

# A Robust Fuzzy c-Means Clustering Algorithm for Incomplete Data

Jinhua Li<sup>1</sup>, Shiji Song<sup>1(✉)</sup>, Yuli Zhang<sup>1,2(✉)</sup>, and Kang Li<sup>3</sup>

<sup>1</sup> Department of Automation, TNList, Tsinghua University,  
Beijing 100084, People's Republic of China

lijinhua11@mails.tsinghua.edu.cn, {shijis,zhangyuli}@mail.tsinghua.edu.cn

<sup>2</sup> Department of Industrial Engineering, Tsinghua University,  
Beijing 100084, People's Republic of China

<sup>3</sup> School of Electronics, Electrical Engineering and Computer Science,  
Queens University Belfast, Belfast, UK  
k.li@qub.ac.uk

**Abstract.** Data sets with missing feature values are prevalent in clustering analysis. Most existing clustering methods for incomplete data rely on imputations of missing feature values. However, accurate imputations are usually hard to obtain especially for small-size or highly corrupted data sets. To address this issue, this paper proposes a robust fuzzy c-means (RFCM) clustering algorithm, which does not require imputations. The proposed RFCM represents the missing feature values by intervals, which can be easily constructed using the K-nearest neighbors method, and adopts a min-max optimization model to reduce the impact of noises on clustering performance. We give an equivalent tractable reformulation of the min-max optimization problem and propose an efficient solution method based on smoothing and gradient projection techniques. Experiments on UCI data sets validate the effectiveness of the proposed RFCM algorithm by comparison with existing clustering methods for incomplete data.

**Keywords:** Robust FCM · Interval data · Robust clustering algorithm

## 1 Introduction

Clustering analysis is a common technique in machine learning and data mining and has wide applications. Traditional clustering methods only tackle “complete” data sets, where no missing feature values exist. However, missing data are common occurrences in practice due to various reasons, such as missing replies in

---

S. Song—This work was supported by the Major Program of the National Natural Science Foundation of China under Grant 41427806, the National Natural Science Foundation of China under Grants 61503211 and 9152002, and the Project of China Ocean Association under Grant DYXM-125-25-02.

© Springer Nature Singapore Pte Ltd. 2017

D. Yue et al. (Eds.): LSMS/ICSEE 2017, Part II, CCIS 762, pp. 3–12, 2017.

DOI: 10.1007/978-981-10-6373-2\_1

questionnaire, high cost to acquire some feature values and improper collection procedure of data.

To address incomplete data clustering problems, a direct method is to use the two-step approach, which first estimates the missing feature values by imputation [16] and then applies the traditional clustering methods for complete data. Besides the imputation-based approaches, four strategies are proposed by [3] to tailor the FCM algorithm to handle incomplete data, including the whole data strategy (WDS), the partial distance strategy (PDS), the optimal completion strategy (OCS) and the nearest prototype strategy (NPS). In addition to these strategies, principal component analysis (PCA) [9] and local PCA [4] have also been used to capture incomplete data structure.

The limitations of these direct and iterative imputation methods are two-fold. First, these methods require an accurate estimation of the missing feature values, which is difficult to obtain in practice. To address this issue, interval data structures have been developed to represent the missing feature values in [6, 7, 12, 17]. Specifically, [6] define new distance function for interval data and extend the classical FCM to handle missing data. [7] also express missing values by interval data, but use the genetic algorithm to search for proper imputations of missing feature values. [12] estimate interval data by an enhanced back-propagation neural network. [17] design an improved interval construction approach using pre-classified cluster results and search for the optimal cluster by particle swarm optimization. Another limitation of existing clustering methods for missing data is that the cluster results are usually sensitive to the estimation of the missing feature values specifically for small-size and highly corrupted data sets [8].

The aim of this paper is to present a robust FCM clustering algorithm for incomplete data based on interval data representation. To guarantee the clustering performance, we introduce the concept of robust cluster objective function, which is the maximum of the traditional cluster objective function when the missing feature values vary in the considered intervals. Different from the existing algorithms based on interval distance function or optimal imputation [6, 7, 17], we formulate the clustering problem as a min-max optimization problem based on the idea of robust optimization, which has been successfully used in the field of operations research [18, 19], and machine learning, including the minimax probability machine [5, 10, 13], robust support vector machines [11, 15] and robust quadratic regression [14].

To solve the proposed min-max optimization problem, we design an efficient iterative solution method. We first give an equivalent reformulation of our robust optimization problem and analyze its relationship with the classical FCM. Then, we show how to update the cluster prototype and membership matrices separately by solving convex optimization problems. Since both problems are non-smooth, we propose a smoothing method to speed up solution process. To tackle the constrained optimization problem in the process of updating the membership matrix, we present an improved gradient projection method. Numerical experiments on UCI data sets validate the effectiveness of the proposed RFCM algorithm by comparison with other algorithms.

The remainder of this paper is organized as follows. Section 2 presents our RFCM model. In Sects. 3, we give an equivalent reformulation and give the solution method. In Sect. 4, numerical experiments are implemented and discussed. Finally, we conclude this research in Sect. 5.

## 2 RFCM Clustering Algorithm

Consider a set of  $n$  objects  $I = \{1, \dots, n\}$  and each object has  $m$  features  $\{x_{ij} : j \in J\}$ , where  $x_{ij}$  describes the  $j$ -th feature of the  $i$ -th object quantitatively. Let  $x_i = (x_{i1}, \dots, x_{im})^T$  be the feature vector of the  $i$ -th object and  $X = (x_1, \dots, x_n)$  be the feature matrix or data set. The task of clustering problems is to assign these objects to  $K$  clusters.

In practice, due to various reasons, the data set  $X$  may contain missing components. A data set  $X$  is referred to as an incomplete data set if it contains at least one missing feature value for some objects, that is, there exists at least one  $i \in I$  and  $j \in J$ , such that  $x_{ij} = ?$ . For an incomplete data set  $X$ , we further partition the feature set of the  $i$ -th object into two subsets:  $J_i^0 = \{j : x_{ij} = ?, \forall j \in J\}$  and  $J_i^1 = J \setminus J_i^0$ .

Since it is usually difficult to obtain accurate estimation of missing feature values. This paper represents missing feature values by intervals. Specifically, for any  $i \in I$ , we use an interval  $[x_{ij}^-, x_{ij}^+]$  to represent unknown missing feature value where  $j \in J_i^0$ , and use  $\bar{x}_{ij}$  to represent known feature value where  $j \in J_i^1$ . To simply notations, in the following, let  $\bar{x}_{ij} = \frac{x_{ij}^- + x_{ij}^+}{2}$ ,  $\delta_{ij} = \frac{x_{ij}^+ - x_{ij}^-}{2}$  for any  $j \in J_i^0$ , and  $\delta_{ij} = 0$  for any  $j \in J_i^1$ . For details on how to construct these intervals for missing feature values, see [6, 17].

To reduce the impact of inaccurate estimation of the missing feature values, this paper considers the following robust cluster objective function:

$$\begin{aligned} \min \max \left\{ \sum_{k=1}^K \sum_{i \in I} u_{ik}^p \|\bar{x}_i + y_i - v_k\|^2 : y_i \in [-\delta_i, \delta_i], \forall i \in I \right\} \\ \text{s.t. } \sum_{k=1}^K u_{ik} = 1, u_{ik} \in [0, 1], \forall i \in I, k = 1, \dots, K, \end{aligned} \quad (1)$$

The Eq. (1) is a difficult nonlinear optimization problem, and it is hard to solve (1) directly. However, by exploiting the structural properties of (1), we will design an efficient algorithm in next section.

## 3 Solution Method

### 3.1 Equivalent Reformulation of (1)

The following Proposition 1 simplifies the two-level min-max optimization problem (1) into a single level minimization problem, and provides the basis of designing effective solution methods for (1).

**Proposition 1.** *Problem (1) is equivalent to the following problem:*

$$\begin{aligned} \min_{U, V} \quad & \sum_{i \in I} \sum_{j \in J} \left\{ \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj})^2 + 2\delta_{ij} \left| \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right| + \sum_{k=1}^K u_{ik}^p \delta_{ij}^2 \right\} \\ \text{s.t.} \quad & \sum_{k=1}^K u_{ik} = 1, \quad u_{ik} \in [0, 1], \quad \forall i \in I, k = 1, \dots, K. \end{aligned} \quad (2)$$

*Proof.* We simplify the robust cluster objective function for given  $U$  and  $V$  as follows:

$$\begin{aligned} J^R(U, V) &= \sum_{i \in I} \sum_{j \in J} \max_{y_{ij} \in [-\delta_{ij}, \delta_{ij}]} \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} + y_{ij} - v_{kj})^2 \\ &= \sum_{i \in I} \sum_{j \in J} \max \left\{ \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} + \delta_{ij} - v_{kj})^2, \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - \delta_{ij} - v_{kj})^2 \right\} \end{aligned}$$

where the last equation uses the fact that  $f(y_{ij}) = (\bar{x}_{ij} + y_{ij} - v_{kj})^2$  is a convex function and attains its maximum over  $[-\delta_{ij}, \delta_{ij}]$  at the endpoints. For given  $i \in I$  and  $j \in J$ , we have

$$\begin{aligned} & \max \left\{ \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} + \delta_{ij} - v_{kj})^2, \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - \delta_{ij} - v_{kj})^2 \right\} \\ &= \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj})^2 + 2\delta_{ij} \left| \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right| + \sum_{k=1}^K u_{ik}^p \delta_{ij}^2, \end{aligned}$$

which completes the proof.

Proposition 1 also shows that when  $\delta_{ij} = 0$  for any  $i \in I$  and  $j \in J$ , our RFCM reduces to the classical FCM.

Let  $h(U, V)$  be the objective function of (2). The following proposition analyzes properties of  $h(U, V)$ .

**Proposition 2.** *For any given  $U \geq 0$ ,  $h(U, V)$  is convex in  $V$  and for any given  $V$ ,  $h(U, V)$  is convex in  $U$ .*

*Proof.* First, for any  $U \geq 0$ , it is easy to see that the first term of  $h(U, V)$  is convex in  $V$ . Note that the absolute function is convex; thus, the second term of  $h(U, V)$  is also convex in  $V$ . Since the last term of  $h(U, V)$  is constant for given value of  $U$ , we have that  $h(U, V)$  is convex in  $V$ .

Second, from the proof of Proposition 1, we have that

$$h(U, V) = \sum_{i \in I} \sum_{j \in J} \max \left\{ \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} + \delta_{ij} - v_{kj})^2, \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - \delta_{ij} - v_{kj})^2 \right\}.$$

Since  $f(x) = ax^p$  is convex in  $x$  when  $a \geq 0$  and  $p \geq 1$ , we have both  $\sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} + \delta_{ij} - v_{kj})^2$  and  $\sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - \delta_{ij} - v_{kj})^2$  are convex in  $U$  for any given  $V$ . Note that the maximum of two convex function is still convex; thus, we complete the proof.

Proposition 2 shows that although (2) is a complex non-convex minimization problem, when either the value of  $U$  or that of  $V$  is fixed, we only need to solve convex minimization subproblems. The next two subsection focuses on designing efficient solution methods to these two subproblems.

### 3.2 Optimizing $V$ for a Fixed Value of $U$

For a fixed value of  $U$ , we need to solve the following convex piecewise quadratic problem:

$$\min_V \sum_{j \in J} \sum_{i \in I} \left\{ \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj})^2 + 2\delta_{ij} \left| \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right| \right\}. \quad (3)$$

Note that (3) is decomposable in index  $j$ . Let  $v_j = \{v_{1j}, \dots, v_{Kj}\}$ . Therefore, it is sufficient to solve the following subproblem separately for each  $j \in J$ :

$$\min_{v_j} \sum_{i \in I} \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj})^2 + 2 \sum_{i \in I} \delta_{ij} \left| \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right|. \quad (4)$$

Due to the non-differentiability of the absolute function  $g(x) = |x|$ , (4) is a non-smooth convex minimization problem. Although the sub-gradient method can be used to solve such non-smooth optimization problems, its search direction may oscillate around non-smooth points. To improve computation efficiency, we introduce the following smoothing upper bound estimation (SUBE) function of  $g$ :

$$g_\epsilon(x) = \begin{cases} x, & \text{if } x \geq \epsilon, \\ \frac{x^2}{2\epsilon} + \frac{\epsilon}{2}, & \text{if } |x| < \epsilon, \\ -x, & \text{if } x \leq -\epsilon, \end{cases}$$

where  $\epsilon > 0$ . It is easy to see that for any  $\epsilon > 0$ ,  $g_\epsilon(x)$  is a smooth function and  $0 \leq g_\epsilon(x) - g(x) \leq \epsilon/2$ , and

$$g'_\epsilon(x) = \begin{cases} 1, & \text{if } x \geq \epsilon, \\ \frac{x}{\epsilon} + \frac{\epsilon}{2}, & \text{if } |x| < \epsilon, \\ -1, & \text{if } x \leq -\epsilon. \end{cases}$$

By introducing the SUBE function  $g_\epsilon(x)$ , instead of directly solving the non-smooth optimization problem (4), we only need to consider the following smooth optimization problem:

$$\min_{v_j} \phi(v_j) = \sum_{i \in I} \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj})^2 + 2 \sum_{i \in I} \delta_{ij} g_\epsilon \left( \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right) \quad (5)$$

The gradient of the objective function  $\phi$  is given as follows:

$$\frac{\partial \phi(v_j)}{\partial v_{kj}} = 2 \sum_{i \in I} u_{ik}^p (v_{kj} - \bar{x}_{ij}) - 2 \sum_{i \in I} \delta_{ij} g'_\epsilon \left( \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right).$$

Therefore, for a fixed value of  $U$ , we design the following smoothing gradient descent method to solve (5).

---

**Algorithm 1.** Smoothing Gradient Descent Algorithm (SGDA)

---

**0** Input  $U$ ,  $j \in J$ , the maximum iteration number  $N_{max}$  and error bound  $Err > 0$ .

**1** Initialization:  $v_{kj}^0 = \frac{\sum_{i \in I} u_{ik}^p \bar{x}_{ij}}{\sum_{i \in I} \sum_{i \in I} u_{ik}^p}$  for  $k = 1, \dots, K$ . Let  $n = 1$ .

**2** Update: set  $\epsilon = \frac{Err}{n}$ ,  $v_{kj}^n = v_{kj}^{n-1} - s_n \frac{\partial \phi(v_j^{n-1})}{\partial v_{kj}}$  for  $k = 1, \dots, K$  and  $n = n + 1$ .

**3** Stop criteria: if  $n \geq N_{max}$  or  $|\phi(v_j^n) - \phi(v_j^{n-1})| \leq Err$ , goto step 4; else goto step 2.

**4** Output solution  $v_j^* = v_j^n$ .

---

During the solution process, SGDA decreases the approximation error  $\epsilon$  repeatedly. In the SGDA, the step size can be either set as  $s_n = a/(b + n)$  with  $a, b > 0$  or selected by the Armijo rule [1].

### 3.3 Optimizing $U$ for a Fixed Value of $V$

For a fixed value of  $V$ , we need to solve the following constrained convex minimization problem:

$$\begin{aligned} \min_U \quad & \sum_{i \in I} \sum_{j \in J} \left\{ \sum_{k=1}^K u_{ik}^p ((\bar{x}_{ij} - v_{kj})^2 + \delta_{ij}^2) + 2\delta_{ij} \left| \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right| \right\} \\ \text{s.t.} \quad & \sum_{k=1}^K u_{ik} = 1, \quad u_{ik} \in [0, 1], \quad \forall i \in I, k = 1, \dots, K. \end{aligned} \quad (6)$$

Note that (6) is decomposable in index  $i$ . Let  $u_i = \{u_{i1}, \dots, u_{iK}\}$ . Using the SUBE function as in the last subsection, we consider the following smooth subproblem separately for each  $i \in I$ .

$$\begin{aligned} \min_{u_i} \psi(u_i) = \sum_{j \in J} \left\{ \sum_{k=1}^K u_{ik}^p ((\bar{x}_{ij} - v_{kj})^2 + \delta_{ij}^2) + 2\delta_{ij} g'_\epsilon \left( \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right) \right\} \\ \text{s.t.} \quad \sum_{k=1}^K u_{ik} = 1, \quad u_{ik} \in [0, 1], \quad \forall k = 1, \dots, K. \end{aligned} \quad (7)$$

The gradient of the objective function  $\psi$  is given as follows:

$$\frac{\partial \psi(u_i)}{\partial u_{ki}} = 2p \sum_{j \in J} u_{ik}^{p-1} ((\bar{x}_{ij} - v_{kj})^2 + \delta_{ij}^2) + 2 \sum_{j \in J} \delta_{ij} p u_{ik}^{p-1} (\bar{x}_{ij} - v_{kj}) g'_\epsilon \left( \sum_{k=1}^K u_{ik}^p (\bar{x}_{ij} - v_{kj}) \right).$$

To solve (7), we design the following smoothing gradient projection descent algorithm.

---

**Algorithm 2.** Smoothing Gradient Projection Descent Algorithm (SGPDA)
 

---

**0** Input  $V$ ,  $i \in I$ , the maximum iteration number  $N_{max}$  and error bound  $Err > 0$ .

**1** Initialization:  $u_{ik}^0 = \left( \sum_{t=1}^K \left( \frac{\|\bar{x}_i - v_k\|^2}{\|\bar{x}_i - v_t\|^2} \right)^{\frac{1}{m-1}} \right)^{-1}$  for  $k = 1, \dots, K$ . Let  $n = 1$ .

**2** Update: set  $\epsilon = \frac{Err}{n}$ ,  $\bar{u}_{ik}^n = u_{ik}^{n-1} - s_n \frac{\partial \psi(u_i^{n-1})}{\partial u_{ik}}$  for  $k = 1, \dots, K$  and  $u_i^n = \text{Proj}_{\Delta}(\bar{u}_i^n)$ . Let  $n = n + 1$ .

**3** Stop criteria: if  $n \geq N_{max}$  or  $|\psi(u_i^n) - \psi(u_i^{n-1})| \leq Err$ , goto step 4; else goto step 2.

**4** Output solution  $u_i^* = u_i^n$ .

---

In SGPDA, the projection operator  $\text{Proj}_{\Delta}$  denotes the projection onto the simplex  $\Delta = \{x \in R^K : \sum_{k=1}^K x_k = 1, x_k \geq 0, \forall k = 1, \dots, K\}$ . [2] has shown that  $\text{Proj}_{\Delta}$  can be computed in  $\mathcal{O}(K \log K)$  time.

Based on the solution methods for subproblems  $(\tilde{P}_U^j)$  and  $(\tilde{P}_V^i)$ , our RFCM algorithm can be summarized as follows.

---

**Algorithm 3.** RFCM Algorithm
 

---

**0** Input data: estimated data  $\bar{X}$ , interval size data  $\delta$ , the number of clustering  $K$ , the maximum iteration numbers  $N_{fcm}$  and  $N_{sub}$ , and the error bounds  $Err_{fcm}$  and  $Err_{sub}$ .

**1** Initialization: Initialize the clustering centers  $V^0$  by randomly choose  $K$  samples, and set  $k = 1$ .

**2** Update  $U$  when  $V$  is fixed as  $V = V^{k-1}$ :

For  $i \in I$

Solve  $(\tilde{P}_V^i)$  by SGPDA with the input:  $V = V^{k-1}$ ,  $i$ ,  $N_{sub}$  and  $Err_{sub}$ .

End

Set  $U^k = \{u_i^* : i \in I\}$ .

**3** Update  $V$  when  $U$  is fixed as  $U = U^k$ :

For  $j \in J$

Solve  $(\tilde{P}_U^j)$  by GPDA with the input:  $U = U^k$ ,  $j$ ,  $N_{sub}$  and  $Err_{sub}$ .

End

Set  $V^k = \{v_j^* : j \in J\}$ .

**4** Calculate objective function value  $h^k = h(U^k, V^k)$ .

**5** Stop criteria: if  $k \geq N_{fcm}$  or  $|h^k - h^{k-1}| \leq Err_{fcm}$ , then return  $(U^*, V^*) = (U^k, V^k)$ ; otherwise, let  $k = k + 1$  and goto step 2.

**6** Output  $U^*$  and  $V^*$ .

---

## 4 Numerical Experiments

### 4.1 Data Sets

We conduct experiments on two data sets of the UCI database: Wine and Seeds. All these data sets are complete data sets. To generate the incomplete data sets, we adopt the method used in [3,6]. Specifically, we use the missing completely at random (MCAR) mechanism to generate the missing feature values. We randomly select a specified percentage of components and designate them as missing. We further make sure the following constraints are satisfied:

1. Each object retains at least one feature;
2. Each feature has at least one value present in the incomplete data set.

### 4.2 Experimental Results and Discussion

We compare the proposed RFCM algorithm with the classical FCM using the WDS, PDS and NPS strategies on Wine and Seeds data sets. The missing rates of these data sets vary from 0% to 20%. We report the number of misclassification and misclassification rate of these algorithms on test data sets.

We generate 100 incomplete data instances of Wine and Seeds data sets and report the averaged performance of different algorithms in Tables 1 and 2, respectively. Results given by the proposed RFCM, the classical FCM using the WDS, PDS and NPS strategies are labeled as “RFCM”, “WDS”, “PDS”, and “NPS”, respectively.

**Table 1.** Performance of different FCM algorithms on the incomplete Wine data set

%	Number of misclassification				Misclassification rate			
	WDS	PDS	NPS	RFCM	WDS	PDS	NPS	RFCM
0	16.2	16.2	16.2	16.2	9.1	9.1	9.1	9.1
5	18.0	21.0	18.8	17.4	10.13	11.82	10.56	9.77
10	19.7	24.1	21.7	18.5	11.05	13.54	12.18	10.37
15	22.4	27.0	23.7	19.6	12.61	15.17	13.29	11.02
20	25.4	32.6	25.1	22.4	14.28	18.33	14.12	12.58

From Tables 1 and 2, we have the following observations.

- (1) For all the test data sets, the proposed RFCM algorithm provides the best clustering performance in terms of misclassification rate. For example, when the missing rate of the Wine data set is 20%, the misclassification rate of RFCM is only 12.58% while the misclassification rate of WDS, PDS and NPS are 14.28%, 18.33% and 14.12%, respectively.



**Table 2.** Performance of different FCM algorithms on the incomplete Seeds data set

%	Number of misclassification				Misclassification rate			
	WDS	PDS	NPS	RFCM	WDS	PDS	NPS	RFCM
0	22.3	22.3	22.3	22.3	10.62	10.62	10.62	10.62
5	22.2	25.4	23.5	19.3	10.58	12.13	11.18	9.17
10	23.6	45.8	31.4	23.1	11.24	21.81	14.95	11.01
15	25.3	42.8	32.2	24.2	12.04	20.40	15.34	11.53
20	33.7	61.4	30.3	27.1	16.07	29.23	14.44	12.94

- (2) When the missing rate is small, missing data have little adverse effect on the performance of the proposed RFCM and missing data even help RFCM to give better performance.
- (3) The cluster prototypes given by RFCM also have smaller cluster prototype errors compared with cluster prototypes given by WDS, PDS and NPS.

## 5 Conclusion

This paper proposes a robust FCM algorithm to cluster data sets with missing feature values. The RFCM algorithm represents the missing feature values by intervals. Different existing interval-based clustering algorithms, which only replace the traditional Euclidean distance with a modified distance function for interval data, the RFCM algorithm adopts the idea of robust optimization and aims to find an optimal cluster with the minimum worst-case cluster objective function value. Therefore, it can guarantee the worst-case performance of the resulted cluster output and minimizes the adverse effect of missing feature values. This paper formulates the robust clustering problem as a two-level min-max optimization problem and provides an equivalent reformulation. An efficient solution method is further designed based on smoothing and gradient projection techniques. Experiments on UCI data sets also validate the effectiveness and robustness of the RFCM algorithm by comparison with existing clustering algorithms.

## References

1. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Belmont (1999)
2. Condat, L.: Fast projection onto the simplex and the l-1 ball. Preprint HAL, 1056171 (2014)
3. Hathaway, R.J., Bezdek, J.C.: Fuzzy c-means clustering of incomplete data. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **31**(5), 735–744 (2001)
4. Honda, K., Ichihashi, H.: Linear fuzzy clustering techniques with missing values and their application to local principal component analysis. *IEEE Trans. Fuzzy Syst.* **12**(2), 183–193 (2004)

5. Lanckriet, G.R.G., Ghaoui, L.E., Bhattacharyya, C., Jordan, M.I.: Minimax probability machine. *Adv. Neural Inf. Process. Syst.* **1**, 801–808 (2002)
6. Li, D., Hong, G., Zhang, L.: A fuzzy c-means clustering algorithm based on nearest-neighbor intervals for incomplete data. *Expert Syst. Appl.* **37**(10), 6942–6947 (2010)
7. Li, D., Hong, G., Zhang, L.: A hybrid genetic algorithm–fuzzy c-means approach for incomplete data clustering based on nearest-neighbor intervals. *Soft. Comput.* **17**(10), 1787–1796 (2013)
8. Li, J., Song, S., Zhang, Y., Zhou, Z.: Robust k-median and k-means clustering algorithms for incomplete data. *Math. Prob. Eng.* **2016**, 1–8 (2016)
9. Shibayama, T.: A PCA-like method for multivariate data with missing values. *Japan. J. Educ. Psychol.* **40**(2), 257–265 (1992)
10. Song, S., Gong, Y., Zhang, Y., Huang, G., Huang, G.-B.: Dimension reduction by minimum error minimax probability machine. *IEEE Trans. Syst. Man Cybern.: Syst.* **47**(1), 58–69 (2017)
11. Trafalis, T., Gilbert, R.: Robust support vector machines for classification and computational issues. *Optim. Methods Softw.* **22**(1), 187–198 (2007)
12. Wang, B.L., Zhang, L.Y., Zhang, L., Bing, Z.H., Xu, X.H.: Missing data imputation by nearest-neighbor trained bp for fuzzy clustering. *J. Inf. Comput. Sci.* **11**(15), 5367–5375 (2014)
13. Wang, Y., Zhang, Y., Yi, J., Qu, H., Miu, J.: A robust probability classifier based on the modified-distance. *Math. Probl. Eng.* **2014**, 1–13 (2014)
14. Wang, Y., Zhang, Y., Zhang, F., Yi, J.: Robust quadratic regression and its application to energy-growth consumption problem. *Math. Probl. Eng.* **2013**, 1–10 (2013)
15. Huan, X., Caramanis, C., Mannor, S.: Robustness and regularization of support vector machines. *J. Mach. Learn. Res.* **10**, 1485–1510 (2009)
16. Yao, L., Weng, K.-S.: Imputation of incomplete data using adaptive ellipsoids with linear regression. *J. Intell. Fuzzy Syst.* **29**(1), 253–265 (2015)
17. Zhang, L., Bing, Z., Zhang, L.: A hybrid clustering algorithm based on missing attribute interval estimation for incomplete data. *Pattern Anal. Appl.* **18**(2), 377–384 (2015)
18. Zhang, Y., Shen, Z.-J.M., Song, S.: Distributionally robust optimization of two-stage lot-sizing problems. *Prod. Oper. Manag.* **25**(12), 2116–2131 (2016)
19. Zhang, Y., Song, S., Shen, Z.-J.M., Wu, C.: Data-driven robust shortest path problem with distributional uncertainty. *IEEE Trans. Intell. Transp. Syst.* (2017). doi:[10.1109/TITS.2017.2709798](https://doi.org/10.1109/TITS.2017.2709798)