



On Some Fuzzy Clustering Algorithms for Time-Series Data

Mizuki Fujita^(✉) and Yuchi Kanzawa^{ORCID}

Shibaura Institute of Technology, 3-7-5 Toyosu, Koto, Tokyo 135-8548, Japan
{ma21131,kanzawa}@shibaura-it.ac.jp

Abstract. Various fuzzy clustering algorithms have been proposed for vectorial data. However, these methods have not been applied to time-series data. This paper presents three fuzzy clustering algorithms for time-series data based on dynamic time warping (DTW). The first algorithm involves Kullback–Leibler divergence regularization of the DTW k -means objective function. The second algorithm replaces the membership of the DTW k -means objective function with its power. The third algorithm involves q-divergence regularization of the objective function of the first algorithm. Theoretical discussion shows that the third algorithm is a generalization of the first and second algorithms, which is substantiated through numerical experiments.

Keywords: Time-series data · Fuzzy clustering · Dynamic time warping

1 Introduction

Hard c -means (HCM) is the most commonly used type of clustering algorithm [1]. The fuzzy c -means (FCM) [2] approach is an extension of the HCM that allows each object to belong to all or some of the clusters to varying degrees. To distinguish the general FCM method from other proposed, such as entropy-regularized FCM (EFCM) [3], it is referred to as the Bezdek-type FCM (BFCM) in this work. The above mentioned algorithms may misclassify some objects that should be assigned to a large cluster as belonging to a smaller cluster if the cluster sizes are not balanced. To overcome this problem, some approaches introduce variables to control the cluster sizes [4, 5]. Such variables have been added to the BFCM and EFCM algorithms to derive the revised BFCM (RBFCM) and revised EFCM (REFCM) [6] algorithms, respectively.

In the aforementioned clustering algorithms, the dissimilarities between the objects and cluster centers are measured as the inner-product-induced squared distances. This measure cannot be used for time-series data because they vary over time. Dynamic time warping (DTW) is a representative dissimilarity with respect to time-series data. Hence, a clustering algorithm using the DTW is proposed [8] herein and referred to as the DTW k -means algorithm.

The accuracies that can be achieved with fuzzy clustered results are better than those using hard clustering. Various kinds of fuzzy clustering algorithms have been proposed in literature for vectorial data [2, 3]. However, this is not true for time-series data, which is the main motivation for this study.

In this work, we propose three fuzzy clustering algorithms for time-series data. The first algorithm involves the Kullback–Leibler (KL) divergence regularization of the DTW k -means objective function, which is referred to as the KL-divergence-regularized fuzzy DTW c -means (KLFDTWCM); this approach is similar to the REFCM obtained by KL divergence regularization of the HCM objective function. In the second algorithm, the membership of the DTW k -means objective function is replaced with its power, which is referred to as the Bezdek-type fuzzy DTW c -means (BFDTWCM); this method is similar to the RBFCM, where the membership of the HCM objective function is replaced with its power. The third algorithm is obtained by q -divergence regularization of the objective function of the first algorithm (QFDTWCM). The theoretical results indicate that the QFDTWCM approach reduces to the BFDTWCM under a specific condition and to the KLFDTWCM under a different condition. Numerical experiments were performed using artificial datasets to substantiate these observations.

The remainder of this paper is organized as follows. Section 2 introduces the notations used herein and the background regarding some conventional algorithms. Section 3 describes the three proposed algorithms. Section 4 presents the procedures and results of the numerical experiments demonstrating the properties of the proposed algorithms. Finally, Sect. 5 presents the conclusions of this work.

2 Preliminaries

2.1 Divergence

For two probability distributions P and Q , the KL divergence of Q from P , $D_{\text{KL}}(P||Q)$ is defined to be

$$D_{\text{KL}}(P||Q) = \sum_k P(k) \ln \left(\frac{P(k)}{Q(k)} \right). \quad (1)$$

KL divergence has been used to achieve fuzzy clustering [3] of vectorial data. KL divergence has been extended by using q -logarithmic function

$$\ln_q(x) = \frac{1}{1-q} (x^{1-q} - 1) \quad (\text{for } x > 0) \quad (2)$$

as

$$D_q(P||Q) = \frac{1}{1-q} \left(\sum_k P(k)^q Q(k)^{1-q} - 1 \right), \quad (3)$$

referred to as q -divergence [7]. In the limit $q \rightarrow 1$, the KL-divergence is recovered. q -divergence has been implicitly used to derive fuzzy clustering only for vectorial data [6] although that is not indicated in the literature.

2.2 Clustering for Vectorial Data

Let $X = \{x_k \in \mathbb{R}^D \mid k \in \{1, \dots, N\}\}$ be a dataset of D -dimensional points. The set of cluster centers is denoted by $v = \{v_i \in \mathbb{R}^D \mid i \in \{1, \dots, C\}\}$. The membership of x_k with respect to the i -th cluster is denoted by $u_{i,k}$ ($i \in \{1, \dots, C\}, k \in \{1, \dots, N\}$) and has the following constraint:

$$\sum_{i=1}^C u_{i,k} = 1. \tag{4}$$

The variable controlling the i -th cluster size is denoted by α_i , and has the constraint

$$\sum_{i=1}^C \alpha_i = 1. \tag{5}$$

The HCM, RBFCM, and REFCM clusters are respectively obtained by solving the following optimization problems:

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - v_i\|_2^2, \tag{6}$$

$$\underset{u,v,\alpha}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (\alpha_i)^{1-m} (u_{i,k})^m \|x_k - v_i\|_2^2, \tag{7}$$

$$\underset{u,v,\alpha}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - v_i\|_2^2 + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\alpha_i} \right). \tag{8}$$

where $m > 1$ and $\lambda > 0$ are the fuzzification parameters. When $m = 1$, the RBFCM is reduced to HCM; the larger the value of m , the fuzzier are the memberships. When $\lambda \rightarrow +\infty$, the REFCM is reduced to HCM; the smaller the value of λ , the fuzzier are the memberships.

2.3 Clustering of Time-Series Data: DTW k -Means

Let $X = \{x_k \in \mathbb{R}^D \mid k \in \{1, \dots, N\}\}$ be a time-series dataset and $x_{k,\ell}$ be its elements at time ℓ . Let $v = \{v_i \in \mathbb{R}^D \mid i \in \{1, \dots, C\}\}$ be the set of cluster centers set $v_{i,\ell}$ be its elements at time ℓ . Let $\text{DTW}_{i,k}$ be the dissimilarities between the objects x_k and cluster centers v_i as below, with $\text{DTW}_{i,k}$ being defined as follows DTW [8]. Denoting $\Omega_{i,k} \in \{0, 1\}^{D \times D}$ as the warping path used to calculate $\text{DTW}_{i,k}$, the membership of x_k with respect to the i -th cluster is given by $u_{i,k}$ ($i \in \{1, \dots, C\}, k \in \{1, \dots, N\}$). The DTW k -means is obtained by solving the following optimization problem

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \text{DTW}_{i,k}. \tag{9}$$

in accordance with Eq. (4), where

$$\text{DTW}_{i,k} = \sqrt{d(v_{i,D}, x_{k,D})},$$

$$d(v_{i,D}, x_{k,D}) = \|x_{k,D} - v_{i,D}\|^2 \quad (10)$$

$$+ \min\{d(v_{i,D-1}, x_{k,D-1}), d(v_{i,D}, x_{k,D-1}), d(v_{i,D-1}, x_{k,D})\}. \quad (11)$$

In addition to DTW, we obtain the warping path that maps the pairs (ℓ, m) for each element in the series to minimize the distance between them. Hence, the warping path is a sequence of pairs (ℓ, m) . Here, we consider matrices $\{\Omega_{i,k} \in \{0, 1\}^{D \times D}\}_{(i,k)=(1,1)}^{(C,N)}$ whose (ℓ, m) -th element is one if (ℓ, m) is an element of the corresponding warping path and zero otherwise then, we have the cluster centers

$$v_i = \left(\sum_{x_k \in G_i}^N \Omega_{i,k} x_k \right) \oslash \left(\sum_{x_k \in G_i} \Omega_{i,k} \mathbf{1} \right), \quad (12)$$

where $\mathbf{1}$ is the D -dimensional vector with all elements equal to one, and \oslash describes element-wise division. The DTW k -means algorithm can be summarized as follows.

Algorithm 1 (DTW k -means). [8]

STEP 1. Set the number of clusters C and initial membership $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$.

STEP 2. Calculate $\{v_i\}_{i=1}^C$ as

$$v_i = \frac{\sum_{k=1}^N u_{i,k} x_k}{\sum_{k=1}^N u_{i,k}}. \quad (13)$$

STEP 3. Calculate $\{\text{DTW}_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ and update $\{v_i\}_{i=1}^C$ as

(a) Calculate $\text{DTW}_{i,k}$ from Eq. (11).

(b) Update v_i from Eq. (25).

(c) Check the limiting criterion for v_i . If the criterion is not satisfied, go to Step (a).

STEP 4. Update $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ as

$$u_{i,k} = \begin{cases} 1 & (i = \arg \min_{1 \leq j \leq C} \{\text{DTW}_{j,k}\}), \\ 0 & (\text{otherwise}). \end{cases} \quad (14)$$

STEP 5. Check the limiting criterion for (u, v) . If the criterion is not satisfied, go to Step 3

3 Proposed Algorithms

3.1 Concept

In this work, we propose three fuzzy clustering algorithms for time-series data.

The first algorithm is similar to the REFCM obtained by KL divergence regularization the DTW k -means objective function, which is referred to as KLFDTWCM. The optimization problem for this is given by

$$\underset{u,v,\alpha}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \text{DTW}_{i,k} + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \ln \left(\frac{u_{i,k}}{\alpha_i} \right) \quad (15)$$

subject to Eqs. (4) and (5).

The second algorithm is similar to the RBFCM obtained by replacing the membership of the HCM objective function with its power, which is referred to as BFDTWCM. The optimization problem is then given by

$$\underset{u,v,\alpha}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (\alpha_i)^{1-m} (u_{i,k})^m \text{DTW}_{i,k} \quad (16)$$

subject to Eqs. (4) and (5).

The third algorithm is obtained by q-divergence regularization of the BFDTWCM, which is referred to as QFDTWCM. The optimization problem in this case is given by

$$\underset{u,v,\alpha}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (\alpha_i)^{1-m} (u_{i,k})^m \text{DTW}_{i,k} + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (\alpha_i)^{1-m} (u_{i,k})^m \quad (17)$$

subject to Eqs. (4) and (5). This optimization problem relates the optimization problems for BFDTWCM and KLFDTWCM because Eq. (17) with $\lambda \rightarrow +\infty$ reduces to the BFDTWCM method and Eq. (17) with $m \rightarrow 1$ reduces to the KLFDTWCM approach. In the next subsection, we present derivation of the update equations for u, v , and α based on of the minimization problem in Eqs. (15), (16), and (17).

3.2 KLFDTWCM, BFDTWCM and QFDTWCM

The KLFDTWCM is obtained by solving the optimization problem in Eqs. (15), (4) and (5), where the Lagrangian $L(u, v, \alpha)$ is defined as

$$\begin{aligned} L(u, v, \alpha) = & \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \text{DTW}_{i,k} + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \ln \left(\frac{u_{i,k}}{\alpha_i} \right) \\ & + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \beta \left(1 - \sum_{i=1}^C \alpha_i \right) \end{aligned} \quad (18)$$

using Lagrangian multipliers $(\gamma_1, \dots, \gamma_{N+1})$. The necessary conditions for optimality are given as

$$\frac{\partial L(u, v, \alpha)}{\partial u_{i,k}} = 0, \quad (19)$$

$$\frac{\partial L(u, v, \alpha)}{\partial \alpha_i} = 0, \quad (20)$$

$$\frac{\partial L(u, v, \alpha)}{\partial \gamma_k} = 0, \quad (21)$$

$$\frac{\partial L(u, v, \alpha)}{\partial \beta} = 0. \quad (22)$$

The optimal membership is obtained from Eqs. (19) and (21) in a manner similar to that of the REFCM as

$$u_{i,k} = \left[\sum_{j=1}^C \frac{\alpha_j}{\alpha_i} \exp(-\lambda(\text{DTW}_{j,k} - \text{DTW}_{i,k})) \right]^{-1}. \quad (23)$$

The optimal variable for controlling the cluster sizes is obtained from Eqs. (20) and (22) in a manner similar to that of the REFCM as

$$\alpha_i = \frac{\sum_{k=1}^N u_{i,k}}{N}. \quad (24)$$

Recall that in the DTW k -means approach, the cluster centers v_i are calculated using $\Omega_{i,k}$ and x_k belonging to cluster $\#i$, as shown in Eq. (12), which can be equivalently written as

$$v_i = \left(\sum_{k=1}^N u_{i,k} \Omega_{i,k} x_k \right) \oslash \left(\sum_{k=1}^N u_{i,k} \Omega_{i,k} \mathbf{1} \right). \quad (25)$$

This form can be regarded as the $u_{i,k}$ -weighted mean of $\Omega_{i,k} x_k$. Similarly, the cluster centers for KLFDTWCM are calculated using Eq. (25). KLFDTWCM can be described as follows:

Algorithm 2 (KLFDTWCM).

STEP 1. Set the number of clusters C , fuzzification parameter $\lambda > 0$, and initial membership $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$.

STEP 2. Calculate v_i from Eq. (13).

STEP 3. Calculate $\{\text{DTW}_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ and update $\{v_i\}_{i=1}^C$ as

(a) Calculate $\text{DTW}_{i,k}$ from Eq. (11).

(b) Update v_i from Eq. (25).

(c) Check the limiting criterion for v_i . If the criterion is not satisfied, go to Step (a).

STEP 4. Update u from Eq. (23)

STEP 5. Calculate α from Eq. (24)

STEP 6. Check the limiting criterion for (u, v, α) . If the criterion is not satisfied, go to Step 3.

The BFDTWCM is obtained by solving the optimization problem in Eqs. (16), (4), and (5). Similar to the derivation of the KLFDTWCM, the optimal membership u , variable for controlling the cluster sizes α , and cluster centers v are obtained as

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\alpha_j}{\alpha_i} \left(\frac{DTW_{j,k}}{DTW_{i,k}} \right)^{1/(1-m)}}, \tag{26}$$

$$\alpha_i = \frac{1}{\sum_{j=1}^C \left(\frac{\sum_{k=1}^N (u_{j,k})^m DTW_{j,k}}{\sum_{k=1}^N (u_{i,k})^m DTW_{i,k}} \right)^{1/m}}, \tag{27}$$

$$v_i = \left(\sum_{k=1}^N (u_{i,k})^m \Omega_{i,k} x_k \right) \oslash \left(\sum_{k=1}^N (u_{i,k})^m \Omega_{i,k} \mathbf{1} \right), \tag{28}$$

respectively. The BFDTWCM can be described as follows:

Algorithm 3 (BFDTWCM).

STEP 1. Set the number of clusters C , fuzzification parameter $m > 1$, and initial membership $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$.

STEP 2. Calculate v_i from Eq. (13).

STEP 3. Calculate $\{DTW_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ and update $\{v_i\}_{i=1}^C$ as

(a) Calculate $DTW_{i,k}$ from Eq. (11).

(b) Update v_i from Eq. (28).

(c) Check the limiting criterion for v_i . If the criterion is not satisfied, go to Step. (a).

STEP 4. Update u from Eq. (26)

STEP 5. Calculate α from Eq. (27)

STEP 6. Check the limiting criterion for (u, v, α) . If the criterion is not satisfied, go to Step. 3.

The QFDTWCM is obtained by solving the optimization problem in Eqs. (17), (4), and (5). Similar to the derivations of BFDTWCM and KLFDTWCM, the optimal membership u and variable for controlling the cluster sizes α are obtained as

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\alpha_j}{\alpha_i} \left(\frac{1-\lambda(1-m)DTW_{j,k}}{1-\lambda(1-m)DTW_{i,k}} \right)^{1/(1-m)}}, \tag{29}$$

$$\alpha_i = \frac{1}{\sum_{j=1}^C \left(\frac{\sum_{k=1}^N (u_{j,k})^m (1-\lambda(1-m)DTW_{j,k})}{\sum_{k=1}^N (u_{i,k})^m (1-\lambda(1-m)DTW_{i,k})} \right)^{1/m}}, \tag{30}$$

respectively. The optimal cluster centers are defined by Eq. (28). The QFDTWCM can be described as follows:

Algorithm 4 (QFDTWCM).

- STEP 1. Set the number of clusters C , fuzzification parameter $m > 1, \lambda > 0$, and initial membership $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$.
- STEP 2. Calculate $\{v_i\}_{i=1}^C$ from Eq. (13).
- STEP 3. Calculate $\{DTW_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ and update $\{v_i\}_{i=1}^C$ as
- (a) Calculate $DTW_{i,k}$ from Eq. (11).
 - (b) Update v_i from Eq. (28).
 - (c) Check the limiting criterion for v_i . If the criterion is not satisfied, go to Step (a).
- STEP 4. Update u from Eq. (29).
- STEP 5. Calculate α from Eq. (30).
- STEP 6. Check the limiting criterion for (u, v, α) . If the criterion is not satisfied, go to Step 4.

In the remainder of this section, we show that the QFDTWCM with $m - 1 \rightarrow +0$ reduces to BFDTWCM and QFDTWCM with $\lambda \rightarrow +\infty$ reduces to KLFDTWCM.

The third step in the QFDTWCM approach is exactly equal to that of the BFDTWCM because Eq. (28) is identical to Eq. (28). In the fourth step of the QFDTWCM, the u value in Eq. (29) reduces to that in Eq. (26) of the BFDTWCM as

$$\begin{aligned}
 & \frac{1}{\sum_{j=1}^C \frac{\alpha_j}{\alpha_i} \left(\frac{1-\lambda(1-m)DTW_{j,k}}{1-\lambda(1-m)DTW_{i,k}} \right)^{1/(1-m)}} \\
 &= \frac{\alpha_i (1/\lambda - (1-m)DTW_{i,k})^{1/(1-m)}}{\sum_{j=1}^C \alpha_j (1/\lambda - (1-m)DTW_{j,k})^{1/(1-m)}} \\
 &\rightarrow \frac{\alpha_i (-(1-m)DTW_{i,k})^{1/(1-m)}}{\sum_{j=1}^C \alpha_j (-(1-m)DTW_{j,k})^{1/(1-m)}} \\
 & \quad (\text{with } \lambda \rightarrow +\infty) \\
 &= \frac{(m-1)\alpha_i (DTW_{i,k})^{1/(1-m)}}{(m-1) \sum_{j=1}^C \alpha_j (DTW_{j,k})^{1/(1-m)}} \\
 &= \frac{\alpha_i (DTW_{i,k})^{1/(1-m)}}{\sum_{j=1}^C \alpha_j (DTW_{j,k})^{1/(1-m)}} \\
 &= \frac{1}{\sum_{j=1}^C \frac{\alpha_j}{\alpha_i} \left(\frac{DTW_{j,k}}{DTW_{i,k}} \right)^{1/(1-m)}}. \tag{31}
 \end{aligned}$$

In the fifth step of the QFDTWCM, the α value in Eq. (30) is reduces to that in Eq. (27) of the BFDTWCM as

$$\begin{aligned}
 & \frac{1}{\sum_{j=1}^C \left(\frac{\sum_{k=1}^N (u_{j,k})^m (1-\lambda(1-m)\text{DTW}_{j,k})}{\sum_{k=1}^N (u_{i,k})^m (1-\lambda(1-m)\text{DTW}_{i,k})} \right)^{1/m}} \\
 &= \frac{\left(\sum_{k=1}^N (u_{i,k})^m (1/\lambda - (1-m)\text{DTW}_{i,k}) \right)^{1/m}}{\sum_{j=1}^C \left(\sum_{k=1}^N (u_{j,k})^m (1/\lambda - (1-m)\text{DTW}_{j,k}) \right)^{1/m}} \\
 &\rightarrow \frac{\left(\sum_{k=1}^N (u_{i,k})^m (-(1-m)\text{DTW}_{i,k}) \right)^{1/m}}{\sum_{j=1}^C \left(\sum_{k=1}^N (u_{j,k})^m (-(1-m)\text{DTW}_{j,k}) \right)^{1/m}} \\
 & \quad (\text{with } \lambda \rightarrow +\infty) \\
 &= \frac{(m-1)^{1/m} \left(\sum_{k=1}^N (u_{i,k})^m (\text{DTW}_{i,k}) \right)^{1/m}}{(m-1)^{1/m} \sum_{j=1}^C \left(\sum_{k=1}^N (u_{j,k})^m (\text{DTW}_{j,k}) \right)^{1/m}} \\
 &= \frac{\left(\sum_{k=1}^N (u_{i,k})^m (\text{DTW}_{i,k}) \right)^{1/m}}{\sum_{j=1}^C \left(\sum_{k=1}^N (u_{j,k})^m (\text{DTW}_{j,k}) \right)^{1/m}} \\
 &= \frac{1}{\sum_{j=1}^C \left(\frac{\sum_{k=1}^N (u_{j,k})^m \text{DTW}_{j,k}}{\sum_{k=1}^N (u_{i,k})^m \text{DTW}_{i,k}} \right)^{1/m}}. \tag{32}
 \end{aligned}$$

From the above discussion, we can conclude that the QFDTWCM with $\lambda \rightarrow +\infty$ reduces to the BFDTWCM.

The third step of the QFDTWCM with $m = 1$ is obviously equal to the third step of the KLFDTWCM because Eq. (28) with $m = 1$ is identical to Eq. (11). In the fourth step of the QFDTWCM, the u value in Eq. (29) reduces to that in Eq. (23) of the KLFDTWCM as

$$\begin{aligned}
 & (1 - \lambda(1 - m)\text{DTW}_{i,k})^{1/(1-m)} \\
 & \rightarrow \exp(-\lambda(\text{DTW}_{i,k})) \quad (\text{with } m = 1). \tag{33}
 \end{aligned}$$

The fifth step of the QFDTWCM reduces to that of the KLFDTWCM because

$$\begin{aligned}
 &= \frac{1}{\sum_{j=1}^C \left(\frac{\sum_{k=1}^N (u_{j,k})^m (1-\lambda(1-m)\text{DTW}_{j,k})}{\sum_{k=1}^N (u_{i,k})^m (1-\lambda(1-m)\text{DTW}_{i,k})} \right)^{1/m}} \\
 &\rightarrow \frac{1}{\sum_{j=1}^C \sum_{k=1}^N \frac{u_{j,k}}{u_{i,k}}} \quad (\text{with } m = 1) \\
 &= \frac{\sum_{k=1}^N u_{i,k}}{\sum_{j=1}^C \sum_{k=1}^N u_{j,k}} \\
 &= \frac{\sum_{k=1}^N u_{i,k}}{N}. \tag{34}
 \end{aligned}$$

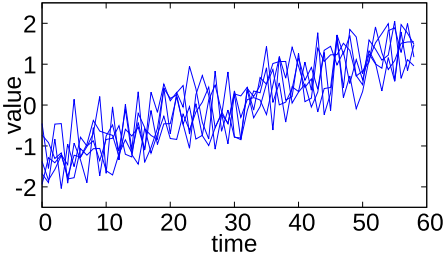


Fig. 1. Sample data group1

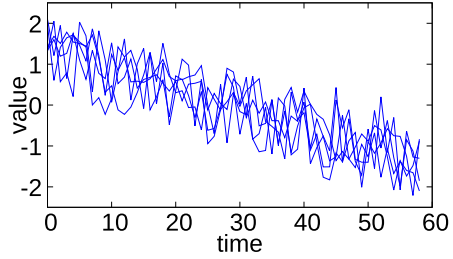


Fig. 2. Sample data group2

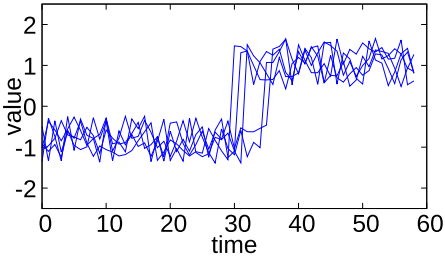


Fig. 3. Sample data group3

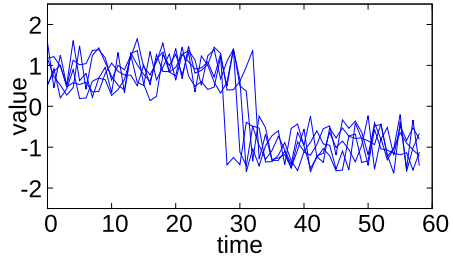


Fig. 4. Sample data group4

From the above discussion, we can conclude that the QFDTWCM with $m-1 \rightarrow 0$ reduces to the KLFDTWCM.

As shown herein, the proposed QFDTWCM includes both the BFDTWCM and KLFDTWCM. Thus, the QFDTWCM is a generalization of the BFDTWCM as well as KLFDTWCM.

4 Numerical Experiments

This section presents some numerical examples based on one artificial dataset. The example compares the characteristic features of the proposed clustering algorithm (Algorithm 4) with those of other algorithms (Algorithms 2 and 3) for an artificial dataset, as shown in Figs. 1, 2, 3 and 4 for four clusters ($C = 4$), with each clusters containing five objects ($N = 4 \times 5 = 20$).

The initialization step assigns the initial memberships according to the actual class labels. All three proposed methods with various fuzzification parameter values were able to classify the data adequately, and the obtained membership values are shown in Tables 1, 2, 3, 4, 5, 6, 7, 8 and 9. Tables 1 and 2 show that for the BFDTWCM, when the fuzzification parameter m is larger, the membership values are fuzzier. Tables 3 and 4 show that for the KLFDTWCM, when the fuzzification parameter λ is smaller, the membership values are fuzzier. Tables 5 and 6

Table 1. Sample data memberships of the BFDTWCM, $m = 1.001$

Cluster	Group			
	1	2	3	4
1	1.00	0.00	0.00	0.00
2	0.00	1.00	0.00	0.00
3	0.00	0.00	1.00	0.00
4	0.00	0.00	0.00	1.00

Table 2. Sample data memberships of the BFDTWCM, $m = 1.35$

Cluster	Group			
	1	2	3	4
1	0.77	0.01	0.10	0.00
2	0.01	0.66	0.00	0.11
3	0.21	0.01	0.89	0.00
4	0.01	0.32	0.00	0.89

Table 3. Sample data memberships of the KLFDTWCM, $\lambda = 1.5$

Cluster	Group			
	1	2	3	4
1	0.84	0.00	0.06	0.00
2	0.00	0.70	0.00	0.06
3	0.15	0.00	0.94	0.00
4	0.00	0.30	0.00	0.94

Table 4. Sample data memberships of the KLFDTWCM, $\lambda = 100$

Cluster	Group			
	1	2	3	4
1	1.00	0.00	0.00	0.00
2	0.00	1.00	0.00	0.00
3	0.00	0.00	1.00	0.00
4	0.00	0.00	0.00	1.00

Table 5. Sample data memberships of the QFDTWCM, $(m, \lambda) = (1.2, 3)$

Cluster	Group			
	1	2	3	4
1	0.78	0.00	0.10	0.00
2	0.00	0.61	0.00	0.09
3	0.21	0.00	0.90	0.00
4	0.00	0.39	0.00	0.91

Table 6. Sample data memberships of the QFDTWCM, $(m, \lambda) = (1.001, 3)$

Cluster	Group			
	1	2	3	4
1	0.99	0.00	0.00	0.00
2	0.00	0.98	0.00	0.01
3	0.01	0.00	1.00	0.00
4	0.00	0.02	0.00	0.99

show that for the QFDTWCM, when the fuzzification parameter m is larger, the membership values are fuzzier. Tables 5 and 7 show that for the QFDTWCM, when the fuzzification parameter λ is smaller, the membership values are fuzzier. Furthermore, Tables 6 and 8 show that the QFDTWCM with large values of λ produces results similar to those of the KLFDTWCM, and Tables 7 and 9 show that the QFDTWCM with smaller values of m produces results similar to those of the BFDTWCM. These results indicate that the QFDTWCM combines the features of both BFDTWCM and KLFDTWCM.

Table 7. Sample data memberships of the QFDTWCM, $(m, \lambda) = (1.2, 100)$

Cluster	Group			
	1	2	3	4
1	0.95	0.00	0.03	0.00
2	0.00	0.90	0.00	0.04
3	0.05	0.00	0.97	0.00
4	0.00	0.10	0.00	0.96

Table 8. Sample data memberships of the KLFDTWCM, $\lambda = 3$

Cluster	Group			
	1	2	3	4
1	0.99	0.00	0.00	0.00
2	0.00	0.98	0.00	0.01
3	0.01	0.00	1.00	0.00
4	0.00	0.02	0.00	0.99

Table 9. Sample data memberships of the BFDTWCM, $m = 1.2$

Cluster	Group			
	1	2	3	4
1	0.95	0.00	0.03	0.00
2	0.00	0.90	0.00	0.04
3	0.05	0.00	0.97	0.00
4	0.00	0.10	0.00	0.96

5 Conclusion

This work, propose three fuzzy clustering algorithms for classifying time-series data. The theoretical results indicate that the QFDTWCM approach reduces to the BFDTWCM as $m - 1 \rightarrow +0$ and to the KLFDTWCM as $\lambda \rightarrow +\infty$. Numerical experiments were performed on an artificial dataset to substantiate these properties.

In the future work, these proposed algorithms will be applied to real datasets.

References

1. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)
2. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
3. Miyamoto, S., Mukaidono, M.: Fuzzy c-means as a regularization and maximum entropy approach. In: Proceedings of the 7th International Fuzzy Systems Association World Congress (IFSA 1997), vol. 2, pp. 86–92 (1997)
4. Miyamoto, S., Kurosawa, N.: Controlling cluster volume sizes in fuzzy c-means clustering. In: Proceedings of the SCIS&ISIS2004, pp. 1–4 (2004)
5. Ichihashi, H., Honda, K., Tani, N.: Gaussian mixture PDF approximation and fuzzy c-means clustering with entropy regularization. In: Proceedings of the 4th Asian Fuzzy System Symposium, pp. 217–221 (2000)

6. Miyamoto, S., Ichihashi, H., Honda, K.: Algorithms for Fuzzy Clustering. Springer, Heidelberg (2008)
7. Chernoff, H.: A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Ann. Math. Statist.* **23**, 493–507 (1952)
8. Petitjean, F., Ketterlin, A., Gancarski, P.: A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recogn.* **44**, 678–693 (2011)