Engineering Applications of Artificial Intelligence 39 (2015) 235–244

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Fuzzy clustering of time series data using dynamic time warping distance

Hesam Izakian [a,*], Witold Pedrycz [a,b,c], Iqbal Jamal [d]

[a] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2V4
[b] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[c] System Research Institute, Polish Academy of Sciences, Warsaw 00-716, Poland
[d] AQL Management Consulting Inc., Edmonton, AB, Canada

ARTICLE INFO

Article history:
Received 11 July 2014
Received in revised form
26 December 2014
Accepted 30 December 2014
Available online 17 January 2015

Keywords:
Clustering time series
Dynamic Time Warping (DTW)
Fuzzy clustering
Hybrid approach

ABSTRACT

Clustering is a powerful vehicle to reveal and visualize structure of data. When dealing with time series, selecting a suitable measure to evaluate the similarities/dissimilarities within the data becomes necessary and subsequently it exhibits a significant impact on the results of clustering. This selection should be based upon the nature of time series and the application itself. When grouping time series based on their shape information is of interest (shape-based clustering), using a Dynamic Time Warping (DTW) distance is a desirable choice. Using stretching or compressing segments of temporal data, DTW determines an optimal match between any two time series. In this way, time series exhibiting similar patterns occurring at different time periods, are considered as being similar. Although DTW is a suitable choice for comparing data with respect to their shape information, calculating the average of a collection of time series (which is required in clustering methods) based on this distance becomes a challenging problem. As the result, employing clustering techniques like K-Means and Fuzzy C-Means (where the cluster centers – prototypes are calculated through averaging the data) along with the DTW distance is a challenging task and may produce unsatisfactory results. In this study, three alternatives for fuzzy clustering of time series using DTW distance are proposed. In the first method, a DTW-based averaging technique proposed in the literature, has been applied to the Fuzzy C-Means clustering. The second method considers a Fuzzy C-Medoids clustering, while the third alternative comes as a hybrid technique, which exploits the advantages of both the Fuzzy C-Means and Fuzzy C-Medoids when clustering time series. Experimental studies are reported over a set of time series coming from the UCR time series database.

1. Introduction

Time series are commonly encountered in numerous application areas including finances, medicine, engineering, and environmental science. Considering high dimensionality and substantial volume of time series collected in different applications, extracting and visualizing available structure in this type of data is highly beneficial and exhibits numerous potential applications in data summarization, anomaly detection, etc.

In this study, we discuss and contrast a number of alternatives for fuzzy clustering of time series to reveal and visualize the available structure within this type of data. Fuzzy clustering is one of the widely used clustering techniques where, instead of assigning data to individual cluster, the Boolean-like nature of assignment is relaxed by assigning membership grades that assume values in the unit interval and quantify a strength of belongingness of a data point to the individual cluster. Fuzzy C-Means (FCM) (Bezdek, 1981) and Fuzzy C-Medoids (FCMdd) (Krishnapuram et al., 2001) are the two well-known and representative fuzzy clustering techniques. In both techniques, the objective is to form a number of cluster centers (prototypes) and a partition matrix so that a given performance index becomes minimized. FCMdd selects the cluster centers as some of the existing data points (medoids) whereas FCM generates a set of cluster centers using a weighted average of data. In both techniques, the intent is to minimize a weighted sum of distances between data points and cluster centers.

Selecting a distance function to evaluate similarities/dissimilarities of time series has a significant impact on the clustering algorithms and their final results produced by them. This selection may depend upon the nature of the data and the specificity of the application. In most partition-based time series data clustering techniques, the Euclidean distance is commonly used to quantify the similarities/

* Corresponding author: Tel: +1 780 7169026.
E-mail addresses: izakian@ualberta.ca (H. Izakian), wpedrycz@ualberta.ca (W. Pedrycz), iqbaljamal@aqlmc.com (I. Jamal).

http://dx.doi.org/10.1016/j.engappai.2014.12.015
0952-1976/

dissimilarities of time series (or their representation). However, in this technique, one compares the points of time series in a fixed order and cannot take into account existing time shifts. Furthermore, this distance function is applicable only when comparing equal-length time series. On the other hand, in most representation-based (feature-based) clustering techniques, the representatives of clusters cannot be reconstructed in the original time series domain and in such a way they are not useful for data summarization.

In this study, we use Dynamic Time Warping (Berndt and Clifford, 1994) (DTW) distance for clustering time series data. DTW is the most well known technique for evaluating similarity/dissimilarity of time series with respect to their shape information. It is a commonly encountered method and different versions of this technique have been reported in the literature (e.g., see (Salvador and Chan, 2007; Jeong et al., 2011; Bankó and Abonyi, 2012; Chu et al., 2002; Keogh and Pazzani, 1999)) for evaluating similarity/dissimilarity of time series. This distance function determines an optimal match between two time series by stretching or compressing some segments of the series. As the result, patterns occurring at different time instances of time series are considered as similar and consequently, this technique evaluates the similarity of time series with respect to their shapes. Moreover, this technique can express the distance between non-equal-length time series.

As mentioned earlier, DTW distance is a suitable measure to evaluate the similarities/dissimilarities of time series with respect to their shape information. However, there are some difficulties in applying this technique to partition-based clustering methods. Among several reasons behind this, time complexity of this technique is quadratic (Salvador and Chan, 2007) and calculating the average of a set of time series based on this measure is a challenging problem.

In this study, we propose and evaluate three alternatives for fuzzy clustering of time series data using DTW distance. These techniques cluster the time series data with respect to their shape information. Furthermore, the prototypes generated during the clustering process can be used for data summarization based on the shape information within the time series.

Using a DTW-based averaging technique reported in the literature (Petitjean et al., 2011), a Fuzzy C-Means (FCM) clustering is proposed. As the second method, a Fuzzy C-Medoids (FCMdd) clustering, that is free from calculating averages of time series, is examined to select a number of optimal cluster centers as well as an optimal partition matrix. As the third alternative, a hybrid of Fuzzy C-Means and Fuzzy C-Medoids technique is considered for clustering time series data. In all these techniques, time series are clustered based on shape similarities (using the DTW distance) and the cluster centers are in time series domain (not their representation). As a result, the centers can be considered as representatives of time series.

Partition-based clustering of time series data using DTW distance is a challenging problem that has been addressed in this study. Employing a DTW-based averaging technique in FCM is a novel idea presented in this work. Moreover, the proposed hybrid technique which exploits the merits of FCM and FCMdd for clustering time series data realizes a novel idea that has been proposed and investigated in this paper.

The study is structured as follows. In Section 2, we review the proposed methods for clustering time series data. In Section 3 the DTW distance along with an averaging technique based on this measure is briefly reviewed. Section 4 discusses different alternatives for fuzzy clustering of time series data using DTW, and Section 5 reports on the experimental studies. Finally, Section 6 concludes the paper.

2. Literature review

In this section, we briefly review some well-known similarity/dissimilarity measures of time series as well as some clustering techniques reported in the literature for this type of data.

Similarity measures used in time series data can be divided into three general categories including L_p-norm distances, elastic measures, and statistical techniques (Izakian et al., 2013). Selecting a similarity measure in time series data mining depends on the nature of data and the nature of the application itself. When comparing two time series based on a fixed mapping of their points is of interest, L_p-norm distances can be used. The most commonly used examples of L_p-norm are L_1 (Manhattan), L_2 (Euclidean), and L_∞ (Tchebyschev). These distances can be applied to compare time series in their original or a representation domain.

In Izakian et al. (2013), Izakian and Pedrycz (2014) and Izakian and Pedrycz (2014), authors presented an augmented version of Euclidean distance function for fuzzy clustering of time series data. The original time series as well as different representation techniques, including Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), and Piecewise Aggregate Approximation (PAA) were examined for clustering purpose. D'Urso and Maharaj (2009), transformed the time series data through their autocorrelation representation, and used the Euclidean distance to compare data in the new feature space. Then, a FCM technique was employed to cluster the transformed data. In Izakian and Pedrycz (2013), a clustering-based technique for anomaly detection in time series data was proposed. For detecting anomalies in the amplitude of time series, a Fuzzy C-Means clustering applied to the original representation of time series and the Euclidean distance function was employed as a dissimilarity measure. For the purpose of detecting anomalies in the shape of time series, first the data are transformed into an autocorrelation representation, and then the Euclidean distance was employed to compare time series in the transformed domain.

In Vlachos et al. (2003), time series data were represented using a Haar wavelet transform, and the K-Means algorithm along with the Euclidean distance employed to cluster data in the new feature space. In Maharaj and D'Urso (2011), time series are represented through a set of cepstral coefficients, and Euclidean distance is employed to quantify the dissimilarity of time series in the process of a Fuzzy C-Means clustering. Möller-Levet et al. (2003), represented time series data through piecewise linear functions, and proposed a short time series distance, measured as the sum of squared Euclidean distances between the corresponding slopes encountered in two time series. The Fuzzy C-Means algorithm was realized to cluster the data in the new feature space. In Nanda et al., (2010), the Euclidean distance was considered to cluster stock market time series using the K-Means, Fuzzy C-Means, and a self organization map for building a portfolio. The experimental results showed that K-Means could generate more compact clusters in comparison with the other clustering techniques.

Dynamic time warping distance (DTW) (Berndt and Clifford, 1994), longest common subsequence (LCSS) (Vlachos et al., 2002), and edit distance of real-number sequences (EDR) (Chen et al., 2005) are located in the elastic measures category. DTW helps to find an optimal match between two time series by stretching or compressing their segments, and evaluate the similarity of time series with respect to their shapes. LCSS employs the length of the longest subsequence occurring in two time series to quantify their similarity, and EDR takes into account the number of insert, delete and replace operations required to convert one sequence to another one to determine their similarity.

Authors in Niennattrakul and Ratanamahatana (2007) examined K-Means and C-Medoids algorithms for clustering time series data using dynamic time warping distance function. Experimental results indicated that the K-Means clustering cannot generate acceptable results when this distance function is considered (because of the problem of averaging time series based on this measure), and instead, C-Medoids technique may generate satisfactory results. In Keogh and Pazzani (1999), authors proposed a hierarchical clustering technique of time series data, and a DTW distance was considered to quantify the dissimilarity of time series. In Petitjean and Gançarski (2012), a DTW-based averaging of time series is proposed using a compact multiple alignments theory. A Genetic Algorithm was considered to scan the space of compact multiple alignments and find the average of time series. Furthermore, in Petitjean et al. (2011), a DTW-based global averaging of time series was proposed and a K-Means clustering was employed to group data using the proposed averaging technique. Authors compared their averaging technique with some others reported techniques in the literature indicating that the proposed method was more accurate in clustering and classification. In (Jeong et al., 2011), a weighted dynamic time warping distance was introduced for time series classification. The intuition behind this technique was that the DTW does not take into account the relative importance regarding the phase difference between a reference point and a testing point, leading to some misclassification, where shape similarity is in concern. As the result, this technique penalizes points with higher phase difference between a reference point and a testing point to prevent minimum distance distortion resulted through outliers. In comparison with the general DTW technique this technique produces more accurate classification results.

In Bankó and Abonyi (2012), a correlation-based DTW distance was developed to compare multivariate time series. The proposed method combines DTW and PCA-based similarity measures to preserve correlation information when comparing multivariate time series. Experimental studies showed that the proposed method is efficient in dealing with datasets with complex correlation structure. In Chu et al. (2002), an iterative deepening dynamic time warping was proposed. The technique employs a piecewise aggregate approximation to reduce the dimensionality of time series and producing a comparison method. In Sakurai et al. (2005), through considering some constraints on warping path, a fast search method for dynamic time warping was proposed. The proposed method was faster than the techniques proposed in the literature.

Pearson coefficients and the Kullback–Liebler distance (Ramoni et al., 2002) are two well-known statistical indicators to quantify the existing similarities between time series. The first one considers the correlation between time series and regards it as a similarity measure whereas the second one is useful in evaluating the similarity of time series represented by their Markov chains. In Golay et al. (1998), two cross correlation-based similarity measures of raw functional MRI data were proposed to provide functional maps of human brain activity using the fuzzy C-Means method. In (Liu and George, 2003), the Pearson correlation coefficient was considered as the similarity measure expressing closeness of weather time series and a FCM technique employed to cluster the data. In Sobhe Bidari et al. (2008), a Pearson correlation between expression patterns of genes was considered as a similarity measure, and a Fuzzy C-Means and K-Means clustering were employed to find functional patterns of time series gene expression data. Moreover Ramoni et al. (2002) proposed a Bayesian method to cluster time series through modeling the time series as Markov chains and using a symmetric Kullback–Liebler distance between transition matrices. The clustering was considered as a Bayesian model selection problem to find the most suitable set of clusters. A number of surveys of time series data clustering and analysis are reported in Refs. Liao (2005), Fu (2011), Ding et al. (2008), and Esling and Agon (2012). Note that selecting a suitable similarity/dissimilarity measure is application-dependent and has to reflect the nature of time series data. In this paper, we propose some techniques for shape-based clustering of time series data. For this purpose, fuzzy clustering along with the DTW distance function has been considered.

3. Dynamic time warping distance

Using stretching and compressing of time series, DTW determines an optimal match between two time series in the calculations of their differences. In this technique, each point coming from the first time series is compared with any arbitrary point of the second time series. As a result, time series with similar patterns occurred in different time periods, are considered similar. To implement this algorithm, a method of dynamic programming is considered. Table 1 presents a pseudocode for calculating the DTW distance between two time series $\boldsymbol{a}$ and $\boldsymbol{b}$ of length n and m, respectively (Petitjean et al., 2011).

As shown in Table 1, the distance between the two time series $\boldsymbol{a}$ and $\boldsymbol{b}$ in their ith and jth coordinates, $cost_{i,j}$, can be calculated using their current distance, and the minimal value of their distance in their previous coordinates. Moreover, min_index() stands for the coordinates corresponding to the minimum value of the previous cost. The algorithm calculates two matrices, called cost and path comprising the cost and warping path of calculating the DTW distance between two time series.

The main difficulty in applying DTW distance to partition-based clustering techniques (like the FCM algorithms) is about computing the average of a set of time series based on this distance function. There are a number of techniques reported in the literature (e.g., Petitjean et al., 2011; Gupta et al., 1996; Niennattrakul and Ratan-

Table 1
Calculations of the DTW distance between time series $\boldsymbol{a}$ and $\boldsymbol{b}$ (Petitjean et al., 2011).

```
Given:
a = a_1, a_2, ..., a_n, the first time series with length n
b = b_1, b_2, ..., b_m, the second time series with length m
Output:
cost: a matrix of size n × m containing the cost values. cost_{n,m} is the DTW distance between a and b
path: a matrix of size n × m containing a warping path
DTW(a, b):
Let δ be a distance between coordinates of sequences
cost_{1,1} = δ(a_1, b_1);
path_{1,1} = (0,0);
for i=2,3,..., n do
  cost_{i,1} = cost_{i-1,1} + δ(a_i, b_1)
end
for j=2,3,..., m do
  cost_{1,j} = cost_{1,j-1} + δ(a_1, b_j)
end
for i=2,3,..., n do
  for j=2,3,..., m do
    cost_{i,j} = min(cost_{i-1,j}, cost_{i,j-1}, cost_{i-1,j-1}) + δ(a_i, b_j)
    path_{i,j} = min_index((i-1,j), (i,j-1), (i-1,j-1));
  end
end
```

Table 2
Calculating the average of time series based on DTW distance.

Given:
$\boldsymbol{a} = a_1, a_2, ..., a_n$, the initial average of time series with length n
$\mathbf{x}_1 = x_{11}, x_{12}, ..., x_{1m_1}$, the first time series with length m_1
$\mathbf{x}_2 = x_{21}, x_{22}, ..., x_{2m_2}$ the second time series with length m_2
⋮
$\mathbf{x}_N = x_{N1}, x_{N2}, ..., x_{Nm_N}$, the Nth time series with length m_N
Output:
$\boldsymbol{y} = y_1, y_2, ..., y_n$, average of time series $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$
DTWavg($\boldsymbol{a}$, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$):
Let$assosTab = [\varphi, \varphi, ..., \varphi]$,an empty vector with n elements;
for $k = 1,2,\ldots,N$ **do**
  path = path matrix in DTW($\boldsymbol{a}$, $\mathbf{x}_k$);
    $i = n$;
    $j = m_k$
  while $i \geq 1$ and $j \geq 1$ **do**
    $assosTab_i = assosTab_i \cup x_{kj}$;
    $(i,j) \leftarrow path_{i,j}$;
  end
end
for $i = 1,2,\ldots,n$ **do**
  y_i = average of elements located in $assosTab_i$;
end

amahatana, 2009) for averaging a set of sequences based on DTW. Among them, the method proposed in Petitjean et al. (2011) is more accurate and efficient than the others. Assume that there are N time series $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N$ of length $m_1, m_2, ..., m_N$, respectively. Table 2 contains a pseudocode for calculating the average of time series based on DTW distance (Petitjean et al., 2011).

To calculate the average of a set of time series using the pseudocode expressed in Table 2, one of the time series $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N$ can be considered as the initial average of the data. However, this selection may impact the final calculated average of time series. In other words, considering different time series for initializing the described averaging technique may produce different results.

4. Fuzzy clustering of time series using DTW distance

In this paper, we consider three alternatives for fuzzy clustering of time series data. In the first method, we take into account the averaging technique discussed in the previous section and employ the Fuzzy C-Means technique for clustering time series data. In the second method, a Fuzzy C-Medoids technique, which is free from the averaging phase, is examined and finally, in the third technique a hybrid of FCM and FCMdd is employed.

4.1. Fuzzy C-means clustering of time series

FCM clusters N time series, $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N$ into c information granules – fuzzy clusters. The result of clustering is a set of c cluster centers (prototypes), $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_c$ and a partition matrix $U = [u_{ik}]$, $i = 1,2,\ldots,c$, $k = 1,2,\ldots,N$ where $u_{ik} \in [0, 1]$, $\sum_{i=1}^{c} u_{ik} = 1 \ \ \forall k$ and $0 < \sum_{k=1}^{n} u_{ik} < n \ \ \forall i$. This structure arises through the minimization of the following objective function:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m d^2(\boldsymbol{v}_i, \boldsymbol{x}_k) \tag{1}$$

where d is a distance function and m $(m > 1)$ is a fuzzification coefficient. Considering the Euclidean distance function in (1), and starting from a randomly initialized partition matrix the cluster centers and partition matrix are calculated in an iterative fashion as follows:

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^m \boldsymbol{x}_k}{\sum_{k=1}^{N} u_{ik}^m} \tag{2}$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d(\boldsymbol{v}_i, \boldsymbol{x}_k)}{d(\boldsymbol{v}_j, \boldsymbol{x}_k)} \right)^{2/(m-1)}} \tag{3}$$

Note that when considering DTW as the distance function in (1), the cluster centers cannot be calculated using (2). In this paper, for clustering time series using FCM and DTW as the distance function, we employ a weighted averaging method, based on the technique covered in Table 2. Moreover, to initialize the average of time series in each cluster, the time series having the maximum membership degree to that cluster center is considered. The intuition behind this selection is that, the weighted average of a set of time series should be more similar to the time series having the maximum weight in averaging. In the proposed technique in this section, the partition matrix is calculated using (3), where d stands for the DTW distance.

4.2. Fuzzy C-Medoids clustering of time series

While using the FCM technique we calculate the cluster centers as a weighted average of time series, FCMdd selects c time series from the dataset as cluster centers (medoids) and in this way it does not require averaging. It is important to stress that the prototypes formed in this way are the elements of the collection of time series and are fully interpretable. Similarly as in the case of the FCM, the objective function used in this technique is expressed by (1). At the first step, a set of c medoids can be selected randomly from the data. Considering the selected medoids as cluster centers, a partition matrix is constructed using (3). In the next step of the algorithm, in order to minimize the objective function expressed in (1), the new medoids $\mathbf{v}_i$, for $i = 1,2,\ldots c$ are selected in the form (Krishnapuram et al., 2001):

$$\mathbf{v}_i \leftarrow \mathbf{x}_q,$$

where

$$q = \arg\min_{1 \leq j \leq N} \sum_{k=1}^{N} u_{ik}^m \mathrm{DTW}^2(\mathbf{x}_j, \mathbf{x}_k) \tag{4}$$

The algorithm proceeds with the calculation of partition matrix and new medoids until there is no a significant change in the partition matrix.

4.3. Hybrid Fuzzy C-Means and Fuzzy C-Medoids for time series clustering

The performance of the proposed FCM technique directly depends on the accuracy of the DTW-based averaging technique proposed in Petitjean et al. (2011) (and described in the previous section). On the other hand, FCMdd is sensitive to initialization and may generate results in local optima. In this section, we propose a hybrid technique that exploits the merits of both FCM and FCMdd techniques. The proposed technique uses the DTW-based averaging technique only once and is less sensitive to initialization. Fig. 1 shows the overall scheme of the hybrid method.

In this approach, FCMdd is used to cluster the data for a large number of clusters r; $r \geq c$, where c is the number of clusters in data selected by the end user or some cluster validity indices. The result of clustering is a set of prototypes (medoids), $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_r$ and a partition matrix U with r rows and N columns (with N being the number of time series). One may consider U as a new dataset with N objects each having r features located in range [0, 1]. As a matter of fact, the FCMdd technique is used here to create a transformed version of the dataset. Considering that n is the length of time series, usually we have $r \ll n$, and as the result, the transformed dataset is positioned in a smaller search space. Moreover, the Euclidean distance can be now used to compare the data formed in the new feature space. In the next step of the algorithm, the FCM with the Euclidean distance is used to cluster the new dataset, U, into c clusters. The result is a partition matrix U' and a set of cluster centers $\mathbf{v}'_1, \mathbf{v}'_2, ..., \mathbf{v}'_c$.

The proposed technique clusters the data into c clusters and generates a partition matrix, U' describing the membership degree of each time series to different clusters. However, the cluster centers in the proposed technique do not assume a form of time series as they are expressed in the new feature space. To reconstruct the cluster centers as time series in the original domain, one may use the calculated cluster centers by FCM and FCMdd. In fact, each calculated cluster center using FCM,$\mathbf{v}'_i$, $i=1,2,\ldots c$ is composed of r weights assuming values in the range [0, 1] corresponding to the selected medoids $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r$through the FCMdd. As the result, one may generate cluster centers in the form of time series using a DTW-based weighted average of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r$ (expressed in Table 2) where the weights are stored as $\mathbf{v}'_1, \mathbf{v}'_2, ..., \mathbf{v}'_c$.

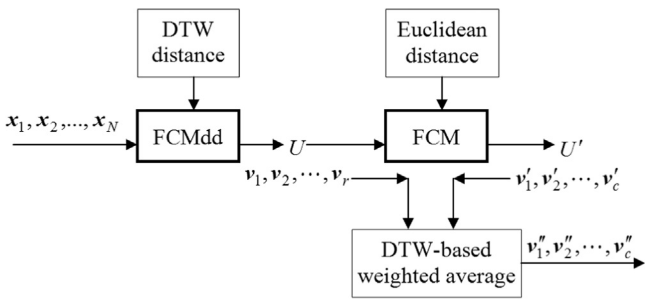

Fig. 1. The overall scheme of the hybrid method proposed in this paper.

Table 3
UCR time series selected for experimental studies.

Dataset	Length of time series	Number of time series	Number of classes
CBF	128	930	3
Synthetic control	60	600	6
Trace	275	200	4
Face four	350	112	4
Olive oil	570	60	4
Gun point	150	200	2
Lighting 7	319	143	7
Beef	60	470	5

5. Experimental studies

In this section, we examined the three alternatives for fuzzy clustering using a number of datasets available online at the UCR time series (www.cs.ucr.edu/~eamonn/time_series_data/). Table 3 summarizes the selected time series.

As shown in this table, datasets with different number of time series, different length of time series, and different number of classes are selected.

For both the FCM and FCMdd techniques the fuzzification coefficient, m, is set to 2 and the maximum number of iterations is set to 40, while the termination condition is considered to be in the form $|U(iter+1)-U(iter)|<0.01$, where $U(iter)$ stands for the partition matrix in iteration $iter$. For the hybrid technique, the number of clusters considered in the first step of the algorithm, r, may have a substantial impact on the performance of the algorithm. Considering a small value for this number (e.g. $r=c$), the FCMdd technique cannot capture the data to be used for clustering at the next step of the algorithm. On the other hand, when increasing the value of this parameter, more structural information about the data can be accommodated and the performance of the algorithm (in terms of accuracy) can be improved. However, assuming a high value for this parameter may result in generating some local optima through FCMdd and this may adversely impact the performance of the algorithm. The optimal value of this parameter depends on the structure available within the dataset. For simplicity, in this paper we set $r=\sqrt{N}$ where N is the number of time series in dataset.

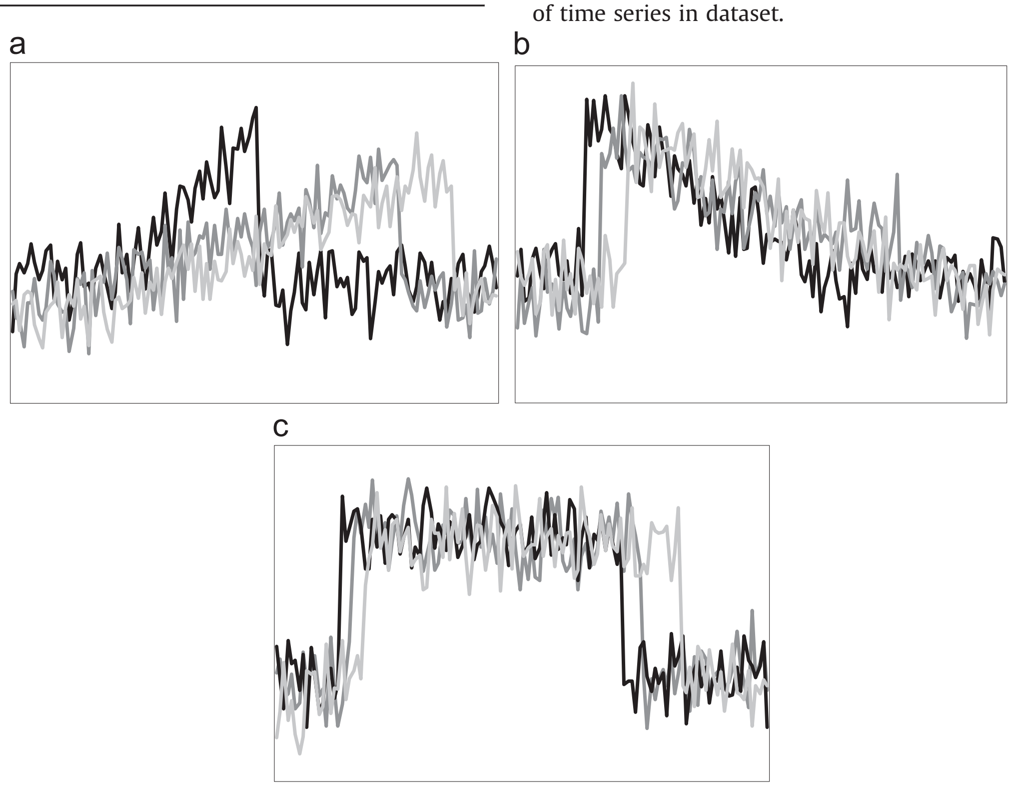

Fig. 2. Example time series belonging to different classes of the CBF dataset.

Figs. 2–4 show a collection of time series belonging to different classes of the CBF, Synthetic control, and Trace datasets (the first three datasets in Table 3).

We set the number of clusters equal to the number of classes present in the above three datasets. Moreover, for purpose of comparative analysis, we considered the FCM technique with the Euclidean distance function (referred here to as FCM-EU). In a collection of figures, Figs. 5–7, we display the obtained prototypes for all the clustering methods and the three described datasets.

Let us consider the available classes in the CBF dataset (Fig. 2), and the obtained prototypes. As shown in Fig. 5 the FCM technique only revealed two classes. All the cluster centers achieved by

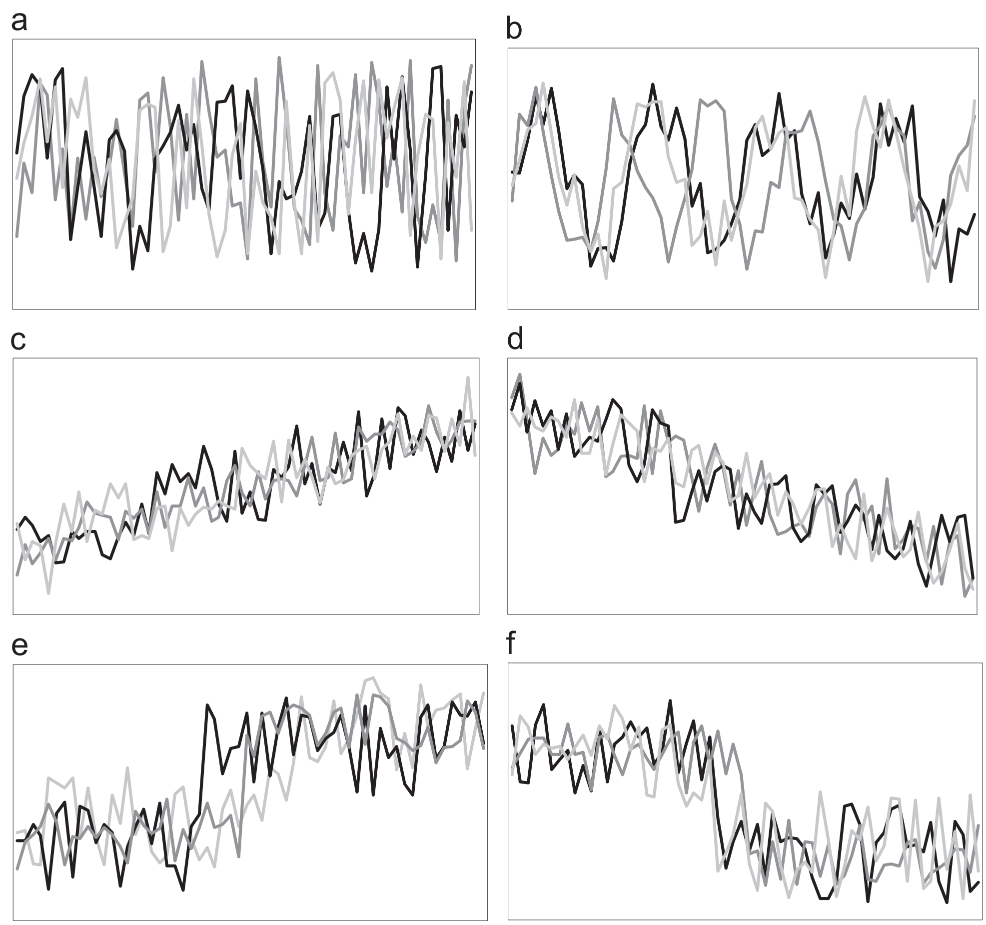

Fig. 3. Selected time series belonging to different classes of the synthetic control dataset.

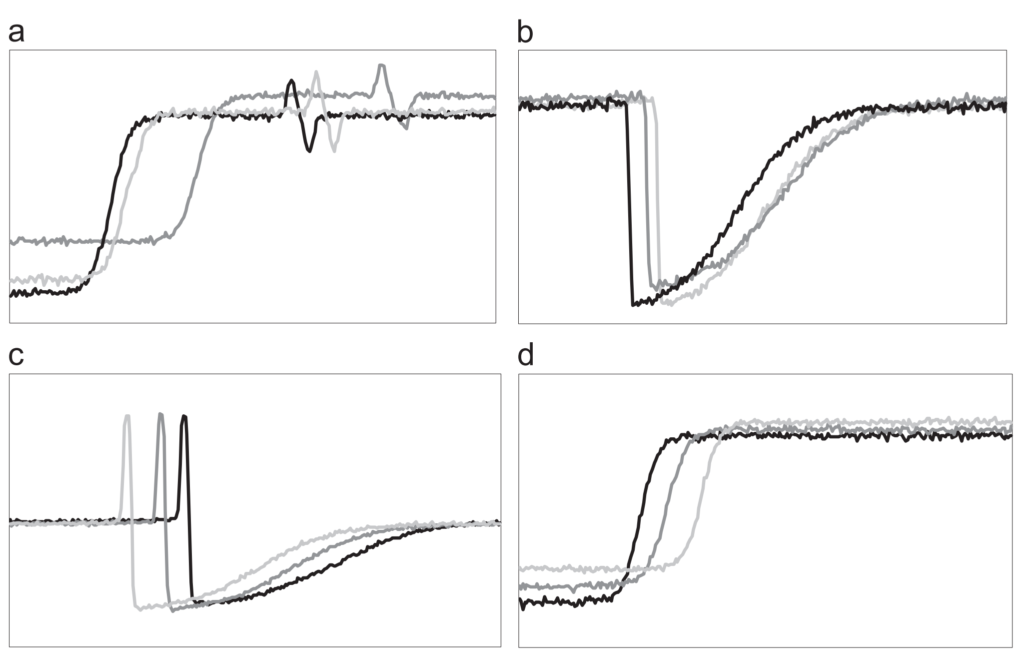

Fig. 4. Example time series belonging to different classes of the trace dataset.

FCMdd are from the same class and the cluster centers coming from FCM-EU are not similar to any classes available in the dataset. The only technique that revealed all the three classes is the hybrid one. Now let us consider the Synthetic control dataset. Since the time series available in this dataset come from six classes, the revealed prototypes using each clustering technique are shown in the separate figures. As shown in Fig. 6, the only technique, which discovers all six classes is the hybrid method. Finally, considering the Trace dataset, all the three proposed techniques studied here are able to reveal four available classes within the dataset, however, the FCM-EU technique which uses the Euclidean distance function cannot produce convincing results.

Fig. 8(a)–(c) shows a process of minimization of the objective function (1) for the FCM and FCMdd techniques for the three datasets.

Both the FCM and FCMdd attained the minimum value of J in the first few iterations of the algorithm.

5.1. Comparative studies

In this section, we compare the three proposed fuzzy clustering techniques along with the FCM method with Euclidean distance function (FCM-EU) over the datasets summarized in Table 3. The number of clusters is considered to be $c = k, k+1, \ldots, 2k$ where k is the number of classes in each dataset. To quantify the ability of the

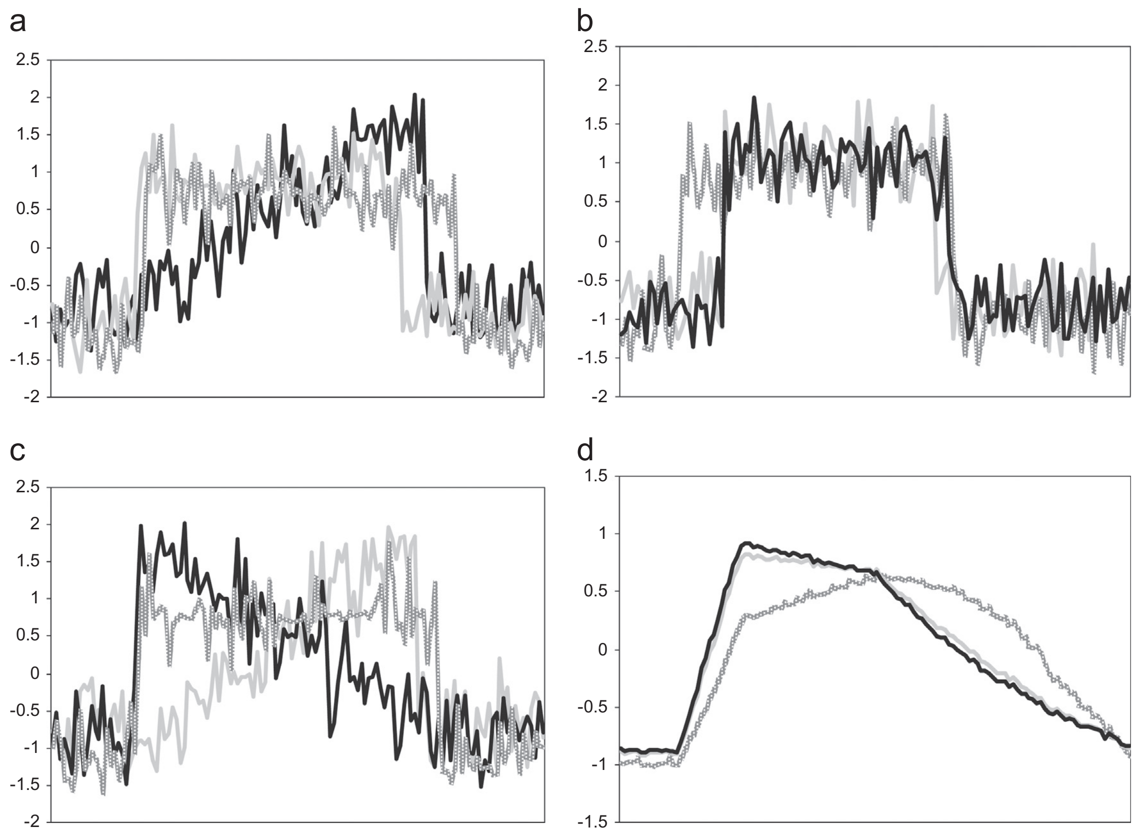

Fig. 5. Prototypes of the CBF dataset produced by (a) FCM, (b) FCMdd, (c) hybrid, and (d) FCM-EU. The number of clusters was set to 3.

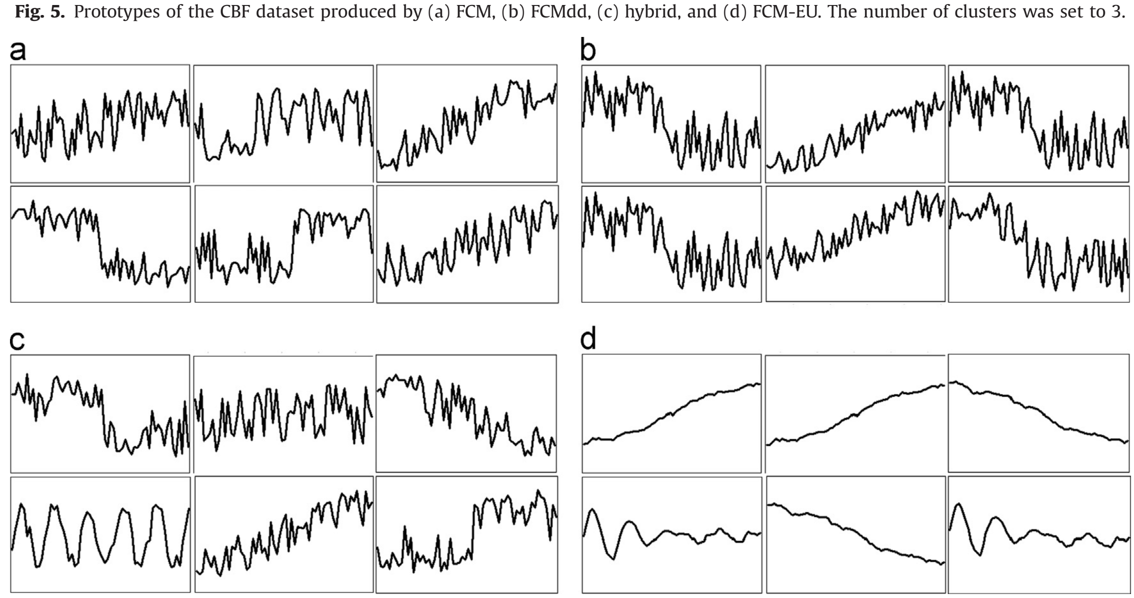

Fig. 6. Prototypes for the Synthetic control dataset using (a) FCM, (b) FCMdd, (c) hybrid, and (d) FCM-EU. The number of clusters was set to 6.

techniques in grouping similar time series, precision has been considered as a suitable evaluation criterion. This criterion is expressed in (5).

$$P = \sum_{j=1}^{c} \frac{|\mathbf{C}_j|}{N} \times \max_{i=1,2,\ldots k} p(\mathbf{K}_i, \mathbf{C}_j), \quad (5)$$

and

$$p(\mathbf{K}_i, \mathbf{C}_j) = \frac{|\mathbf{K}_i \cap \mathbf{C}_j|}{|\mathbf{C}_j|}, \quad (6)$$

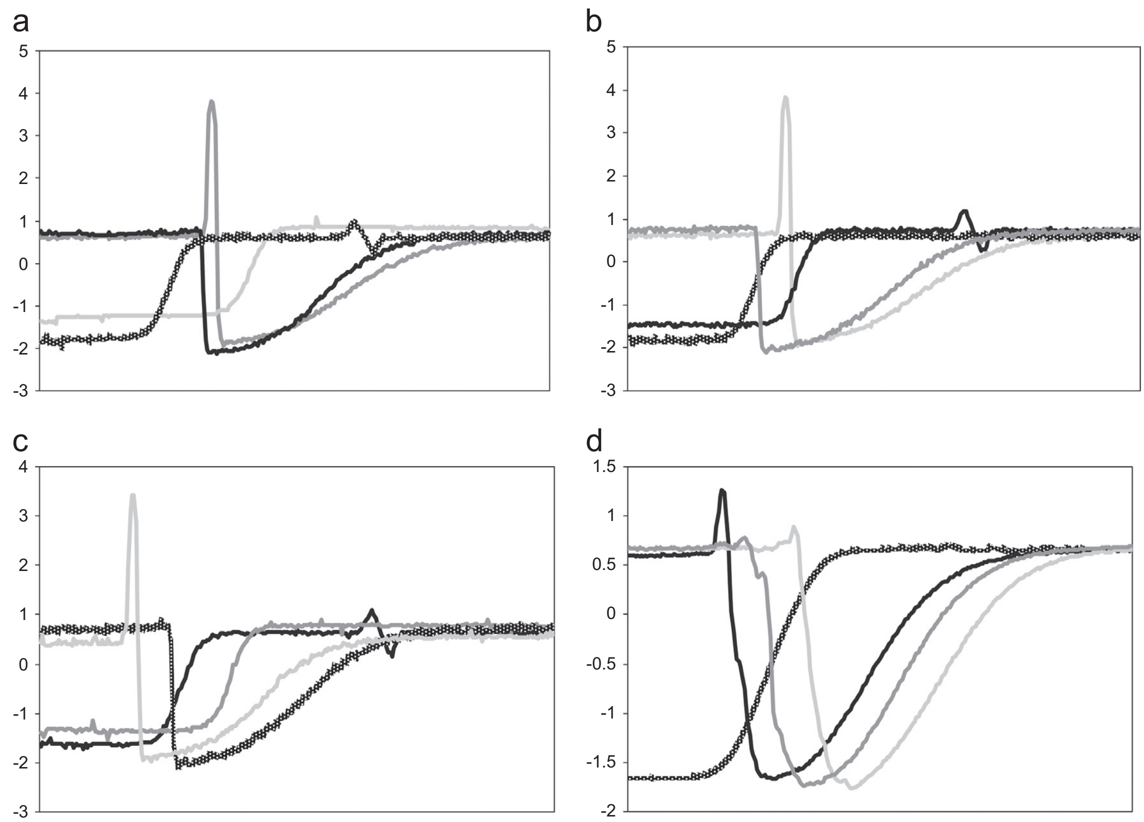

Fig. 7. Prototypes over the Trace dataset using (a) FCM, (b) FCMdd, (c) hybrid, and (d) FCM-EU. The number of clusters set to 4.

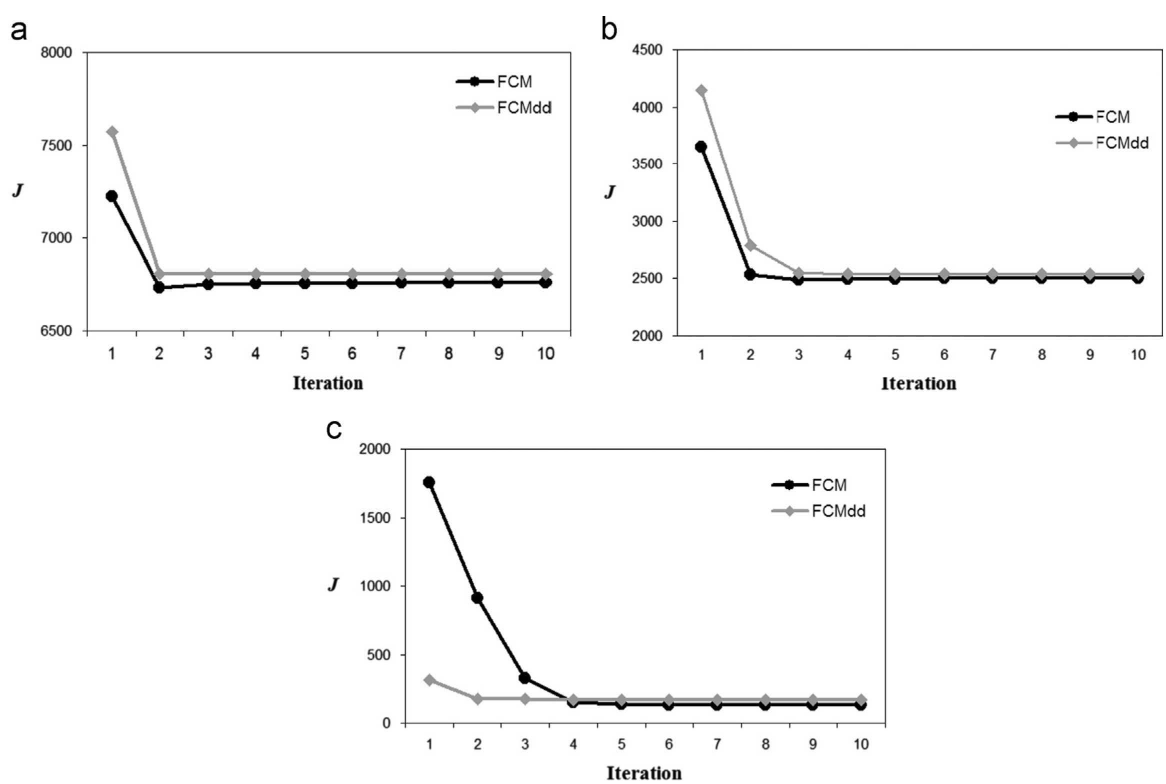

Fig. 8. Values of the objective function in 10 iterations: (a) CBF, (b) synthetic control, and (c) trace.

where c is the number of clusters, k is the number of classes (in the labeled data), N is the number of time series, $\mathbf{K}_i$is the set of objects (time series) belonging to ith class (labeled as i), $\mathbf{C}_j$ is the set of objects belonging to the jth cluster, and $|.|$denotes the number of elements in a set. Table 4 shows the average and standard deviation of the obtained precision; the clustering process has been repeated 20 times (viz. the experiment completed over 20 independent runs).

As shown in this table, FCM-EU generates weak results for most of the data sets. However, for the Olive-oil data set, this technique produces acceptable results. The reason is that in the Olive-oil data set there is no so much time shifting in time series and therefore the Euclidean distance function can be sought as a suitable choice to compare time series. In 70% of cases, the hybrid technique produced higher precision, and the FCMdd, FCM, and FCM-EU techniques achieved better results (higher precision) in 11%, 14%, and 5% cases, respectively.

The t-test carried out at $\alpha = 0.05$ (95% confidence) was employed to assess whether the differences between the achieved results produced by different techniques are statistically significant. Entries marked with asterisk (⁎) indicate that the differences between the achieved precision value produced by the proposed method and the achieved values by the other techniques are statistically significant. In 47% of cases, the hybrid method produced results exhibiting statistically significant differences with the results produced by other techniques. These values for FCMdd, FCM, and FCM-EU are 0%, 2%, and 2%, respectively. In fact, the hybrid technique proposed in this paper, builds on the merits of both FCMdd and FCM techniques to generate results with higher quality.

6. Conclusions

In this paper, to capture shape similarities between time series, dynamic time warping distance is considered and three alternatives, namely a Fuzzy C-Means, a Fuzzy C-Medoids, and a hybrid Fuzzy C-Means and Fuzzy C-Medoids is introduced for clustering time series data with respect to their shape information. A weighted average technique is employed for calculating cluster centers in the

Table 4
Comparison of the fuzzy clustering techniques in terms of the precision value. The results are reported in terms of average and standard deviation obtained over 20 independent runs for $c=k, k+1, \dots 2k$, where k is the number of classes in each dataset. The entries in boldface highlight the best results obtained in each case.

Dataset	c	FCM-EU	FCM	FCMdd	Hybrid
CBF	3	0.631 ± 0.006	0.795 ± 0.166	0.704 ± 0.159	**0.914 ± 0.021⁎**
	4	0.643 ± 0.013	0.828 ± 0.109	0.748 ± 0.150	**0.978 ± 0.017⁎**
	5	0.652 ± 0.012	0.869 ± 0.084	0.795 ± 0.092	**0.974 ± 0.025⁎**
	6	0.660 ± 0.006	0.923 ± 0.079	0.805 ± 0.113	**0.976 ± 0.023⁎**
Synthetic control	6	0.583 ± 0.009	0.690 ± 0.092	0.632 ± 0.175	**0.929 ± 0.048⁎**
	7	0.512 ± 0.008	0.745 ± 0.087	0.767 ± 0.111	**0.933 ± 0.025⁎**
	8	0.534 ± 0.010	0.817 ± 0.097	0.833 ± 0.100	**0.936 ± 0.025⁎**
	9	0.581 ± 0.041	0.826 ± 0.061	0.769 ± 0.110	**0.949 ± 0.018⁎**
	10	0.528 ± 0.009	0.889 ± 0.058	0.887 ± 0.045	**0.933 ± 0.025⁎**
	11	0.573 ± 0.029	0.871 ± 0.080	0.863 ± 0.076	**0.921 ± 0.030⁎**
	12	0.570 ± 0.046	0.879 ± 0.056	0.905 ± 0.068	**0.936 ± 0.030**
Trace	4	0.539 ± 0.018	0.755 ± 0.006	0.743 ± 0.050	**0.761 ± 0.009⁎**
	5	0.536 ± 0.016	0.767 ± 0.032	0.778 ± 0.121	**0.829 ± 0.029**
	6	0.549 ± 0.002	0.801 ± 0.051	0.820 ± 0.094	**0.895 ± 0.056⁎**
	7	0.558 ± 0.009	0.820 ± 0.053	0.813 ± 0.088	**0.877 ± 0.071⁎**
	8	0.562 ± 0.005	0.866 ± 0.061	0.776 ± 0.081	**0.914 ± 0.055⁎**
Face four	4	0.585 ± 0.044	0.611 ± 0.099	0.610 ± 0.091	**0.699 ± 0.061⁎**
	5	0.623 ± 0.042	0.704 ± 0.116	**0.747 ± 0.120**	0.745 ± 0.039
	6	0.637 ± 0.034	0.760 ± 0.085	0.742 ± 0.110	**0.780 ± 0.027**
	7	0.657 ± 0.024	0.737 ± 0.097	0.674 ± 0.038	**0.811 ± 0.041⁎**
	8	0.651 ± 0.011	0.739 ± 0.103	0.730 ± 0.107	**0.819 ± 0.049⁎**
Olive-oil	4	**0.783 ± 0.000**	0.770 ± 0.038	0.728 ± 0.055	0.735 ± 0.045
	5	0.783 ± 0.008	**0.798 ± 0.033**	0.737 ± 0.073	0.742 ± 0.038
	6	0.775 ± 0.009	**0.800 ± 0.026⁎**	0.742 ± 0.042	0.763 ± 0.030
	7	0.827 ± 0.009	**0.837 ± 0.028**	0.792 ± 0.046	0.805 ± 0.042
	8	**0.847 ± 0.022⁎**	0.815 ± 0.020	0.815 ± 0.030	0.813 ± 0.028
Gun point	2	0.500 ± 0.000	0.505 ± 0.004	0.507 ± 0.003	**0.515 ± 0.025**
	3	0.576 ± 0.016	0.615 ± 0.028	0.604 ± 0.056	**0.637 ± 0.037⁎**
	4	0.650 ± 0.063	0.651 ± 0.035	0.653 ± 0.059	**0.727 ± 0.024⁎**
Lighting 7	7	0.384 ± 0.024	0.502 ± 0.044	0.514 ± 0.060	**0.539 ± 0.037**
	8	0.372 ± 0.018	0.515 ± 0.042	0.513 ± 0.037	**0.545 ± 0.042⁎**
	9	0.376 ± 0.017	**0.555 ± 0.029**	0.528 ± 0.026	0.547 ± 0.042
	10	0.377 ± 0.015	0.569 ± 0.033	0.557 ± 0.056	**0.574 ± 0.033**
	11	0.376 ± 0.023	0.550 ± 0.036	0.560 ± 0.053	**0.580 ± 0.039**
	12	0.385 ± 0.029	**0.588 ± 0.058**	0.578 ± 0.021	0.576 ± 0.032
	13	0.364 ± 0.008	**0.584 ± 0.046**	0.564 ± 0.024	0.577 ± 0.035
	14	0.383 ± 0.032	0.591 ± 0.030	**0.604 ± 0.035**	0.588 ± 0.047
Beef	5	0.525 ± 0.009	0.515 ± 0.030	**0.530 ± 0.017**	0.513 ± 0.040
	6	0.533 ± 0.000	0.522 ± 0.029	0.525 ± 0.025	**0.535 ± 0.050**
	7	0.533 ± 0.000	0.530 ± 0.025	0.540 ± 0.029	**0.552 ± 0.023**
	8	0.538 ± 0.008	0.562 ± 0.025	**0.570 ± 0.033**	0.558 ± 0.018
	9	0.563 ± 0.015	0.553 ± 0.030	**0.577 ± 0.042**	0.570 ± 0.023
	10	0.562 ± 0.019	0.572 ± 0.024	0.570 ± 0.022	**0.573 ± 0.043**

FCM method. For the FCMdd technique, one may select a number of time series within the dataset as cluster centers (medoids) and this technique is free from averaging. In the hybrid technique, first the time series are clustered using the FCMdd, transforming the data into a new feature space, and then a FCM technique along with a Euclidean distance function is used to cluster the transformed data. Using the DTW-based averaging technique and the cluster centers revealed using FCMdd and FCM, the cluster centers of the hybrid method can be constructed in the original feature space.

Experimental studies carried out over eight datasets coming from UCR time series repository. In comparison with the conventional FCM-EU method, the proposed techniques generate more acceptable results in terms of the shape of generated cluster centers and the precision of clustering results. Moreover, among the three proposed techniques, in most cases the hybrid technique produces superior results. Furthermore, in 47% of cases of experiments the difference between the achieved precision by the hybrid technique and the obtained precision using the other methods is statistically significant.

Acknowledgment

This work was supported in part by Alberta Innovates – Technology Futures and Alberta Advanced Education & Technology, Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada Research Chair Program.

References

Bankó, Z., Abonyi, J., 2012. Correlation based dynamic time warping of multivariate time series. Expert Syst. Appl. 39 (17), 12814–12823.

Berndt, D., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: Proceedings of the AAAI-94 Workshop Knowledge Discovery in Databases, pp. 359–370.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Functions. Plenum, New York.

Chen, L., Özsu, M.T., Oria, V., 2005. Robust and fast similarity search for moving object trajectories. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 491–502.

Chu, S., Keogh, E., Hart D., Pazzani, M., 2002. Iterative deepening dynamic time warping for time series. In: Proceedings of the 2nd SIAM International Conference on Data Mining.

Ding, H., Trajcevski, G., Scheuermann, P., X. Wang, Keogh, E., 2008. Querying and mining of time series data: experimental comparison of representations and distance measures. In: Proceedings of the VLDB Endowment, pp. 1542–1552.

D'Urso, P., Maharaj, E.A., 2009. Autocorrelation-based fuzzy clustering of time series. Fuzzy Sets Syst. 160 (24), 3565–3589.

Esling, P., Agon, C., 2012. Time-series data mining. ACM Comput. Surv. 45 (1), Article 12.

Fu, T.-C., 2011. A review on time series data mining. Eng. Appl. Artif. Intell. 24 (1), 164–181.

Golay, X., Kollias, S., Stoll, G., Meier, D., Valavanis, A., Boesiger, P., 1998. A new correlation-based fuzzy logic clustering algorithm for fMRI. Magn. Reson. Med. 40 (2), 249–260.

Gupta, L., Molfese, D., Tammana, R., Simos, P., 1996. Nonlinear alignment and averaging for estimating the evoked potential. IEEE Trans. Biomed. Eng. 43 (4), 348–356.

Izakian, H., Pedrycz, W., 2013. Anomaly detection in time series data using a Fuzzy C-Means clustering. Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), pp. 1513–1518.

Izakian, H., Pedrycz, W., 2014. Agreement-based Fuzzy C-Means for clustering data with blocks of features. Neurocomputing 127, 266–280.

Izakian, H., Pedrycz, W., 2014. Anomaly detection and characterization in spatial time series data: a cluster-centric approach. IEEE Trans. Fuzzy Syst. 22 (6), 1063–6706.

Izakian, H., Pedrycz, W., Jamal, I., 2013. Clustering spatio–temporal data: an augmented fuzzy C-Means. IEEE Trans. Fuzzy Syst. 21 (5), 855–868.

Jeong, Y.-S., Jeong, M.K., Omitaomu, O.A., 2011. Weighted dynamic time warping for time series classification. Pattern Recognit. 44 (9), 2231–2240.

Keogh, E., Pazzani, M.J., 1999. Scaling up dynamic time warping to massive datasets. In: Proceedings of the of the Third European Conference on Principles of Data Mining and Knowledge Discovery, pp. 1–11.

Krishnapuram, R., Joshi, A., Nasraoui, O., Yi, L., 2001. Low-complexity fuzzy relational clustering algorithms for web mining. IEEE Trans. Fuzzy Syst. 9 (4), 595–607.

Liao, T.W., 2005. Clustering of time series data – a survey. Pattern Recognit. 38 (11), 1857–1874.

Liu, Z., George, R., 2003. Fuzzy cluster analysis of spatio-temporal data. In: Proceedings of the of the ISCIS 18th International symposium, pp. 984–991.

Maharaj, E.A., D'Urso, P., 2011. Fuzzy clustering of time series in the frequency domain. Inf. Sci. 181 (7), 1187–1211.

Möller-Levet, C.S., Klawonn, F., Cho, K.-H., Wolkenhauer O., 2003. Fuzzy clustering of short time series and unevenly distributed sampling points LNCS. In: Proceedings of the of the IDA, pp. 28–30.

Nanda, S.R., Mahanty, B., Tiwari, M.K., 2010. Clustering Indian stock market data for portfolio management. Expert Syst. Appl. 37 (12), 8793–8798.

Niennattrakul, V., Ratanamahatana, C.A., 2007. On clustering multimedia time series data using K-Means and dynamic time warping. In: Proceedings of the International Conference on Multimedia and Ubiquitous Engineering, pp. 733–738.

Niennattrakul, V., Ratanamahatana, C.A., 2009. Shape averaging under time warping. In: Proceedings of the International Conference on Electrical Engineering/ Electronics, Computer, Telecommunications, and Information Technology, pp. 626–629.

Petitjean, F., Gançarski, P., 2012. Summarizing a set of time series by averaging: from Steiner sequence to compact multiple alignment. Theor. Comput. Sci. 414, 76–91.

Petitjean, F., Ketterlin, A., Gancarski, P., 2011. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognit. 44 (3), 678–693.

Ramoni, M., Sebastiani, P., Cohen, P., 2002. Bayesian clustering by dynamics. Mach. Learn. 47 (1), 91–121.

Sakurai, Y., Yoshikawa, M., Faloutsos, C., 2005. FTW: fast similarity search under the time warping distance In: Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 326–337.

Salvador, S., Chan, P., 2007. Toward accurate dynamic time warping in linear time and space. Intell. Data Anal. 11 (5), 561–580.

Sobhe Bidari, P., Manshaei, R., Lohrasebi, T., Feizi, A., Malboobi, M.A., Alirezaie, J., 2008. Time series gene expression data clustering and pattern extraction in Arabidopsis thaliana phosphatase-encoding genes. In: Proceedings of the 8th IEEE International Conference on BioInformatics and BioEngineering, pp. 1–6.

Vlachos, M., Gunopulos, D., Kollios, G., 2002. Discovering similar multidimensional trajectories. In: Proceedings of the International Conference on Data Engineering, pp. 673–684.

Vlachos, M., Lin, J., Keogh, E., Gunopulos, D., 2003. A wavelet based anytime algorithm for K-Means clustering of time series. In: Proceedings of the Workshop on Clustering High Dimensionality Data and Its Applications, pp. 23–30.