


Fuzzy c-Means Directional Clustering (FCMDC) algorithm using trigonometric approximation

Orhan KESEMEN, Özge TEZEL, Eda ÖZKUL, Buğra Kaan TİRYAKİ* 

Department of Statistics and Computer Sciences, Faculty of Science, Karadeniz Technical University, Trabzon, Turkey

Received: 20.03.2019

Accepted/Published Online: 07.08.2019

Final Version: 27.01.2020

Abstract: Cluster analysis is widely used in data analysis. Statistical data analysis is generally performed on the linear data. If the data has directional structure, classical statistical methods cannot be applied directly to it. This study aims to improve a new directional clustering algorithm which is based on trigonometric approximation. The trigonometric approximation is used for both descriptive statistics and clustering of directional data. In this paper, the fuzzy clustering algorithms (FCD and FCM4DD) improved for directional data and the proposed method are carried out on some numerical and real data examples, and the simulation results are presented. Consequently, these results indicate that the fuzzy c-means directional clustering algorithm gives the better results from the points of the mean square error and the standard deviation for cluster centers.

Key words: Directional data, fuzzy directional clustering, trigonometric mean, angular distance

1. Introduction

Clustering analysis is one of the popular topics of recent days. Clustering methods are commonly used in many fields such as statistical analysis, pattern recognition, data mining, and machine learning [1–6]. Many clustering algorithms have been developed in the literature to separate data into groups. Most of these algorithms work on linear data systems, while others work on directional data. Clustering algorithms working on linear data systems cannot be applied directly on directional data by their nature. Otherwise, this situation may lead to dilemmas [7]. The most important reason for this is having a modular structure of the directional data. Therefore, if directional data are characterized in the range $[0, 2\pi)$, they are continuous at the points (0) ; if directional data are characterized in the range $[-\pi, \pi)$, they are continuous at the points $(-\pi)$. Although the directional data are continuous at these points, they show discontinuities in mathematical operations. Two commonly used directional fuzzy clustering methods have been presented to overcome these discontinuities. The first of them is the fuzzy c-directions (FCD) algorithm developed by [8]. This algorithm is a model-based clustering method that uses von Misses distribution [9]. The biggest deficiency of this method is that the membership values exceed the definition range of $[0, 1]$. The other method is the fuzzy c-means clustering for directional data (FCM4DD) algorithm developed by [10]. This algorithm was developed by adapting the operations in the linear data system to the directional data system.

The motivation of this study comes from the fact that contrary to the FCM4DD, the proposed method calculates the cluster centers using trigonometric approach instead of the linear approach. Moreover, the clustering which uses the trigonometric approach will have less mean square errors due to the structure of the

*Correspondence: bugrakaantiryaki@gmail.com

directional data. Therefore, in this study, the clustering is performed with a trigonometric approach which is appropriate to the structure of directional data instead of clustering which uses a linear approach such as the FCM4DD.

Analyzing directional data is as old as the subject of mathematical statistics but there are still gaps in this field. In order to fill these gaps, several studies have been presented and the directional data has been increasingly popular [3, 11–21]. Statistical studies on directional data are becoming more and more important. There are several books on the circular data analysis such as those by Batschelet [22], Fisher [23], and Mardia and Jupp [24]. Directional data have applications in many areas such as medicine, biology, geology, oceanography, and meteorology. In all studies on directional data, each data is accepted as a unit vector and the statistical calculations are made by using their trigonometric components. The main basis for achieving distinctive results with this approach is the usage of the trigonometric functions.

In this study, the trigonometric approach was used for the directional clustering. Some existing clustering algorithms which are the FCD and the FCM4DD algorithm, and the proposed fuzzy c-means directional clustering algorithm were performed on several artificial and real-life data. Then, the results were compared.

2. Directional clustering

In this study, the trigonometric approach used in the statistical calculations of the directional data was also used for the proposed directional fuzzy clustering algorithm. The trigonometric distance and the trigonometric mean were used in this clustering method [24].

Directional clustering algorithms are based on the angular distances between data as a similarity measure. The clockwise and the counter-clockwise are angularly two distances between two angles θ_a and θ_b on the circle. In general, the shortness of these two distances is preferred [24]. Ackermann [25] proposed Eq. (1) in order to remedy this confusion in the similarity measurement. The measure δ yields the smaller of the two angles between θ_a and θ_b .

$$\delta = \pi - |\pi - |\theta_a - \theta_b||. \quad (1)$$

The other similarity measure is given in Eq. (2) which was proposed by Lund [26].

$$d = 1 - \cos(\theta_a - \theta_b). \quad (2)$$

The distance d in Eq. (2) ranges from 0 to 2. Although the similarity measure given in Eq. (1) gives the real angular distance, it is inadequate in several circular processes. For this reason, generally, the equality in Eq. (2) is used in circular data processing. The following equation is used to displace the measure of the angular distance in Eq. (2) to the real distance value.

$$d = \frac{\pi}{2} [1 - \cos(\theta_a - \theta_b)]. \quad (3)$$

In the FCM4DD algorithm, the angular difference between two angles is used instead of the absolute distance when measuring the distance between two angles. The angular difference between the two angles is calculated as in Eq. (4).

$$\psi = \left(((\theta_a - \theta_b) + \pi) \bmod 2\pi \right) - \pi. \quad (4)$$

The angular difference, unlike the absolute distance, can also take minus values. If the angle (θ_a) is in front of the counter-clockwise direction, the ψ value is positive otherwise it is negative.

2.1. The fuzzy c-directions clustering algorithm

The fuzzy c-directions clustering algorithm (FCD) was presented by Yang and Pan [8]. It is a directional fuzzy clustering algorithm that allows the data point belonging to multiple clusters with different membership values in the interval [0,1]. The general assumption here is that the data set $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ is a random sample which is generated from a mixture von Mises distribution. The following equation is objective function of the FCD algorithm.

$$\beta_{m,w}(\mu, \alpha, \nu, \kappa) = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m (-\log_e 2\pi I_0(\kappa_j) + \kappa_j \cos(\theta_i - \nu_j)) + w \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \log_e a_j. \tag{5}$$

μ_{ij} is the membership value of the i^{th} data point to the j^{th} cluster. $I_0(\kappa)$ denotes the modified Bessel function of the first kind and order zero [24]. ν and κ are called the mean direction and the concentration parameter of von Mises distribution, respectively. m and w are fixed constants, m is chosen as 2 and w is chosen as 1, generally. α is the proportion value of each data point [8]. Here, α_j and μ_{ij} in this objective function are given in Eqs. (6) and (7). Here, α_j and μ_{ij} must satisfy the following two conditions:

$$\sum_{j=1}^C \alpha_j = 1. \tag{6}$$

$$\sum_{j=1}^C \mu_{ij} = 1, \quad (i = 1, 2, \dots, N). \tag{7}$$

The FCD algorithm for circular data can be given in Algorithm 1.

Algorithm 1 The FCD clustering algorithm

- Step 1. Fix $c \in [2, N)$, $m > 1$, $\varepsilon > 0$, $w > 0$.
- Step 2. Give initial membership values, randomly $\mu_{ij}^{(0)} \sim U(0, 1)$.
- Step 3. Let $t = 1$.
- Step 4. Calculate $\alpha^{(t)}$ with $\mu^{(t-1)}$ by using Eq. (8).

$$\alpha_j = \frac{\sum_{i=1}^N \mu_{ij}^m}{\sum_{j=1}^C \sum_{i=1}^N \mu_{ij}^m}, \quad (j = 1, 2, \dots, C). \tag{8}$$

- Step 5. Compute $\phi^{(t)}$ with $\mu^{(t-1)}$ by using Eq. (9),

$$\phi_j = \text{atan2} \left(\sum_{i=1}^N \mu_{ij}^m \sin(\theta_i), \sum_{i=1}^N \mu_{ij}^m \cos(\theta_i) \right), \quad (j = 1, 2, \dots, C), \tag{9}$$

The arc tangent function (atan2) is defined in the range $[-\pi, \pi)$.

- Step 6. Compute $\kappa^{(t)}$ with $\mu^{(t-1)}$, $\phi^{(t)}$ and θ by using Eq. (10),

$$\kappa_j = A^{-1} \left(\frac{(\sum_{i=1}^N \mu_{ij}^m \cos(\theta_i - \phi_j))}{\sum_{i=1}^N \mu_{ij}^m} \right), \quad (j = 1, 2, \dots, C), \tag{10}$$

$A^{-1}(x)$ can be determined from the Batschelet's table (see [23]).

Step 7. Update $\mu^{(t)}$ with $\alpha^{(t)}$, $\phi^{(t)}$, $\kappa^{(t)}$ and θ by using Eq. (11).

$$\mu_{ij} = \left(\frac{\sum_{k=1}^C (\log(2\pi I_0(\kappa_j)) - \kappa_j \cos(\theta_i - \phi_j) - w \log(\alpha_j))^{\frac{1}{m-1}}}{\sum_{k=1}^C (\log(2\pi I_0(\kappa_k)) - \kappa_k \cos(\theta_i - \phi_k) - w \log(\alpha_k))^{\frac{1}{m-1}}} \right)^{-1} \quad (i = 1, 2, \dots, N; j = 1, 2, \dots, C) \quad (11)$$

Step 8. IF $\|\mu^t - \mu^{t-1}\| < \varepsilon$, STOP

ELSE $t = t + 1$ and return to Step 4.

2.2. Fuzzy c-means clustering algorithm for directional data

The fuzzy c-means clustering algorithm for directional data (FCM4DD) algorithm was developed by adapting the fuzzy c-mean clustering algorithm for directional data. The most important point of this method is that using angular difference between two angles in Eq. (4) instead of the distance measure in Eq. (2). This method is similar to the FCM algorithm in terms of the distance measure.

Circular data is usually represented by $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ but sometimes can be given periodically by $X = x_1, x_2, \dots, x_i, \dots, x_N$. If directional data is given periodically, it can be converted to the directional form with the help of Eq. (12),

$$\theta_i = \frac{2\pi x_i}{T} - \pi \quad (12)$$

T is the period of the variable x_i .

In this study, the ranges of directional data are $[-\pi, \pi)$ or $[0, 2\pi)$. The improved clustering algorithm endeavors to minimize generalized form of the least-squared errors function in Eq. (13) [27].

$$J_m = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \left\| \left(((\theta_i - \phi_j) + \pi) \bmod 2\pi \right) - \pi \right\|^2 \quad (13)$$

m is weighting exponent parameter and generally is chosen as 2. ϕ_j is the cluster center and μ_{ij} is the membership value of the i^{th} data to the j^{th} cluster. μ_{ij} must satisfy the following conditions given in Eqs. (14)–(16).

$$\mu_{ij} \in [0, 1], \quad \forall i, j \quad (14)$$

$$\sum_{j=1}^C \mu_{ij} = 1, \quad \forall i \quad (15)$$

$$0 < \sum_{i=1}^N \mu_{ij} < N, \quad \forall j \quad (16)$$

The fuzzy c-means clustering algorithm for directional data can be given in Algorithm 2.

Algorithm 2 The FCM4DD clustering algorithm

Step 1. Fix $c \in [2, N)$, $m > 1$, $\varepsilon > 0$.

Step 2. Give initial cluster centers $\phi_j^{(0)} \sim U(0, 1)$.

Step 3. Let $t = 1$.

Step 4. Calculate the angular difference (ψ_{ij}) between two angles according to the temporary cluster center of each data and reduce to the range of $[-\pi, \pi)$ by using Eq. (17).

$$\psi_{ij} = \left(((\theta_i - \phi_j) + \pi) \bmod 2\pi \right) - \pi. \quad (17)$$

Step 5. Compute the membership values (μ_{ij}) with ψ_{ij} by using Eq. (18),

$$\mu_{ij} = \left(\sum_{k=1}^C \left(\frac{\|\psi_{ij}\|}{\|\psi_{ik}\|} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, C. \quad (18)$$

Step 6. Update the cluster centers (ϕ_j) with ψ_{ij} and μ_{ij} by using Eq. (19),

$$\phi_j^{(t+1)} = \left(\left(\phi_j^{(t)} + \frac{\sum_{i=1}^N \mu_{ij}^m \psi_{ij}}{\sum_{i=1}^N \mu_{ij}^m} + \pi \right) \bmod 2\pi \right) - \pi, \quad (j = 1, 2, \dots, C). \quad (19)$$

Step 7. IF $\|\mu^t - \mu^{t-1}\| < \varepsilon$, STOP

ELSE $t = t + 1$ and return to Step 4.

2.3. Fuzzy c-means directional clustering algorithm

The desired optimal criterion of clustering is to minimize the objective function. The fuzzy c-means directional clustering algorithm (FCMDC) minimizes the following generalized form of the least-squared errors function.

$$J_m = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \left(\sum_{d=1}^p 1 - \cos(\theta_{id} - \phi_{jd}) \right)^2, \quad 1 < m < \infty. \quad (20)$$

Here m is weighting exponent parameter, which is selected as 2, generally. μ_{ij} is the membership value of the i^{th} data to the j^{th} cluster. μ_{ij} must satisfy the conditions given in Eqs. (14)–(16) [28]. In the circumstances, the FCMDC with the help of the trigonometric distance and trigonometric mean can be given in Algorithm 3.

Algorithm 3 The FCMDC clustering algorithm

Step 1. Fix directional data (θ_i), $c \in [2, N)$, $m > 1$, $\varepsilon > 0$.

Step 2. Give initial membership values randomly $\mu_{ij}^{(0)} \sim U(0, 1)$.

Step 3. Let $t = 1$.

Step 4. Compute the cluster centers (ϕ_{jd}) with μ_{ij} by using Eq. (21),

$$\phi_{jd} = \text{atan2} \left(\sum_{i=1}^N \mu_{ij}^m \sin(\theta_{id}), \sum_{i=1}^N \mu_{ij}^m \cos(\theta_{id}) \right), \quad (j = 1, 2, \dots, C). \quad (21)$$

Step 5. Update (μ_{ij}) by using Eq. (22),

$$\mu_{ij} = \left(\sum_{k=1}^C \left(\frac{\sum_{d=1}^p 1 - \cos(\theta_{id} - \phi_{jd})}{\sum_{d=1}^p 1 - \cos(\theta_{id} - \phi_{kd})} \right)^{\frac{2}{m-1}} \right)^{-1}. \quad (22)$$

Step 6. IF $\|\mu^t - \mu^{t-1}\| < \varepsilon$, STOP
 ELSE $t = t + 1$ and return to Step 4.

3. Comparing membership functions

The performances of fuzzy clustering algorithms depend on the weighting exponent parameter (m). This parameter forms a bridge between soft clustering and hard clustering. A large weighting exponent parameter, the larger m , an object belongs to the more than one cluster and vice versa. Clustering is an optimization problem and the solution of this depends on the m parameter. In other words, the selection of the different parameter results in differences in clustering. Although there are many studies on the selection of this parameter in the literature, there is no generally accepted criterion for this. A range of m based on experience, $1 \leq m \leq 5$, was proposed by Bezdek [28]. Pal and Bezdek [29] developed a heuristic rule for optimal selection of the parameter m . The value of m was limited to $[1.5, 2.5]$. Generally, m parameter is chosen as 2.

In this section, the behavior of these clustering methods is investigated for $m = \{1.5, 2, 2.5\}$ and cluster centers $v = \{-\frac{2\pi}{3}, -\frac{\pi}{6}, \frac{\pi}{2}\}$. The values of $\mu_1(\theta)$, $\mu_2(\theta)$, and $\mu_3(\theta)$ in the graphs represent the membership values for the first, second, and third clusters, respectively.

The change of FCD algorithm membership values according to the weighting exponent parameter (m) is shown in Figure 1. The parameter values for this algorithm are determined as $\alpha = \{0.4, 0.3, 0.3\}$ and $\kappa = \{7, 15, 3\}$.

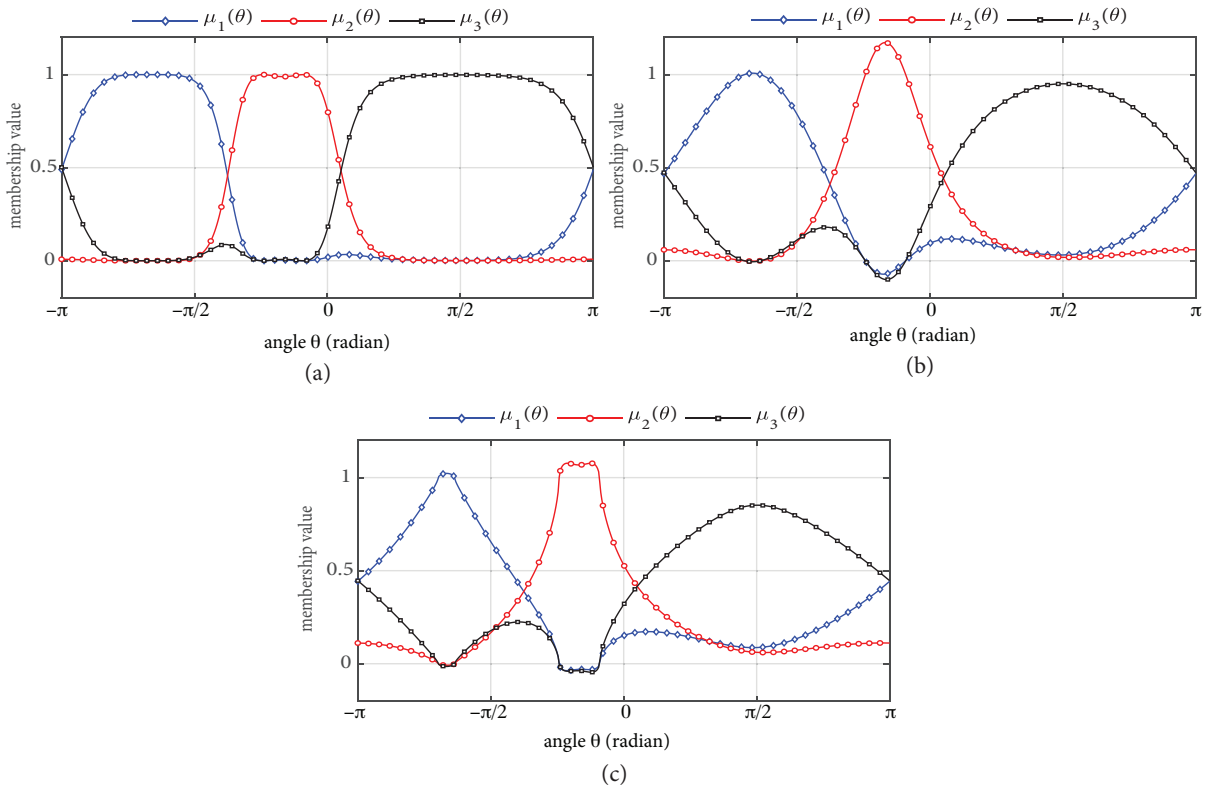


Figure 1. Calculated membership functions by FCD clustering algorithm; (a) Membership functions by FCD using $m = 1.5$, (b) Membership functions by FCD using $m = 2$, (c) Membership functions by FCD using $m = 2.5$.

The membership values calculated by the FCD algorithm with parameter $m = 1.5$ are shown Figure 1a. When the membership values are examined, it is seen that each membership function has reached 1 value within its own cluster center and has reached 0 value within the other clusters. When the membership functions are investigated graphically, it is seen that the membership values are outside of the interval $[0, 1]$ when the weighting exponent parameters are $m = \{2, 2.5\}$ in the Figures 1b and 1c. Therefore, membership values obtained by using FCD algorithm do not provide Eq. (15).

Figure 2 shows the change of the FCM4DD algorithm membership values according to the weighting exponent parameter (m).

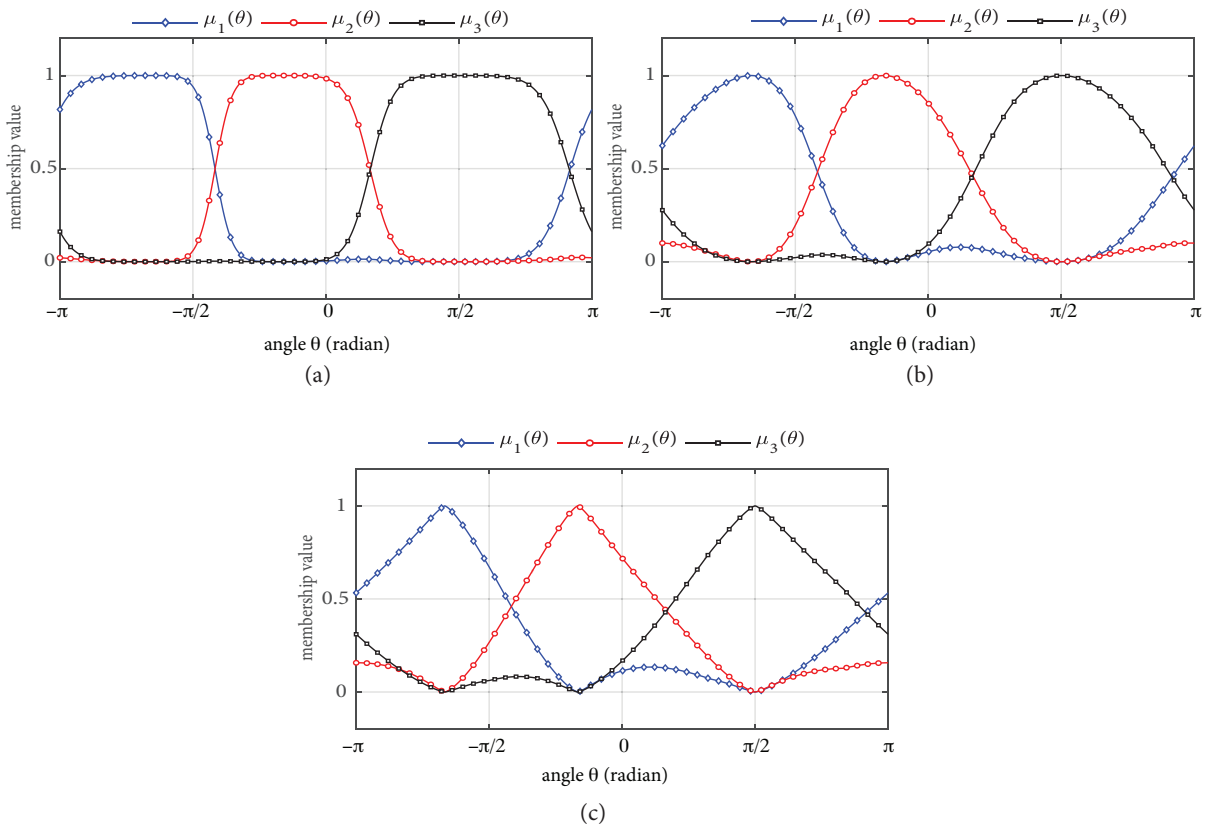


Figure 2. Calculated membership functions by FCM4DD clustering algorithm; (a) Membership functions by FCM4DD using $m = 1.5$, (b) Membership functions by FCM4DD using $m = 2$, (c) Membership functions by FCM4DD using $m = 2.5$.

According to membership value in Figure 2, it is seen that each membership function’s value is 1 within its own cluster center and 0 in the other cluster centers. Furthermore, the membership values range from 0 to 1 for each m . Depending on the increase in the value m , membership functions have changed from trapezoid membership function to triangle membership function.

Figure 3 shows the change of the proposed FCMDC algorithm according to the weighting exponent parameter (m).

According to membership value in Figure 3, it is seen that each membership function’s value is 1 within its own cluster center and 0 in the other cluster centers. In addition, the membership values range between 0 and 1 for each m .

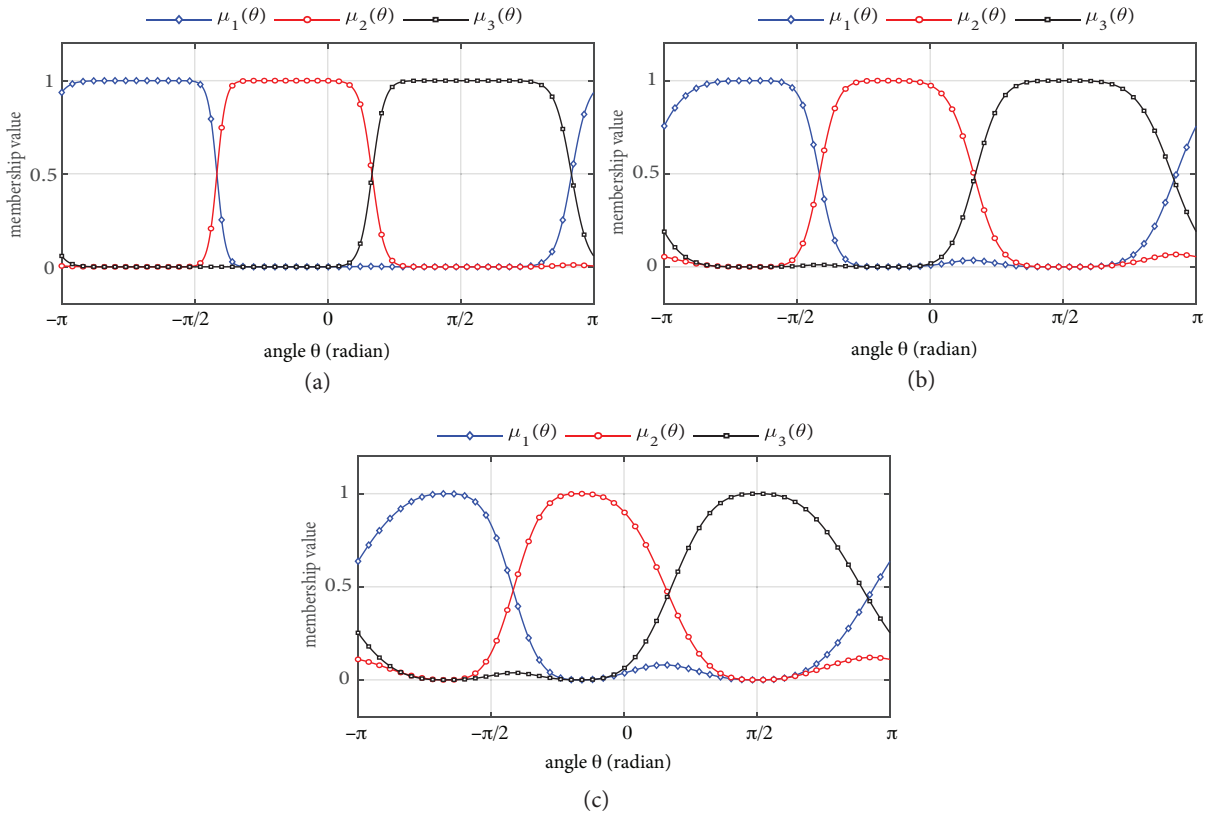


Figure 3. Calculated membership functions by FCMDC clustering algorithm; (a) Membership functions by FCMDC using $m = 1.5$, (b) Membership functions by FCMDC using $m = 2$, (c) Membership functions by FCMDC using $m = 2.5$.

4. Experiments

In order to compare the performance of the FCMDC algorithm, it has been tested on some numerical examples with the FCD and FCM4DD algorithms.

Example 1 The turtle data set is about the directions of 76 turtles which is given by Stephens [30]. The directions of the turtles were recorded after laying their eggs, which is shown in Table 1.

Table 1. Directions of 76 turtles after laying eggs.

8	9	13	13	14	18	22	27	30	34
38	38	40	44	45	47	48	48	48	48
50	53	56	57	58	58	61	63	64	64
64	65	65	68	70	73	78	78	78	83
83	88	88	88	90	92	92	93	95	96
98	100	103	106	113	118	138	153	153	155
204	215	223	226	237	238	243	244	250	251
257	268	285	319	343	350				

The data has two cluster centers that are about 60° and 240° , which is supported by the previous studies [21].

In order to cluster the turtle data set, FCD, FCM4DD, and FCMDC algorithms are used. The cluster centers are estimated by repeating these algorithms 10 times; subsequently the estimated cluster centers are compared with the real cluster centers to evaluate the performances of these algorithms. The comparison results are given in Table 2.

Table 2. Comparison of algorithms for Example 1.

Algorithms	FCD	FCM4DD	FCMDC
$\phi_1(deg)$	63.0749	62.5633	62.3815
$\phi_2(deg)$	238.3317	239.0884	239.6002
MSE(rad)	0.00372298	0.00225552	0.00253515
STD(rad)	(0.0567, 0.0305)	(0.0474, 0.0163)	(0.0524, 0.0085)

In Table 2, MSE is the mean square error and STD is the standard deviation for cluster centers. When the FCMDC algorithm is compared with the FCD and FCM4DD, we find that the FCM4DD and the FCMDC clustering algorithms give almost the same results according to the MSE values. Thus, the FCMDC is a consistent clustering algorithm like the FCM4DD. According to the standard deviation for ϕ (STD), it is seen that the STD value of the FCMDC algorithm for ϕ_1 is only less than the FCD algorithm. However, we find that the FCMDC has the smallest standard deviation for ϕ_2 among the three algorithms.

Considering the maximum crisp membership values, it is seen that 61 turtles move in the direction of 60° and that 15 turtles move in the direction of 240° when the FCD algorithm is used. Fifty-nine turtles move in the direction of the 60° and 17 turtles move in the direction of the 240° when the FCM4DD and the FCMDC algorithms are used in all trials. The FCMDC and the FCM4DD algorithms separate the same number of turtles into two clusters.

Example 2 We generate 100 simulation data from the mixture von Mises distributions whose probability density function is $f(\theta) = 0.3VM\left(\theta; \frac{\pi}{2}, 25\right) + 0.3VM\left(\theta; \pi, 25\right) + 0.4VM\left(\theta; \frac{\pi}{3}, 25\right)$ using Best and Fisher's simulation method [31] for the performance comparisons of the FCD, FCM4DD, and FCMDC algorithms. The cluster centers are estimated by repeating these algorithms 100 times and the means of estimated cluster centers ($\phi_1; \phi_2; \phi_3$) for each algorithm were given in Table 3.

Table 3. Comparison of algorithms for Example 2.

Algorithms	FCD	FCM4DD	FCMDC
$\phi_1(deg)$	98.1729	89.3028	89.3489
$\phi_2(deg)$	204.8087	180.7776	180.7660
$\phi_3(deg)$	299.9412	299.9722	299.9743
MSE(rad)	0.20783142	0.00033246	0.00030804
STD(rad)	(0.3739, 0.9879, 0.0421)	(0.0307, 0.0369, 0.0413)	(0.0311, 0.0369, 0.0418)

In this example, when these algorithms were compared according to the MSE values, the FCMDC algorithm showed better results than the other algorithms. It is seen that the FCM4DD and the FCMDC clustering algorithms give almost the same results according to the STD values. On the other hand, the FCD algorithm is an unstable method because this algorithm gives very bad results according to the standard deviation for cluster centers.

Example 3 In this example, sudden infant death syndrome (SIDS) data set which is given by Mooney et al. [1] is used to compare the clustering algorithms. The data set indicates the numbers of sudden infant deaths, according to the month of birth, month of death, age at death, and gender in the UK in the years 1983–1998. The SIDS data for the year 1998 can be fitted by a mixture of two von Mises distribution which is periodic data, so that we may apply the directional clustering algorithms to these data. The data were month-corrected to 31 days by multiplying February by 1.097 and the 30-day months by 1.033. The numbers and the corrected numbers of the SIDS cases for the year 1998 are given in Table 4.

Table 4. The numbers and the corrected numbers of SIDS cases for the year 1998.

1998	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Numbers	40	28	25	25	29	32	25	20	26	40	42	63
Corrected numbers	40	31	25	26	29	33	25	20	27	40	43	63

The numbers of cases given per month are converted into circular data by dividing 360° into 12. The months of the 1 year can be divided into the angular range of a circle into 12 intervals ($0^\circ, 360^\circ$). For example, the circular range of the January is $(0^\circ, 30^\circ)$, the circular range of the February is $(30^\circ, 60^\circ)$ and so on. Random numbers are generated from uniform distribution as the number of cases for each month in the defined circular range. Thus, all data has been converted into the circular data in the range of $(0^\circ, 360^\circ)$. The FCD, FCM4DD, and FCMDC circular clustering algorithms are applied to these circular data. The cluster centers are estimated by repeating these algorithms 10 times; subsequently the estimated cluster centers are compared with the real cluster centers to evaluate the performances of these algorithms. The comparison results are given in Table 5.

Table 5. Comparison of algorithms for Example 3.

Algorithms	FCD	FCM4DD	FCMDC
$\phi_1(deg)$	168.8667	150.7914	150.8867
$\phi_2(deg)$	336.9319	334.2690	335.2671
MSE(rad)	0.0926	0.0033	0.0016
STD(rad)	(1.0572, 0.0734)	(0.0503, 0.0672)	(0.0645, 0.0549)

The data has two cluster centers that are about 154.41° and 340.34° , which is supported by the previous studies [1].

In this example, when these algorithms were compared according to the MSE values, the FCMDC algorithm showed better results than the other algorithms. The estimate of the two cluster centers ϕ_1 and ϕ_2 are 150.8867° and 335.2671° by using the FCMDC algorithm. Thus, we say that there are two peaks of SIDS cases for the year 1998. One of them is between May and June, the other is between November and December. In other words, this study has given similar results as the previous studies [21]. According to standard deviation for ϕ (STD), it is seen that the STD value of the FCMDC algorithm for ϕ_1 is only less than the FCD algorithm. However, we find that the FCMDC has the smallest standard deviation for ϕ_2 among the three algorithms.

Example 4 We generate 100 simulation data from the mixture triangular distributions whose probability density function is $f(\theta) = 0.33\Lambda\left(\frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}\right) + 0.33\Lambda\left(\frac{2\pi}{3}, \pi, \frac{7\pi}{6}\right) + 0.34\Lambda\left(\frac{7\pi}{6}, \frac{5\pi}{3}, \frac{11\pi}{6}\right)$ using Best and Fisher’s simulation method [31] for the performance comparisons of the FCD, FCM4DD, and FCMDC algorithms. The cluster centers are estimated by repeating these algorithms 100 times and the means of estimated cluster centers ($\phi_1; \phi_2; \phi_3$) for each algorithm were given in Table 6.

Table 6. Comparison of algorithms for Example 4.

Algorithms	FCD	FCM4DD	FCMDC
$\phi_1(deg)$	101.1628	92.0186	91.9506
$\phi_2(deg)$	166.3027	175.3544	175.0455
$\phi_3(deg)$	270.2519	285.7891	286.0619
MSE(rad)	0.36468053	0.06933259	0.06781499
STD(rad)	(0.5420, 0.4912, 0.812)	(0.0541, 0.0973, 0.2787)	(0.0535, 0.1026, 0.2764)

In this example, when these algorithms were compared according to the MSE values, the FCMDC algorithm showed better results than the other algorithms. According to the standard deviation for ϕ (STD), it is seen that the STD values of the FCMDC and FCM4DD algorithms for ϕ_2 are similar. However, we find that the FCMDC has the smallest standard deviation for ϕ_1 and ϕ_3 among the three algorithms.

Example 5 In this example, two different up-to-date wind direction data are used. Firstly, 285 diary wind direction data were taken for the 01/Jan/2016-12/Oct/2016 at a meteorology station in Somió, Spain. It is situated at a latitude of $43^\circ 32' 17'' N$, a longitude of $5^\circ 37' 26'' W$ and 30 m above sea level. The data are available from InfoMet ¹. Wind directions are measured from 0° to 360° . The optimal number of clusters is determined as 2 by using validity indices for directional data [32]. The FCD, FCM4DD, and FCMDC circular clustering algorithms are applied to the Somió wind direction data. The cluster centers are estimated by repeating these algorithms 10 times. The estimated cluster centers ($\phi_1; \phi_2$) for each algorithm were given in Table 7.

Table 7. Comparison of Algorithms for Somió Data.

Algorithms	FCD	FCM4DD	FCMDC
$\phi_1(deg)$	57.1942	56.1258	56.1648
$\phi_2(deg)$	260.6955	255.6693	257.2199

The proposed method gives almost the same results as the FCM4DD. These algorithms calculated the wind direction centers as nearly northeast and southwest. Moreover, the FCD algorithm calculates the wind centers same as the others but it is an unstable clustering method in terms of membership values.

Secondly, Bodrum-Milas airport is considered for exemplifying the analysis of wind data. For 01/Jan/2019–20/May/2019, 5978 wind directions were examined. The airport is situated at a latitude of $37^\circ 15' 1.2'' N$, a longitude of $27^\circ 39' 30.59'' E$ and 6 m above sea level. The data are available from Iowa State University ². Wind directions are measured from 0° to 360° and recorded per approximately 20 min. The optimal number of clusters is determined as 2 by using validity indices for directional data [32]. The FCD, FCM4DD, and FCMDC circular clustering algorithms are applied to the Bodrum-Milas airport data. The cluster centers are estimated by repeating these algorithms 10 times. The estimated cluster centers ($\phi_1; \phi_2$) for each algorithm were given in Table 8.

These algorithms calculated the wind direction centers as nearly northwest and south. On the other hand, in this example, MSE and STD values cannot be calculated because the real wind direction centers are not known in the literature.

¹InfoMet (2019). Data meteorológicos [online]. Website <http://infomet.am.ub.es/clima/gijon> [accessed 22 May 2019].

²Iowa State University (2019). Iowa Environmental Mesonet [online]. Website <https://mesonet.agron.iastate.edu> [accessed 22 May 2019].

Table 8. Comparison of algorithms for Bodrum-Milas data.

Algorithms	FCD	FCM4DD	FCMDC
$\phi_1(deg)$	118.9138	114.8924	115.8384
$\phi_2(deg)$	268.8101	283.3208	277.9160

5. Conclusion

In this study, the FCMDC clustering algorithm is developed by using trigonometric approach for directional data. It is a distribution-free algorithm like the FCM4DD algorithm. In this paper, numerical examples are performed on one-dimensional directional data sets. The proposed algorithm can be performed on both circular data and N-dimensional directional data. When the results are compared, the FCM4DD and FCMDC algorithms gave almost the same results in terms of the MSE values. Experimental results demonstrated that the FCD method shows failure results in contrast to other methods. The FCM4DD and FCMDC algorithms maintain their superiority according to the standard deviation for cluster centers but the FCD algorithm does not. This is because the FCD algorithm is an unstable clustering method. When the methods are compared in terms of membership values, the FCM4DD and FCMDC algorithms provide the condition of being in the range of $[0, 1]$. However, the obtained membership values by the FCD algorithm are outside the range $[0, 1]$. The weighted exponent parameter (m) of the FCM4DD is smaller than FCMDC's for the same membership values. When the FCM4DD's weighting exponent parameter increases, the membership values of these algorithms become similar. The results show the accuracy and consistency of the FCMDC so that it is recommended as a new method for the clustering of the directional data.

References

- [1] Mooney JA, Helms PJ, Jolliffe IT. Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics & Data Analysis* 2003; 41: 505-513.
- [2] Carta JA, Bueno C, Ramirez P. Statistical modelling of directional wind speeds using mixtures of von Mises distributions: case study. *Energy Conversion and Management* 2008; 49 (5): 897-907. doi: 10.1016/j.enconman.2007.10.017
- [3] Lee A. Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics* 2010; 2 (4): 477-486. doi: 10.1002/wics.98
- [4] Abraham C, Molinari N, Servien R. Unsupervised clustering of multivariate circular data. *Statistics in Medicine* 2013; 32 (8): 1376-1382. doi: 10.1002/sim.5589
- [5] Chen L, Singh VP, Guo S, Fang B, Liu P. A new method for identification of flood seasons using directional statistics. *Hydrological Sciences Journal* 2013; 58 (1): 28-40. doi: 10.1080/02626667.2012.743661
- [6] Tasdan F, Cetin M. A simulation study on the influence of ties on uniform scores test for circular data. *Journal of Applied Statistics* 2014; 41 (5): 1137-1146. doi: 10.1080/02664763.2013.862224
- [7] Abuzaid AH. Analysis of mother's day celebration via circular statistics. *The Philippine Statistician* 2012; 61 (2): 39-52.
- [8] Yang MS, Pan JA. On fuzzy clustering of directional data. *Fuzzy Set and Systems* 1997; 91 (3): 319-326. doi: 10.1016/S0165-0114(96)00157-1
- [9] von Mises R. Uber die die "Ganzzahligkeit" der atomgewicht und verwandte fragen. *Physikal* 1918; 19: 490-500 (in German).
- [10] Kesemen O, Tezel Ö, Özkul E. Fuzzy c-means clustering algorithm for directional data (FCM4DD). *Expert Systems with Applications* 2016; 58: 76-82. doi: 10.1016/j.eswa.2016.03.034

- [11] Hamilton LJ. Characterising spectral sea wave conditions with statistical clustering of actual spectra. *Applied Ocean Research* 2010; 32 (3): 332-342. doi: 10.1016/j.apor.2009.12.003
- [12] Pewsey A, Neuhäuser M, Ruxton GD. *Circular statistics in R*. New York, NY, USA: Oxford University Press, 2013.
- [13] Bellasio R. Analysis of wind data for airport runway design. *Journal of Airline and Airport Management* 2014; 4 (2): 97-116. doi: 10.3926/jairm.26
- [14] Maruotti A. Analyzing longitudinal circular data by projected normal models: a semi-parametric approach based on finite mixture models. *Environmental and Ecological Statistics* 2016; 23 (2): 257-277. doi: 10.1007/s10651-015-0338-3
- [15] Patterson TA, Parton A, Langrock R, Blackwell PG, Thomas L et al. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis* 2017; 101 (4): 399-438. doi: 10.1007/s10182-017-0302-7
- [16] Roy A, Pal A, Garain U. JCLMM: A finite mixture model for clustering of circular-linear data and its application to psoriatic plaque segmentation. *Pattern Recognition* 2017; 66 (4): 160-173. doi: 10.1016/j.patcog.2016.12.016
- [17] Mastrantonio G, Pollice A, Fedele F. Distributions-oriented wind forecast verification by a hidden Markov model for multivariate circular-linear data. *Stochastic Environmental Research and Risk Assessment* 2018; 32 (1): 169-181. doi: 10.1007/s00477-017-1416-x
- [18] Luengo-Sanchez S, Larrañaga P, Bielza C. A directional-linear bayesian network and its application for clustering and simulation of neural somas. *IEEE Access* 2019; 7 (1): 69907-69921. doi: 10.1109/ACCESS.2019.2918494
- [19] Euán C, Sun Y. Directional spectra-based clustering for visualizing patterns of ocean waves and winds. *Journal of Computational and Graphical Statistics* 2019; 1-15. doi: 10.1080/10618600.2019.1575745
- [20] Mastrantonio G, Lasinio GJ, Maruotti A, Calise G. Invariance properties and statistical inference for circular data. *Statistica Sinica* 2019; 29 (1): 67-80.
- [21] Chang-Chien SJ, Hung WL, Yang MS. On mean shift-based clustering for circular data. *Soft Computing* 2012; 16 (6): 1043-1060. doi: 10.1007/s00500-012-0802-z
- [22] Batschelet E. *Circular Statistics in Biology*. New York, NY, USA: Academic Press, 1981.
- [23] Fisher NI. *Statistical Analysis of Circular Data*. New York, NY, USA: Cambridge University Press, 1993.
- [24] Mardia KV, Jupp PE. *Directional Statistics*. New York, NY, USA: Wiley, 2000.
- [25] Ackermann H. A note on circular nonparametrical classification. *Biometrical Journal* 1997; 39 (5): 577-587. doi: 10.1002/bimj.4710390506
- [26] Lund U. Cluster analysis for directional data. *Communications in Statistics-Simulation and Computation* 1999; 28 (4): 1001-1009. doi: 10.1080/03610919908813589
- [27] Höppner F, Klawonn F, Kruse R, Runkler T. *Fuzzy Cluster Analysis*. Chichester, UK: Wiley, 2000.
- [28] Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum Press, 1981.
- [29] Pal NR, Bezdek JC. On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy systems* 1995; 3 (3): 370-379. doi: 10.1109/91.413225
- [30] Stephens MA. *Techniques for Directional Data*. Stanford, CA, USA: Stanford University, 1969.
- [31] Best DJ, Fisher NI. Efficient simulation of the von mises distribution. *Journal of the Royal Statistical Society. Series C(Applied Statistics)* 1979; 28 (2): 152-157. doi: 10.2307/2346732
- [32] Kesemen O, Tezel Ö, Özkul E, Tiryaki BK, Ağayev E. A comparison of validity indices on fuzzy c-means clustering algorithm for directional data. In: *IEEE 2017 25th Signal Processing and Communications Applications Conference (SIU)*; Antalya, Turkey; 2017. pp. 1-4. doi: 10.1109/SIU.2017.7960557