, Geluk J, Wang X, Xie L (2014) Detecting abnormal situations using the Kullback–Leibler divergence. Automatica 50(11):2777–2786
37. Muñoz A, Martin I, Moguerza JM (2003) Support vector machine classifiers for asymmetric proximities. In: ICANN (Lecture notes in computer science), vol. 2714, pp 217–224
38. Heskes T (2001) Self-organizing maps, vector quantization, and mixture modeling. IEEE Trans Neural Netw 12(6):1299–1305
39. Frank A, Asuncion A (2010) UCI machine learning repository. http://archive.ics.uci.edu/ml
40. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220. http://circ.ahajournals.org/cgi/content/full/101/23/e215. Circulation Electronic Pages
41. Halkidi M, Batistakis Y, Vazirgiannis M (2001) On clustering validation techniques. J Intell Inf Syst 17(2/3):107–145
42. Handl J, Knowles J, Kell DB (2005) Computational cluster validation in post-genomic data analysis. Bioinformatics 21(15):3201–3212
43. Chengalvarayan R, Deng L (1997) HMM-based speech recognition using state-dependent, discriminatively derived transforms on Mel-warped DFT features. IEEE Trans Speech Audio Process 2(3):243–256