A Learning-Based EM Clustering for Circular Data with Unknown Number of Clusters

Shou-Jen Chang-Chien, Wajid Ali, Miin-Shen Yang[*]

Department of Applied Mathematics, Chung Yuan Christian University, Taoyuan, Taiwan

Abstract

Clustering is a method for analyzing grouped data. Circular data were well used in various applications, such as wind directions, departure directions of migrating birds or animals, etc. The expectation & maximization (EM) algorithm on mixtures of von Mises distributions is popularly used for clustering circular data. In general, the EM algorithm is sensitive to initials and not robust to outliers in which it is also necessary to give a number of clusters a priori. In this paper, we consider a learning-based schema for EM, and then propose a learning-based EM algorithm on mixtures of von Mises distributions for clustering grouped circular data. The proposed clustering method is without any initial and robust to outliers with automatically finding the number of clusters. Some numerical and real data sets are used to compare the proposed algorithm with existing methods. Experimental results and comparisons actually demonstrate these good aspects of effectiveness and superiority of the proposed learning-based EM algorithm.

1. Introduction

Since von Mises [1] proposed a distribution on circular data, Watson and Williams [2] considered statistical inferences for von Mises distributions. Afterward, circular data were widely applied in biology, geology, medicine, oceanography, and meteorology [3-5]. Clustering is a useful tool for data mining. According to the statistical point of view, clustering methods can be generally divided into two categories. One is a model-based approach. Another is a nonparametric approach. In the model-based approach, the expectation and maximization (EM) algorithm [6-8] is the most used method. For a nonparametric approach, an objective function of dissimilarity measures is generally considered in which partitional clustering is the most popular, such as k-means [9-11], fuzzy c-means (FCM) [12-13], and possibilistic c-means (PCM) [14-15].

Generally, the EM algorithm was popularly used in the analysis of circular data based on mixtures of von Mises distributions [16-19]. However, these EM clustering algorithms are sensitive to initials and not robust to outliers with a given number of clusters a priori. In this paper, we construct a learning-based schema for EM, and then propose a learning-based EM algorithm on mixtures of von Mises distributions for clustering circular data that is free of initials and robust to outliers. The proposed clustering method can also automatically find an optimal number of clusters. We apply the proposed algorithm to real circular data and several comparisons are given to demonstrate its effectiveness and superiority.

2. Learning-Based EM Clustering for Circular Data

In Yang et al. [20], they proposed a robust EM clustering for Gaussian mixture models. The algorithm can automatically find an optimal cluster number of data sets and solve the drawbacks of EM that is sensitive to initial values. In this paper, we

---

[*] Corresponding Author. E-mail address: msyang@math.cycu.edu.tw

will modify the approach proposed by Yang et al. [20] such that it can handle circular data. We mention that Yang et al. [21] used the idea of the robust EM clustering on the mixtures of von Mises-Fisher distributions. We know that von Mises-Fisher distributions are well used to handle data higher than three dimensions, i. e. data on the unit hypersphere. However, the mixtures of von Mises distributions are well for handling circular data. In this paper, we especially focus on circular data based on the mixtures of von Mises distributions with angles measured by sine and cosine that is different from Yang et al. [21]. Suppose that the circular data set $X = \{\theta_1, \ \theta_2, \cdots, \theta_n\}$ is a random sample of size n from a mixture model

$$f(\theta; \ a, \ \phi) = \sum_{i=1}^{c} a_i f_i(\theta; \ \phi_i) \tag{1}$$

where a_i denotes the proportion of the ith class with

$$\sum_{i=1}^{c} a_i = 1 \tag{2}$$

and f_i is the pdf of ith class with parameter $\emptyset_i$. Let class variable $z = \{z_1, \ z_2, \cdots, \theta_c\}$ with $z_i = \{z_{i1}, \ z_{i2}, \cdots, z_{in}\}^T$ where $z_{ij} = z_i(\theta_j) = 1$ if θ_j arises from the ith class and $z_{ij} = 0$ if x_j arises from other class, for $i = \{1, 2, \cdots, c\}$ and $j = \{1, \ 2, \cdots, n\}$. Thus, the joint probability density of the complete data $\{ \theta_1, \ \theta_2, \cdots, \theta_n, \ z_1, \ z_2, \cdots, z_c\}$ shall be

$$f(\theta_1, \ \theta_2, \ \theta_n, z_1, \ z_2, \cdots, \ z_c; \ a, \ \phi) = \prod_{j=1}^{n} \prod_{i=1}^{c} [a_i f_i(\theta_j; \phi_i)]^{z_{ij}} \tag{3}$$

The log-likelihood for the complete data is given by

$$L_{EM}(a, \ \phi; \ \theta_1, \ \theta_2, \ \theta_n, z_1, \ z_2, \cdots, \ z_c) = \sum_{j=1}^{n} \sum_{i=1}^{c} z_{ij}(\ln a_i f_i(\theta_j; \phi_i)) \tag{4}$$

The EM algorithm consists of two steps with "Expectation", shorten by E, and "Maximum", shorten by M. In the E step, the expectation

$$E(z_{ij}|\theta_j) = a_i f_i(\theta_j|\phi_i) / \sum_{s=1}^{c} a_s f_s(\theta_j|\phi_s) \tag{5}$$

is used to substitute for missing data z_{ij}. In the M step, estimates of parameters are used by maximizing

$$E(L_{EM}) = \sum_{i=1}^{c} \sum_{j=1}^{n} E(z_{ij}|\theta_j)(\ln a_i f_i(\theta_j; \phi_i)) \tag{6}$$

with the restriction of Eq. (2).

The von Mises distribution is the most common used distribution on circular data. If we consider that the density $f_i = (\theta; \emptyset_i)$ is the von Mises distribution $VM = (v_i; \ k_i)$ with mean direction v_i and the concentration k_i with

$$f_i(\theta; \phi_i) = f_i(\theta; v_i, k_i) = (2\pi I_0(k_i))^{-1} \exp(k_i \cos(\theta - v_i)) , \ 0 \le \theta < 2\pi, 0 \le k_i < \infty \tag{7}$$

$$I_0(k_i) = (2\pi)^{-1} \int_0^{2\pi} \exp(k_i \cos(\theta - v_i) d\theta \tag{8}$$

is the modified Bessel function of order zero, then update equations for a_i, v_i, and k_i can be obtained as follows:

$$a_i = \sum_{j=1}^{n} z_i j / n \tag{9}$$

$$v_i = \tan^{-1}\left(\frac{\sum_{j=1}^{n} z_{ij} \sin(\theta_j)}{\sum_{j=1}^{n} z_{ij} \cos(\theta_j)}\right) \tag{10}$$

$$k_i = A^{-1}\left(\frac{\sum_{j=1}^{n} z_{ij} \cos(\theta_j - v_i)}{\sum_{j=1}^{n} z_{ij}}\right) \tag{11}$$

where $A^{-1}(x)$ a function that can be computed from Batschelet's table (see Fisher [3]). In order to solve drawbacks for EM, Yang et al. [20] give an entropy term of the proportion to the EM log-likelihood function. According to their idea, we propose the learning-based EM objective function for mixtures of von Mises distributions as follows:

$$J(z,\ a,\ v,\ k,\ \theta_1, \cdots,\ \theta_n) = \sum_{i=1}^{c} \sum_{j=1}^{c} z_{ij}(\ln a_i f_i(\theta_j; v_i,\ k_i)) + n\beta \sum_{i=1}^{c} \ln a_i \tag{12}$$

Thus, in the E step, the expectation $\hat{z}_{ij} = E(z_{ij}|\theta_j)$ is used to substitute for missing data z_{ij} with

$$\hat{z}_{ij} = E(z_{ij}|\theta_j) = a_i f_i(\theta_j; v_i, k_i) / \sum_{s=1}^{c} a_s f_s(\theta_j; v_s, k_s) \tag{13}$$

In the M-step, under the constraint Eq. (2), we need to maximize

$$E(J(z,\ a,\ v,\ k;\ \theta_1, \cdots,\ \theta_n)) = \sum_{i=1}^{c} \sum_{j=1}^{c} z_{ij}(\ln a_i f_i(\theta_j; v_i,\ k_i)) + n\beta \sum_{i=1}^{c} \ln a_i \tag{14}$$

Taking partial derivatives with respect to a_i, and then the following update equation is obtained

$$a_i^{(new)} = a_i^{EM} + \beta a_i^{(old)} + (\ln a_i^{(old)} - \sum_{s=1}^{c} a_s^{(old)} \ln a_s^{(old)}) \tag{15}$$

where

$$a_i^{EM} = \sum_{j=1}^{n} \hat{z}_{ij} / n \tag{16}$$

Eq. (15) provides us a method to seek the number of clusters. Because the number of clusters is not larger than n,

$$a_i^{(new)} \leq 1/n \tag{17}$$

or

$$a_i^{(new)} \leq 0 \tag{18}$$

is not reasonable. In this situation, we must discard the ith class. Thus, the new number of clusters can be obtained with

$$c^{(new)} = c^{(old)} - |\{a_i^{(new)} | a_i^{(new)} < 1/n,\ i = 1,\ 2, \cdots,\ c^{(old)}\}| \tag{19}$$

In order to satisfy the constraints

$$\sum_{k'=1}^{c^{(new)}} a_{k'} = 1 \tag{20}$$

and

$$\sum_{k'=1}^{c^{(new)}} \hat{z}_{k'j} = 1 \tag{21}$$

we adjust $a_{k'}$ and $\hat{z}_{k'j}$ by

$$a_{k'} = a_{k'} \Big/ \sum_{s=1}^{c^{(new)}} a_s \tag{22}$$

$$\hat{z}_{k'j} = \hat{z}_{k'j} \Big/ \sum_{s=1}^{c^{(t+1)}} \hat{z}_{sj} \tag{23}$$

The purpose of the parameter β is to control competition of proportions. Following Yang et al. [20], we set β as follows:

$$\beta = \min \Big\{ \frac{\sum_{i=1}^{c^{(new)}} \exp(-n|a_i^{(new)} - a_i^{(old)})}{c^{(old)}}, \quad \frac{1 - a_{(c)}^{EM}}{-a_{(c)}^{(old)} E} \Big\} \tag{24}$$

where

$$a_{(c)}^{(old)} = \max_i a_i^{(old)} \tag{25}$$

$$a_{(c)}^{EM} = \max_i a_i^{EM} \tag{26}$$

$$E = \sum_{i=1}^{c^{(old)}} a_i^{(old)} \ln a_i^{(old)} \tag{27}$$

When the cluster number c is stable, let $\beta = 0$. We also use the M-step to estimate v_i and k_i. We take partial derivatives with respect to v_i and k_i, respectively. Then, update equations are obtained as follows:

$$v_i = \tan^{-1} \Big(\sum_{j=1}^n z_{ij} \sin(\theta_j) \Big/ \sum_{j=1}^n z_{ij} \cos(\theta_j) \Big) \tag{28}$$

$$k_i = A^{-1} \Big(\sum_{j=1}^n z_{ij} \cos(\theta_j - v_j) \Big/ \sum_{j=1}^n z_{ij} \Big) \tag{29}$$

To solve the initialization problem, we let the initial number of clusters be n. It means that

$$(v_1^{(0)}, v_2^{(0)}, \cdots, v_n^{(0)}) = (\theta_1, \theta_2, \cdots, \theta_n) \tag{30}$$

$$(a_1^{(0)}, a_2^{(0)}, \cdots, a_n^{(0)}) = (1/n, 1/n, \cdots, 1/n) \tag{31}$$

We then consider the maximum likelihood estimate of the concentration of von Mises distribution as initial value $k_i^{(0)}$ with

$$k_i^{(0)} = A^{-1}(\bar{R}), \quad i = 1, \cdots, c \tag{32}$$

where

$$\bar{R} = \Big[\Big(\frac{1}{n} \sum_{j=1}^n \cos \theta j \Big)^2 + \Big(\frac{1}{n} \sum_{j=1}^n \sin \theta j \Big)^2 \Big]^{1/2} \tag{33}$$

is the sample mean resultant length. Thus, the proposed learning-based EM for circular data can be summarized as follows:

Learning-based EM algorithm for circular data

Step 1: Fix $\varepsilon > 0$. Give initials $\beta^{(0)} = 1$, $c^{(0)} = n$, $a_i^{(0)} = \frac{1}{n}$, and assign $(v_1^{(0)}, v_2^{(0)}, \cdots, v_n^{(0)}) = (\theta_1, \theta_2, \cdots, \theta_n)$.

Step 2: Compute $k_i^{(0)}$ using Eq. (32) and set $t = 1$.

Step 3: Compute $\hat{z}_{ij}^{(0)}$ with $a_i^{(0)}$, $v_i^{(0)}$, and $\kappa_i^{(0)}$ using Eq. (13).

Step 4: Compute $v_i^{(t)}$ with $\hat{z}_{i1}^{(t-1)}, \hat{z}_{i2}^{(t-1)}, \cdots, \hat{z}_{in}^{(t-1)}$ using Eq. (28).

Step 5: Update to $a_i^{(t)}$ with $\hat{z}_{i1}^{(t-1)}, \hat{z}_{i2}^{(t-1)}, \cdots, \hat{z}_{in}^{(t-1)}$ and $a_i^{(t-1)}$ using Eq. (15).

Step 6: Compute $\beta^{(t)}$ with $a_i^{(t)}$ and $a_i^{(t-1)}$ using Eq. (24).

Step 7: Update $c^{(t-1)}$ to $c^{(t)}$ by discard those clusters with $a_i^{(t)} \leq 1/n$ and adjust $a_i^{(t)}$ and $\hat{z}_{ij}^{(t-1)}$ by Eqs. (22) and (23).

　　　　IF $t \geq 60$ and $c^{(t-60)} - c^{(t)} = 0$, THEN let $\beta^{(t)} = 0$.

Step 8: Compute $\hat{k}_i^{(t)}$ with $\hat{z}_{i1}^{(t-1)}, \hat{z}_{i2}^{(t-1)}, \cdots, \hat{z}_{in}^{(t-1)}$ using Eq. (29).

Step 9: Compute $\hat{z}_{ij}^{(t)}$ with $a_i^{(t)}, v_i^{(t)},$ and $\hat{k}_i^{(t)}$ using Eq. (13).

Step 10: Compute $v_i^{(t+1)}$ with $\hat{z}_{i1}^{(t)}, \hat{z}_{i2}^{(t)}, \cdots, \hat{z}_{in}^{(t)}$ using Eq. (28).

Step 11: Compute $d(v^{(t+1)}, v^{(t)})$ in a convenient norm d.

　　　　IF $d(v^{(t+1)}, v^{(t)}) < \varepsilon$, STOP

　　　　ELSE $t = t + 1$ and return to Step 5.

3. Examples and Comparisons

In this section, we make comparisons between the proposed learning-based EM and EM algorithms on numerical and real data sets.

Example 1: In Fig. 1(a), there is a 2-cluster data set which is generated from a mixture of two von Mises distributions $0.4VM(60°, 6.5)+0.6VM(180°, 7.5)$ with 200 data points. If the angle is an observation, then x-coordinate represents cos and y-coordinate represents sin. We first implement EM with c = 2 for this data set. The clustering results of EM are shown in Fig. 1(b). The data set is well separated, and so we obtain the same clustering results of EM with most random initial values. The estimates parameters are as: $\hat{a}_1 = 0.4274, \hat{a}_2 = 0.5726, \hat{v}_1 = 61.0254°, \hat{v}_2 = 178.2695°, \hat{k}_1 = 6.54, \hat{k}_2 = 7.426$. We can see that EM has good clustering results and estimates close to the given parameters. We use the learning-based EM without a priori cluster number. Figs. 1(c)-1(f) are the states of the cluster centers (denoted by "*" symbol) at iteration = 0, 5, 10, 20, respectively. The cluster centers also imply the mean directions. We can see that the cluster number decreases from 200 to 2. Therefore, the learning-based EM can find an optimal cluster number with c^*=2. The clustering results of the learning-based EM are the same with EM (with a priori cluster number c=2), as shown in Fig. 1(b). The parameter estimates of the learning-based EM are as: $\hat{a}_1 = 0.4257, \hat{a}_2 = 0.5743, \hat{v}_1 = 60.8349°, \hat{v}_2 = 178.1072°, \hat{k}_1 = 6.54, \hat{k}_2 = 7.426$. The results are closed to those of EM. To compare the performance of the proposed learning-based EM with EM, we use the criterion of mean squared error (MSE) to evaluate the accuracy of the two algorithms. We generate 100 data sets from $0.4VM (60°, 6.5)+0.6VM(180°, 7.5)$ and computer MSEs from EM and the learning-based EM. The MSE is defined as

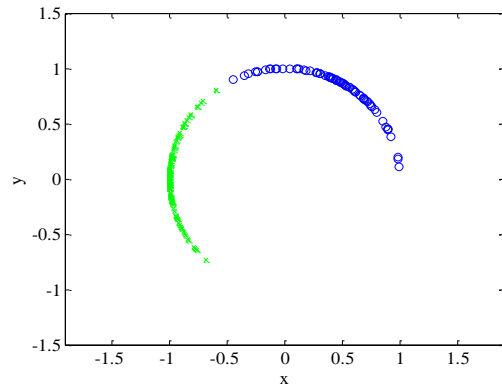$$\text{MSE} = \frac{\text{sum of SE of each data set for } v}{100}$$

where the squared error (SE) for v is defined as

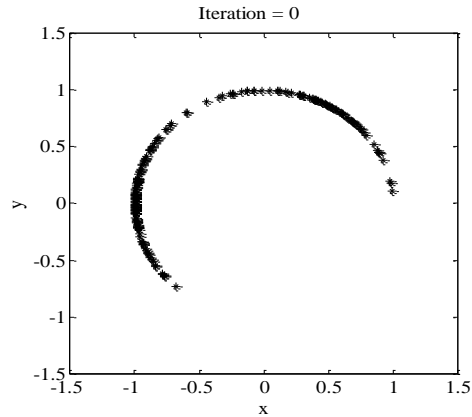$$\text{SE} = \frac{(\hat{v}_1 - 60)^2 + (\hat{v}_2 - 180)^2}{2}$$

From the 100 data sets, we obtain that the MSE value from EM is 2.3561, and the MSE value from the learning-based EM is 2.2193. The proposed learning-based EM performs better than EM.
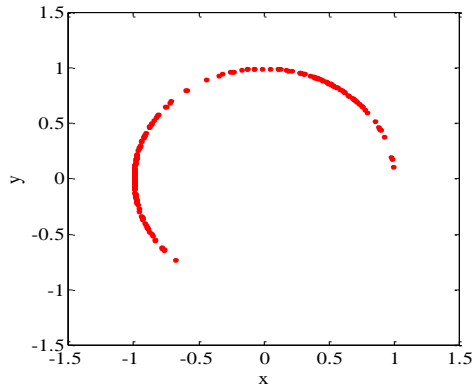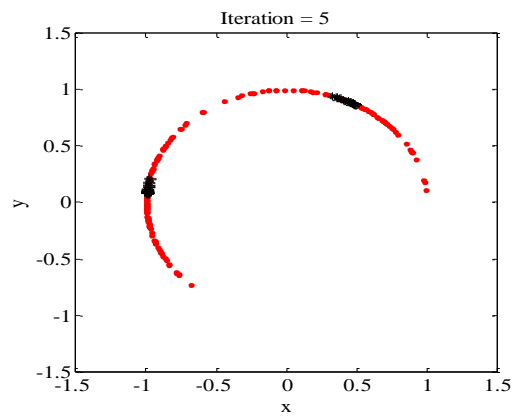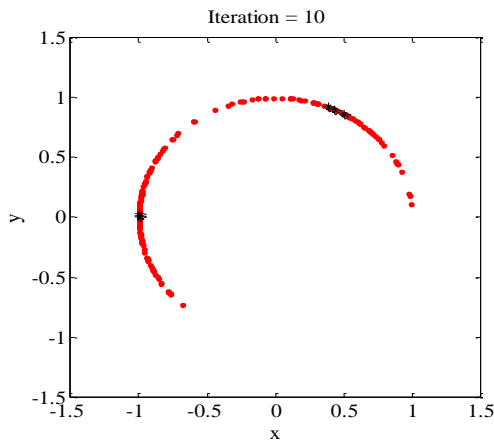
(a) 2-clusters data set

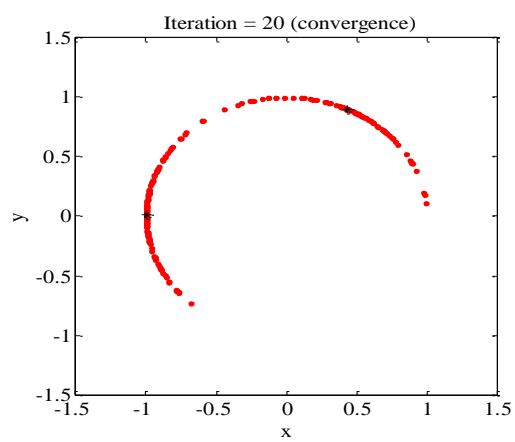(b) Clustering results of EM and the learning-ased EM

(c) The states of cluster centers from the
learning-based EM at iteration = 0

(d) The states of cluster centers from the
learning-based EM at iteration = 5

(e) The states of cluster centers from the
learning-based EM at iteration = 10

(f) The states of cluster centers from the
learning-based EM at iteration = 20

Fig. 1 2-cluster circular data set with their clustering results

Furthermore, we consider the influence of noisy points on the proposed learning-based EM and EM algorithms. We add 50 randomly noisy points which are generated from the interval [0°, 240°] into 2-cluster data set and call noisy data set, as shown in Fig. 2(a). The noisy points are denoted by the "+" symbol. We first implement EM with $c = 2$ for the noisy data set. Fig. 2(b) shows the clustering results of EM. Compare Fig. 2(b) with Fig. 1(b), we find that one of the original 200 data points is classified into another cluster. The estimates parameters are as follows: $\hat{a}_1 = 0.4223$, $\hat{a}_2 = 0.5477$, $\hat{v}_1 = 61.9240°$, $\hat{v}_2 = 178.4338°$, $\hat{k}_1 = 4.8590$, $\hat{k}_2 = 6.5400$. We also implement the learning-based EM with unknown number c of clusters for the noisy data set. The final states of the cluster centers are shown in Fig. 2(c). The learning-based EM find the optimal cluster number $c^* = 2$ after 51 iterations. The clustering results of the learning-based EM are shown in Fig. 2(d). Compare Fig. 2(d) with Fig. 1(b), we find that two clusters in the original 200 data points are the same. The estimates parameters are as

follows: $\hat{a}_1 = 0.4490$, $\hat{a}_2 = 0.5510$, $\hat{v}_1 = 61.1672°$, $\hat{v}_2 = 178.1286°$, $\hat{k}_1 = 4.8590$, $\hat{k}_2 = 5.8520$. To compare the influences of noisy points on the two algorithms, we compute the increasing rate of MSE. We generate 50 randomly noisy points from the interval [0°, 240°] for the data set. Then we compute MSEs for EM and the learning-based EM from 100 noisy data sets. The increasing rate is defined as

$$\text{increasing rate} = \frac{\text{MSE}_{\text{(the noisy data set)}} - \text{MSE}_{\text{(the original data set)}}}{\text{MSE}_{\text{(the original data set)}}} \times 100\% \tag{34}$$

Table 1 shows these MSEs with its increasing rate. From Table 1, we find the increasing rate of the learning-based EM is smaller than EM, which means that the noisy points have a smaller influence on the learning-based EM.
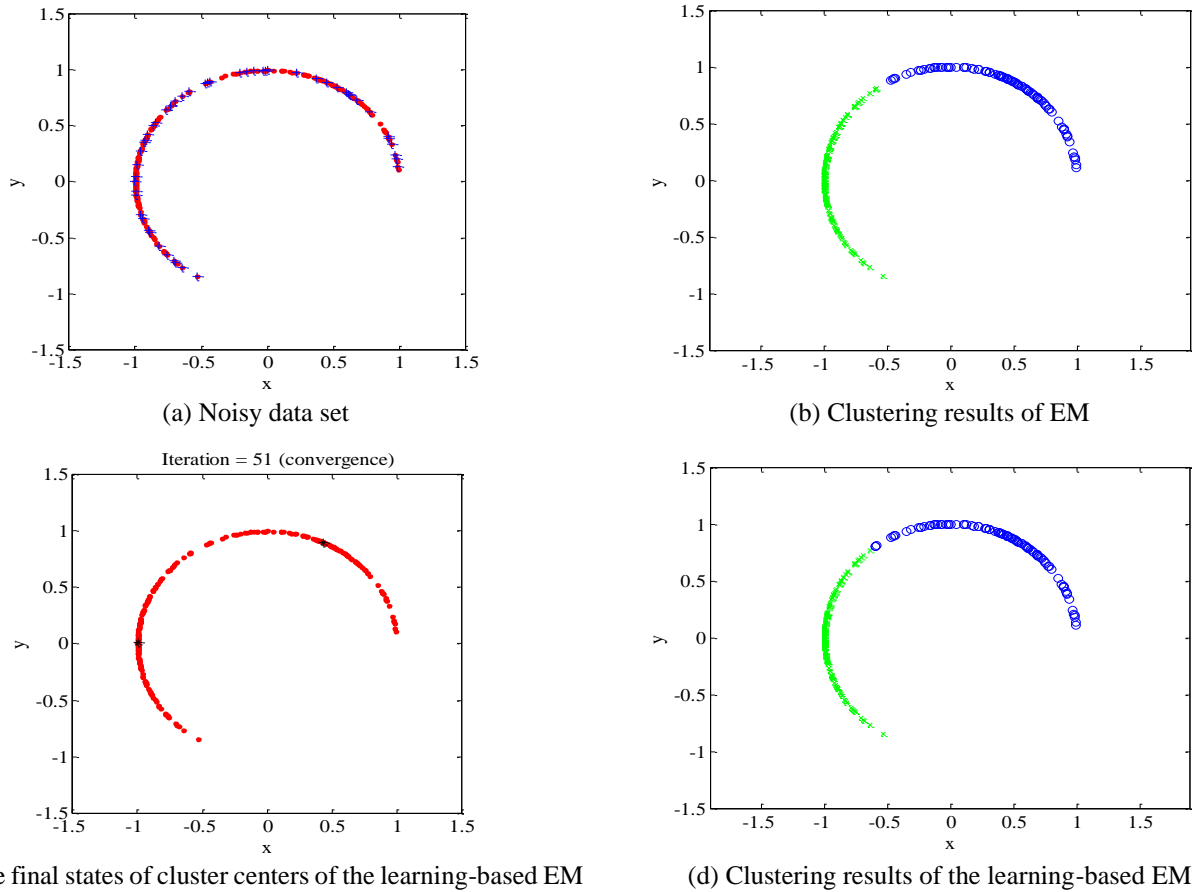

(a) Noisy data set

(b) Clustering results of EM

(c) The final states of cluster centers of the learning-based EM

(d) Clustering results of the learning-based EM

Fig. 2 2-cluster data set with 50 noisy points and their clustering results

Table 1 MSEs for EM and the learning-based EM with its increasing rates

Algorithms	EM	Learning-based EM
MSE(the original data set)	2.3561	2.2193
MSE (the noisy data set)	2.8719	2.4621
Increasing rate	21.89%	10.94%

***Example 2*:** In this example, we apply the proposed learning-based EM and EM algorithms to a real data set, sudden infant death syndrome (SIDS) data of Mooney et al. [17]. The SIDS data set consists of the SIDS cases in the UK by a month of death for the years 1983-1998, exhibiting a seasonal pattern with a winter peak. Mooney et al. [17] modified all of 12 months to 31 days, i.e. the SIDS data were month-corrected to 31 days by multiplying February by 1.097 and the 30 day months by 1.033. Then they found that the SIDS data for the year 1998 could be fitted by a mixture of two von Misses distributions. Table 2 shows the corrected number of SIDS cases for the year 1998. We transform Table 2 to circular data. We map the 12 months from 0 to 360 in degree. Every month has equal length (31 days), and so [0, 360] should be divided into 12 equal length intervals. January corresponds to the interval [0°, 30°], February corresponds to the interval [30°, 60°] and so on. Based on

frequency in each month, we generate random numbers from the uniform distribution on the corresponding interval. 402 observations and its rose diagram (circular histogram) are shown in Fig. 3(a) and Fig. 3(b), respectively. Clearly, we can see two modes on the intervals [150°, 180°] and [330°, 360°] from Fig. 3(b). We implement EM with cluster number $c=2$ for the data set in Fig. 3(a). The clustering results of EM are affected by different initials. Figs. 3(c)-3(e) are results with different initials. If we implement EM for this data set with 100 random initials, then we obtain 23 of 100 results as shown in Fig. 3(c), 72 of 100 results as shown in Fig. 3(d), and 5 of 100 results as shown in Fig. 3(e). We also implement the learning-based EM without giving c for the data set in Fig. 3(a). Fig. 3(f) shows the cluster number decreases from 402 to 2. Therefore, the learning-based EM finds an optimal cluster number $c^*=2$. The final states of cluster centers are shown in Fig. 3(g). We can see that the locations of two mean directions are consistent with two modes in Fig. 3(b). The clustering results of the learning-based EM are shown in Fig. 3(h). Obviously, the results are different from EM. Note that the learning-based EM sets all data points as initials. Thus, an initial condition is not necessary for the learning-based EM. If we perform the learning-based EM for 100 times, then 100 clustering results are the same. The estimated parameters of the learning-based EM, EM (with Fig. 3(d)) and Mooney et al. [17] are shown in Table 3. We find the estimates of the learning-based EM are closed to Mooney et al. [17]. According to mean direction estimates, we know that there are two peaks in SIDS cases for the year 1998. One peak is in early June and the other is in early December.

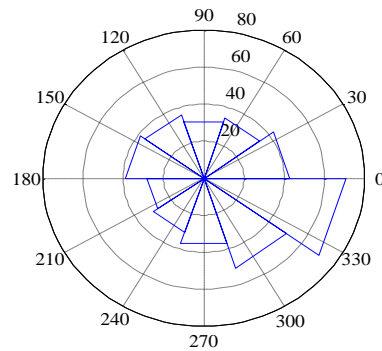Table 2 Corrected number of SIDS case for the year 1998

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
40	31	25	26	29	33	25	20	27	40	43	63	402

Table 3 Estimates of parameters for SIDS data

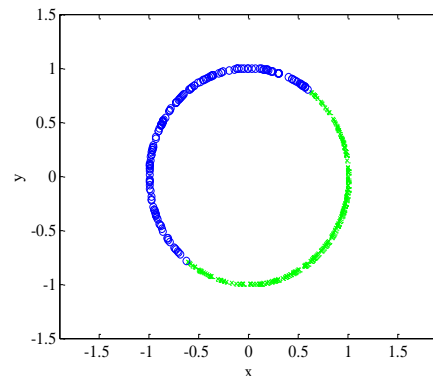| Algorithms	EM	Learning-based EM	Mooney et al. [17]
$\hat{v}_1$	153.32°	151.84°	154.48°
$\hat{v}_2$	337.66°	340.33°	340.52°
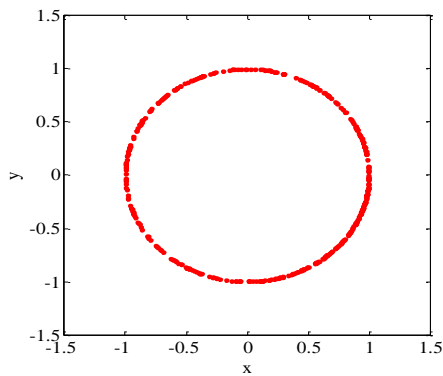
(a) The SIDS data set

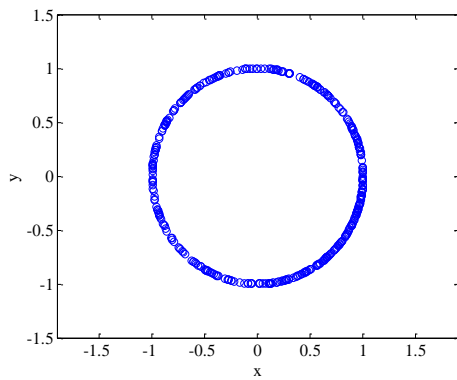(b) Rose diagram to SIDS data for the year 1998
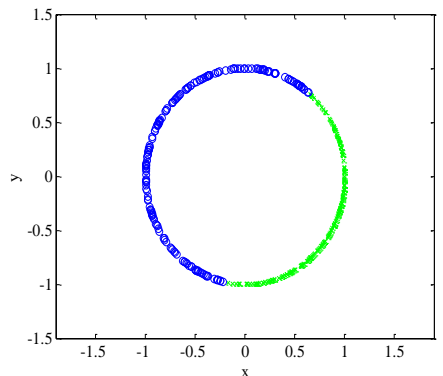
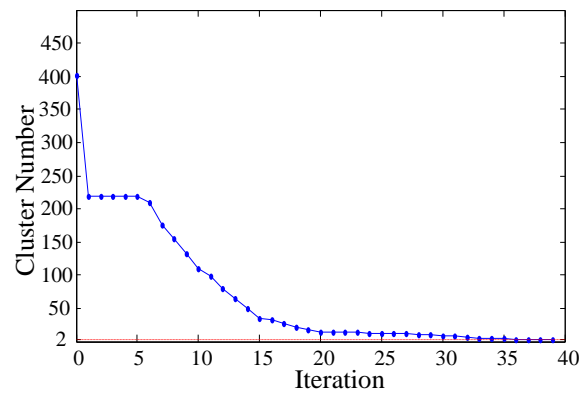(c) Clustering results of EM with different initials

(d) Clustering results of EM with different initials

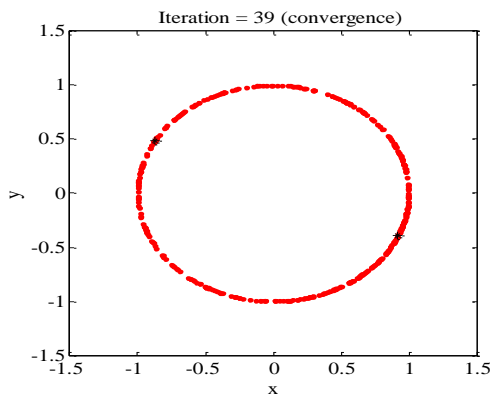Fig. 3 Clustering results of learning-based EM and EM algorithms for SIDS data set
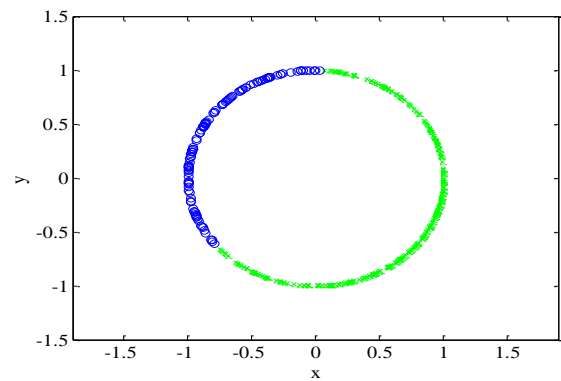
(e) Clustering results of EM with different initials

(f) Cluster number obtained by the learning-based EM under different iterations

(g) Final states of cluster centers of the learning-based EM

(h) Clustering results of learning-based EM

Fig. 3 Clustering results of learning-based EM and EM algorithms for SIDS data set (continued)

4. Conclusions

In general, the EM algorithm for clustering circular data quite depends on initializations with a given number of clusters a priori. To solve these drawbacks of EM, a learning-based EM algorithm is proposed for circular data in this paper. The proposed learning-based EM can automatically find an optimal number of clusters for different circular data sets without any initialization and also robust to outliers. Some numerical data from mixtures of von Mises distributions and real circular data are used for comparisons of the proposed learning-based EM with EM algorithms. Furthermore, we also consider noisy points and outliers in data sets. These comparisons and experimental results actually demonstrate the effectiveness and usefulness of the proposed learning-based EM clustering algorithm for circular data. Although the proposed learning-based EM works well for circular data, it is only fitted for 2-dimensional angular data sets. In our future work, we will further construct a learning-based EM algorithm for spherical data sets, i.e. 3-dimensional angular data, and then apply it to extrasolar planet taxonomy.

Conflicts of Interest

The authors declare no conflict of interest

References

[1] R. Von Mises, "Uber die 'Ganzzahligkeit' der atomgewicht und verwandte fragen," Physikal Z., vol. 19, pp. 490-500, 1918.
[2] G. S. Watson and E. J. Williams, "On the construction of significance tests on the circle and the sphere," Biometrika, vol. 43, no. 3/4, pp. 344-352, December 1956.
[3] N. I. Fisher, Statistical analysis of circular data, Cambridge:Cambridge University, in press, October 1995.

[4] N. Masseran, A. M. Razali, K. Ibrahim, and M. T. Latif, "Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia," Energy Conversion and Management, vol. 72, pp. 94-102, April 2013.

[5] L. P. Rivest and S. Kato, "A random-effects model for clustered circular data," Canadian Journal of Statistics, vol. 47, no. 4, pp. 712-728, August 2019.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, Series B, vol. 39, no. 1, pp. 1-22, September 1977.

[7] G. J. McLachlan and K. E. Basford, Mixture models: Inference and Applications to clustering, New York: Marcel Dekker, 1988.

[8] J. Yu, C. Chaomurilige, and M. S. Yang, "On convergence and parameter selection of the EM and DA-EM algorithms for Gaussian mixtures," Pattern Recognition, vol. 77, pp. 188-203, December 2017.

[9] J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California, in press, vol. 1, no. 14, pp. 281-297, June 1967.

[10] D. Pollard, "Quantization and the method of k-means," IEEE Transaction on Information Theory, vol. 28, no. 2, pp. 199-205, March 1982.

[11] M. S. Yang and K. P. Sinaga, "A feature-reduction multi-view k-means clustering algorithm," IEEE Access, vol. 7, pp. 114472-114486, August 2019.

[12] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms, New York: Plenum, in press, July 1981.

[13] M. S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm with feature-weighted entropy," IEEE Transactions on Fuzzy Systems, vol. 26, no. 2, pp 817-835, April 2018.

[14] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," IEEE Transaction on Fuzzy Systems, vol. 1, no. 2, pp. 98-110, May 1993.

[15] M. S. Yang, S. J. Chang-Chien, and Y. Nataliani, "A fully-unsupervised possibilistic c-means clustering method," IEEE Access, vol. 6, pp. 78308-78320, December 2018.

[16] R. Bartels, "Estimation in a bidirectional mixture of von Mises distributions," Biometrics, vol. 40, no. 3, pp. 777-784, September 1984.

[17] J. A. Mooney, P. J. Helms, and I. T. Jolliffe, "Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome," Computational Statistics and Data Analysis, vol. 41, no. 3/4, pp. 505-513, October 2002.

[18] N. Sanusi, A. Zaharim, S. Mat, and K. Sopian, "A Weibull and finite mixture of the von Mises distribution for wind analysis in Mersing, Malaysia," International Journal of Green Energy, vol. 14, no. 12, pp. 1057-1062, September 2017.

[19] Y. Ban, X. Alameda-Pineda, C. Evers, and R. Horaud, "Tracking multiple audio sources with the von Mises distribution and variational EM," IEEE Signal Processing Letters, vol. 26, no. 6, pp. 798-802, March 2019.

[20] M. S. Yang, C. Y. Lai, and C. Y. Lin, "A robust EM clustering algorithm for Gaussian mixture models," Pattern Recognition, vol. 45, no. 11, pp. 3950-3961, May 2012.

[21] M. S. Yang, S. J. Chang-Chien, and W. L. Hung, "Learning-based EM clustering for data on the unit hypersphere with application to exoplanet data," Applied Soft Computing, vol. 60, pp. 101-114, June 2017.