The Effect of Noise and Outliers on Fuzzy Clustering of High Dimensional Data

Ludmila Himmelspach and Stefan Conrad
Institute of Computer Science, Heinrich-Heine-Universität Düsseldorf, 40225 - Düsseldorf, Germany

Keywords: Fuzzy Clustering, c-Means Models, High Dimensional Data, Noise, Possibilistic Clustering.

Abstract: Clustering high dimensional data is still a challenging problem for fuzzy clustering algorithms because distances between each pair of data items get similar with the increasing number of dimensions. The presence of noise and outliers in data is an additional problem for clustering algorithms because they might affect the computation of cluster centers. In this work, we analyze the effect of different kinds of noise and outliers on fuzzy clustering algorithms that can handle high dimensional data: FCM with attribute weighting, the multivariate fuzzy c-means (MFCM), and the possibilistic multivariate fuzzy c-means (PMFCM). Additionally, we propose a new version of PMFCM to enhance its ability handling noise and outliers in high dimensional data. The experimental results on different high dimensional data sets show that the possibilistic versions of MFCM produce accurate cluster centers independently of the kind of noise and outliers.

1 INTRODUCTION

Clustering algorithms are used in many fields like bioinformatics, image processing, text mining, and many others. Data sets in these applications usually contain many features. Therefore, there is a need for clustering algorithms that can handle high dimensional data. The hard *k-means* algorithm (MacQueen, 1967) is still mostly used for clustering high dimensional data although it is comparatively unstable and sensitive to the initialization. It is not able to distinguish data items belonging to clusters from noise and outliers. This is another issue of the hard k-means algorithm because noise and outliers might influence the computation of cluster centers leading to inaccurate clustering results.

In the case of low dimensional data, the *fuzzy c-means* algorithm (FCM) (Bezdek, 1981), (Dunn, 1973) which assigns data items to clusters with membership degrees might be a better choice because it is more stable and less sensitive to initialization (Klawonn et al., 2015). The *possibilistic fuzzy c-means* algorithm (PFCM) (Pal et al., 2005) partitions data items in presence of noise and outliers. However, when FCM is applied on high dimensional data, it tends to produce cluster centers close to the center of gravity of the data set (Winkler et al., 2011), (Klawonn, 2013). In this work, we analyze three fuzzy clustering methods that are suitable for clustering high dimensional data. The first approach is the *attribute weighting fuzzy clustering algorithm* (Keller and Klawonn, 2000) that uses a new attribute weighting function to determine attributes that are important for each single cluster. This method was recommended in (Klawonn, 2013) for fuzzy clustering high dimensional data. The second approach is the *multivariate fuzzy c-means* (MFCM) (Pimentel and de Souza, 2013) that computes membership degrees of data items to each cluster in each feature. The third method is the *possibilistic multivariate fuzzy c-means* (PMFCM) (Himmelspach and Conrad, 2016) which is an extension of MFCM in a possibilistic clustering scheme. Additionally, we propose a new version of PMFCM to enhance its ability distinguishing data items belonging to clusters from noise and outliers in high dimensional data. The main objective of this work is to analyze the effect of noise and outliers on fuzzy clustering algorithms for high dimensional data. Our aim is to ascertain which fuzzy clustering algorithms are resistant to which kind of noise and outliers when clustering high dimensional data.

The rest of the paper is organized as follows: In the next section we give a short overview of fuzzy clustering methods for high-dimensional data. The evaluation results on artificial data sets containing different kinds of noise and outliers are presented in Section 3. Section 4 closes the paper with a short summary and the discussion of future research.

Himmelspach, L. and Conrad, S.
The Effect of Noise and Outliers on Fuzzy Clustering of High Dimensional Data.
DOI: 10.5220/0006070601010108
In *Proceedings of the 8th International Joint Conference on Computational Intelligence (IJCCI 2016) - Volume 2: FCTA*, pages 101-108
ISBN: 978-989-758-201-1

2 FUZZY CLUSTERING ALGORITHMS

Fuzzy c-means (FCM) (Bezdek, 1981), (Dunn, 1973) is a partitioning clustering algorithm that assigns data objects to clusters with membership degrees. The objective function of FCM is defined as:

$$J_m(U,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m d^2(v_i,x_k), \quad (1)$$

where c is the number of clusters, $u_{ik} \in [0,1]$ is the membership degree of data item x_k to cluster i, $m > 1$ is the fuzzification parameter, $d(v_i,x_k)$ is the distance between cluster prototype v_i and data item x_k. The objective function of FCM is usually minimized in an alternating optimization (AO) scheme (Bezdek, 1981) under constraint (2).

$$\begin{aligned} &\sum_{i=1}^{c} u_{ik} = 1 \ \ \forall k \in \{1,...,n\} \text{ and} \\ &\sum_{k=1}^{n} u_{ik} > 0 \ \ \forall i \in \{1,...,c\}. \end{aligned} \quad (2)$$

The algorithm begins with initialization of cluster prototypes with random values in the data space. In each iteration of the algorithm, the membership degrees and the cluster prototypes are alternating updated according to Formulae (3) and (4).

$$u_{ik} = \frac{\left(d^2(v_i,x_k)\right)^{\frac{1}{1-m}}}{\sum\limits_{l=1}^{c}\left(d^2(v_l,x_k)\right)^{\frac{1}{1-m}}}, \ 1 \le i \le c, \ 1 \le k \le n. \quad (3)$$

$$v_i = \frac{\sum\limits_{k=1}^{n} u_{ik}^m x_k}{\sum\limits_{k=1}^{n} u_{ik}^m}, \quad 1 \le i \le c \ . \quad (4)$$

The iterative process continues as long as the cluster prototypes change up to a chosen limit.

The FCM algorithm has several advantages over the hard k-means algorithm (MacQueen, 1967) in low dimensional data. It is more stable, less sensitive to initialization, and is able to model soft transitions between clusters (Klawonn et al., 2015). However, the hard k-means algorithm is still mostly used in real world applications for clustering high dimensional data because fuzzy c-means does not provide useful results on high dimensional data. It usually computes equal membership degrees of all data items to all clusters which results in the computation of final cluster prototypes close to the center of gravity of the entire data set. This is due to the *concentration of distance phenomenon* described in (Beyer et al., 1999). It says that the distance to the nearest data item approaches the distance to the farthest one with increasing number of dimensions.

2.1 Fuzzy Clustering Algorithms for High Dimensional Data

In (Klawonn, 2013), the author recommended to use the *attribute weighting fuzzy clustering algorithm* (Keller and Klawonn, 2000) for clustering high dimensional data. This method uses a distance function that weights single attributes for each cluster:

$$d^2(v_i,x_k) = \sum_{j=1}^{p} \alpha_{ij}^t (x_{kj} - v_{ij})^2, \ 1 \le i \le c, \ 1 \le k \le n, \quad (5)$$

where p is the number of attributes, $t > 1$ is a fixed parameter that determines the strength of the attribute weighting, and

$$\sum_{j=1}^{p} \alpha_{ij} = 1 \ \ \forall i \in \{1,...,c\}. \quad (6)$$

This approach works in the same way as FCM but it circumvents the concentration of distance phenomenon using a distance function that gives more weight to features that determine a particular cluster. The objective function of the attribute weighting fuzzy clustering algorithm is defined as:

$$J_{m,t}(U,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \sum_{j=1}^{p} \alpha_{ij}^t (v_{ij} - x_{kj})^2. \quad (7)$$

The attribute weights are updated in an additional iteration step according to Formula (8).

$$\alpha_{ij} = \left(\sum_{r=1}^{p}\left(\frac{\sum\limits_{k=1}^{n} u_{ik}^m (x_{kj} - v_{ij})^2}{\sum\limits_{k=1}^{n} u_{ik}^m (x_{kr} - v_{ir})^2}\right)^{\frac{1}{t-1}}\right)^{-1} \quad (8)$$

$$1 \le i \le c, \ 1 \le k \le n.$$

The second approach that we analyze in this work is the *multivariate fuzzy c-means* (MFCM) algorithm. This fuzzy clustering method computes membership degrees of data items to clusters for each feature (Pimentel and de Souza, 2013). The objective function of MFCM is defined as follows:

$$J_m(U,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p} u_{ikj}^m (v_{ij} - x_{kj})^2, \quad (9)$$

where $u_{ikj} \in [0,1]$ is the membership degree of data object x_k to cluster i for feature j. The objective function of MFCM has to be minimized under constraint (10).

$$\begin{aligned} &\sum_{i=1}^{c}\sum_{j=1}^{p} u_{ikj} = 1 \ \ \forall k \in \{1,...,n\} \text{ and} \\ &\sum_{j=1}^{p}\sum_{k=1}^{n} u_{ikj} > 0 \ \ \forall i \in \{1,...,c\}. \end{aligned} \quad (10)$$

The multivariate membership degrees and the cluster prototypes are updated in the iterative process of the algorithm according to (11) and (12).

$$u_{ikj} = \left(\sum_{l=1}^{c} \sum_{j=1}^{p} \left(\frac{(x_{kj} - v_{ij})^2}{(x_{kj} - v_{lj})^2} \right)^{\frac{1}{m-1}} \right)^{-1} \qquad (11)$$
$$1 \leq i \leq c,\ 1 \leq k \leq n,\ 1 \leq j \leq p.$$

$$v_{ij} = \frac{\sum_{k=1}^{n} u_{ikj}^m x_{kj}}{\sum_{k=1}^{n} u_{ikj}^m} \quad 1 \leq i \leq c,\ 1 \leq j \leq p. \qquad (12)$$

MFCM is not influenced by the concentration of distance phenomenon because it computes membership degrees in each feature depending on the partial distances in single dimensions. Therefore, this approach is suitable for clustering high dimensional data.

2.2 Possibilistic Clustering Algorithms for High Dimensional Data

The fuzzy clustering algorithms described before are not designed to cluster data in presence of noise and outliers. They assign such data items to clusters in the same way as data items within clusters. In this way, noise and outliers might affect the computation of cluster centers which leads to inaccurate partitioning results. There are different ways for avoiding this problem. The mostly used method is determining outliers before clustering. There are different methods for outlier detection but the most of them compare the distances from data points to their neighbors (Kriegel et al., 2009). Another method called *noise clustering* introduces an additional cluster that contains all data items that are located far away from any of cluster centers (Dave and Krishnapuram, 1997), (Rehm et al., 2007). In this work, we extend the clustering algorithms described in the previous subsection using the *possibilistic fuzzy c-means* (PFCM) clustering model proposed in (Pal et al., 2005). This approach extends the basic FCM by typicality values that express a degree of typicality of each data item to the overall clustering structure of data set. The advantage of using typicality values is that outliers get less weight in the computation of cluster centers. The objective function of PFCM is defined as:

$$J_{m,\eta}(U,T,V;X) = \sum_{k=1}^{n} \sum_{i=1}^{c} (a u_{ik}^m + b t_{ik}^{\eta}) d^2(v_i, x_k) + \sum_{i=1}^{c} \gamma_i \sum_{k=1}^{n} (1 - t_{ik})^{\eta}, \qquad (13)$$

where $t_{ik} \leq 1$ is the typicality value of data item x_k to cluster i, $m > 1$ and $\eta > 1$ are user defined constants. Similarly to FCM, the first term in (13) ensures that distances between data items and cluster centers are minimized, where constants $a > 0$ and $b > 0$ control the relative influence of fuzzy membership degrees and typicality values. The second term ensures that typicality values are determined as large as possible. The second summand is weighted by the parameter $\gamma_i > 0$. In (Krishnapuram and Keller, 1993), the authors recommended to run the basic FCM algorithm before PFCM and to choose γ_i by computing:

$$\gamma_i = K \frac{\sum_{k=1}^{n} u_{ik}^m d^2(v_i, x_k)}{\sum_{k=1}^{n} u_{ik}^m} \quad 1 \leq i \leq c, \qquad (14)$$

where the $\{u_{ik}\}$ are the terminal membership degrees computed by FCM and $K > 0$ (usually $K = 1$). The objective function of PFCM has to be minimized under constraints (2) and (15).

$$\sum_{k=1}^{n} t_{ik} > 0,\ \forall i \in \{1, ..., c\} \qquad (15)$$

In (Himmelspach and Conrad, 2016), we extended the MFCM algorithm in the possibilistic clustering scheme to make it less sensitive to outliers. Since we considered outliers as data points that have a large overall distance to any cluster, we did not compute typicality values of data points to clusters for each feature. The objective function of the resulting approach that we refer here as *possibilistic multivariate fuzzy c-means* (PMFCM) is defined as

$$J_{m,\eta}(U,T,V;X) = \sum_{k=1}^{n} \sum_{i=1}^{c} \sum_{j=1}^{p} (a u_{ikj}^m + b t_{ik}^{\eta})(v_{ij} - x_{kj})^2 + p \sum_{i=1}^{c} \gamma_i \sum_{k=1}^{n} (1 - t_{ik})^{\eta}. \qquad (16)$$

The objective function of PMFCM has to be minimized under constraint (17).

$$\sum_{i=1}^{c} u_{ikj} = 1 \ \ \forall k, j \ \text{ and } \sum_{k=1}^{n} u_{ikj} > 0 \ \ \forall i, j, \text{ and } \sum_{k=1}^{n} t_{ik} > 0 \ \ \forall i. \qquad (17)$$

In MFCM, the sum of membership degrees over all clusters and features to a particular data item was constrained to be 1. Since we want to retain equal influence of membership degrees and typicality values, we only constrain the sum of membership degrees over all clusters to a particular data item to be 1.

The membership degrees and the typicality values have to be updated according to (18) and (19). In PMFCM, the cluster centers are updated in a similar way as in PFCM according to Formula (20).

$$u_{ikj} = \left(\sum_{l=1}^{c} \left(\frac{(x_{kj} - v_{ij})^2}{(x_{kj} - v_{lj})^2} \right)^{\frac{1}{m-1}} \right)^{-1} \tag{18}$$
$$1 \le i \le c,\ 1 \le k \le n,\ 1 \le j \le p.$$

$$t_{ik} = \left(1 + \left(\frac{b \sum_{j=1}^{p} (x_{kj} - v_{ij})^2}{\gamma_i p} \right)^{\frac{1}{\eta-1}} \right)^{-1} \tag{19}$$
$$1 \le i \le c,\ 1 \le k \le n.$$

$$v_{ij} = \frac{\sum_{k=1}^{n} (a u_{ikj}^m + b t_{ik}^{\eta}) x_{kj}}{\sum_{k=1}^{n} (a u_{ikj}^m + b t_{ik}^{\eta})} \tag{20}$$
$$1 \le i \le c,\ 1 \le j \le p.$$

The membership degrees of data objects to clusters can be computed in this model as the average of the multivariate membership degrees over all variables, $u_{ik} = \frac{1}{p} \sum_{j=1}^{p} u_{ikj}$.

Due to the concentration of distance phenomenon, PMFCM might not produce meaningful typicality values because it uses the Euclidean distances between data items and cluster prototypes. Since distances between data items within clusters and cluster centers and distances between outliers and cluster centers might be similar in high dimensional data, the typicality values as they are computed in PMFCM will not be helpful for distinguishing between data items within clusters and outliers. Therefore, in this work, we propose another version of PMFCM that computes typicality values of data items to clusters in each dimension. We refer this approach here as *possibilistic multivariate fuzzy c-means for high dimensional data* (PMFCM_HDD). The objective function of PMFCM_HDD is given in Formula (21).

$$J_{m,\eta}(U,T,V;X) = \sum_{k=1}^{n} \sum_{i=1}^{c} \sum_{j=1}^{p} (a u_{ikj}^m + b t_{ikj}^{\eta})(v_{ij} - x_{kj})^2 + \sum_{i=1}^{c} \gamma_i \sum_{k=1}^{n} \sum_{j=1}^{p} (1 - t_{ikj})^{\eta}. \tag{21}$$

The objective function of PMFCM_HDD has to be minimized under constraint (22).

$$\sum_{i=1}^{c} u_{ikj} = 1 \ \ \forall k, j \ \text{ and } \ \sum_{k=1}^{n} u_{ikj} > 0 \ \ \forall i, j, \text{ and } \ \sum_{k=1}^{n} t_{ikj} > 0 \ \ \forall i, j. \tag{22}$$

The membership degrees are updated in the same way as in PMFCM according to Formula (18). The update Formulae for typicality values and cluster centers are given in (23) and (24).

$$t_{ikj} = \left(1 + \left(\frac{b (x_{kj} - v_{ij})^2}{\gamma_i} \right)^{\frac{1}{\eta-1}} \right)^{-1} \tag{23}$$
$$1 \le i \le c,\ 1 \le k \le n,\ 1 \le j \le p.$$

$$v_{ij} = \frac{\sum_{k=1}^{n} (a u_{ikj}^m + b t_{ikj}^{\eta}) x_{kj}}{\sum_{k=1}^{n} (a u_{ikj}^m + b t_{ikj}^{\eta})} \tag{24}$$
$$1 \le i \le c,\ 1 \le j \le p.$$

The typicality values of data objects to clusters can also be computed as average of the multivariate typicality values over all variables, $t_{ik} = \frac{1}{p} \sum_{j=1}^{p} t_{ikj}$.

3 DATA EXPERIMENTS

We tested the four fuzzy clustering methods for high dimensional data described in Sections 2 on artificial data sets containing different kinds of noise and outliers. The main data set was generated similarly to one that was used in (Keller and Klawonn, 2000). It is a two-dimensional data set that consists of 1245 data points unequally distributed on one spherical cluster and three clusters that have a low variance in one of the dimensions. The data set is depicted in Figure 1. We generated the second and the third data sets that are depicted in Figures 2 and 3 by adding 150 and 300 noise points to the main data set.

In order to generate high dimensional data sets, in the first experiment, we added 18 additional dimensions containing feature values close to zero. Additionally, we generated the fourth data set by adding 300 noise points containing values different from zero in all 20 dimensions. We had to modify PMFCM and PMFCM_HDD. Instead of running FCM at the beginning, we only computed membership values to compute γ_i. If we ran FCM at the beginning, it output cluster centers close to the center of gravity of the data set which is a bad initialization for possibilistic clustering algorithms. In order to provide the best starting conditions for the clustering algorithms, in this preliminary work, we initialized the cluster centers with the original means of clusters in all experiments.

Table 1 shows the experimental results for FCM with attribute weighting (FCM_AW), MFCM and its possibilistic versions PMFCM and PMFCM_HDD for $a = 0.5$ and different values of b on the 20-dimensional data set without noise and outliers. In order to evaluate the experimental results, we computed the Frobenius distance d_{orig} between the original cluster means and the cluster centers obtained by the clustering algorithms. Moreover, we computed the sum of distances d_{means} between cluster centers produced by the clustering algorithms. The most accurate cluster centers were produced by MFCM. This is indicated by a low d_{orig} and high d_{means} (good separation between clusters). The FCM algorithm with attribute weighting produced the least accurate cluster centers because it recognized the additional dimensions containing feature values close to zero as the important ones and gave small weights to important dimensions. This method was introduced by the authors as a clustering algorithm that performs the dimensionality reduction while clustering. Our results have shown that this method is not always able to distinguish between important and unimportant dimensions. The second best results were produced by the possibilistic clustering methods PMFCM and PMFCM_HDD for $b = 1000$. Tables 2 and 3 show the experimental results on the second and the third data sets with 150 and 300 noise points. The clustering algorithms performed similarly as on the data set without noise and outliers. Table 4 shows the performance results of the algorithms on the fourth data set where noise points contained values different from zero in all features. MFCM produced cluster centers that were farther from each other than the real means of clusters. FCM_AW produced cluster centers that were not close to the center of gravity of the entire data set anymore but they were too far from each other. The most accurate cluster centers were obtained by PMFCM and PMFCM_HDD for $b = 1000$. In all experiments, PMFCM_HDD produced more accurate results than PMFCM because the last uses the Euclidean distances for computation of typicality values which is adversely in high dimensional domain.

In the second experiment, we distributed clusters in the data space so that clusters were determined by different features. For example, data items of the first cluster had real values in the third and the fourth features, data items of the second cluster had real values in the sevenths and the eighth features etc. Furthermore, we completely distributed all noise points among the dimensions so that each dimension pair contained some noise points. Table 5 shows the experimental results on the 20-dimensional data sets with four clusters distributed among dimensions without noise and outliers. The clustering algorithms similarly performed as on the data set where data points of all clusters had real values in the first two features. Tables 6 and 7 show performance results of the clustering algorithms on the same data set as in

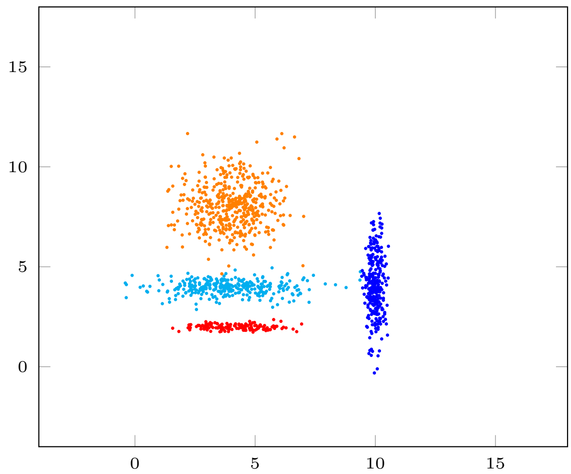

Figure 1: Test data with four clusters.

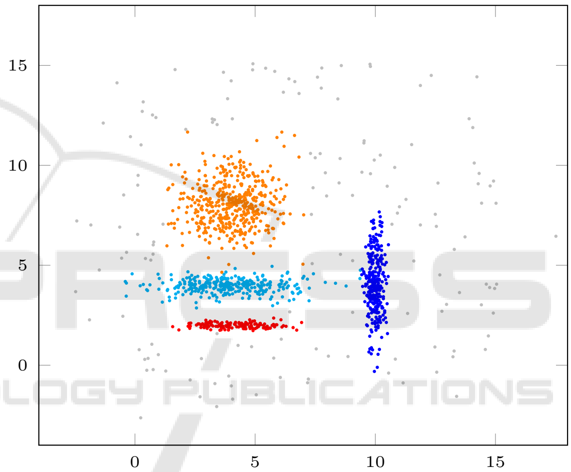

Figure 2: Test data with four clusters and 150 noise points.

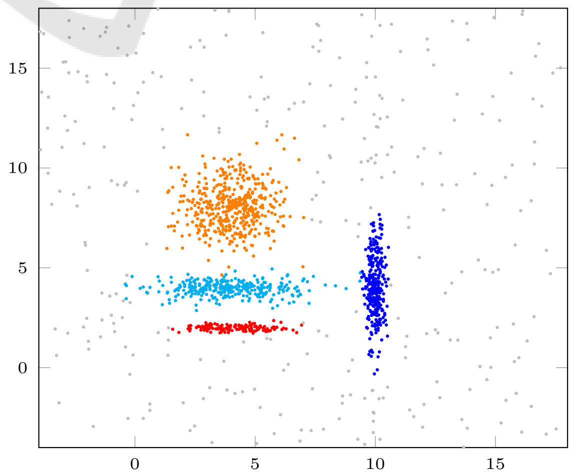

Figure 3: Test data with four clusters and 300 noise points.

the previous experiment but with 150 and 300 noise points. As in the previous experiments, MFCM pro-

Table 1: Comparison between clustering methods on 20-dimensional data set with four clusters.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	13.41	0.829	0.356	0.310
d_{means}	0.0004	32.58	32.05	31.82

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	0.019	0.762	1.035	0.748
d_{means}	31.58	31.45	32.28	32.10

Table 2: Comparison between clustering methods on 20-dimensional data set with four clusters and 150 noise points.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	13.67	0.853	0.655	0.376
d_{means}	0.0023	32.68	32.63	32.19

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	0.031	2.456	0.911	1.051
d_{means}	31.61	27.34	32.05	32.36

Table 3: Comparison between clustering methods on 20-dimensional data set with four clusters and 300 noise points.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	13.99	0.874	0.899	0.674
d_{means}	0.0004	32.67	32.82	32.67

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	0.019	2.418	0.720	0.912
d_{means}	31.58	26.73	31.34	32.10

Table 4: Comparison between clustering methods on 20-dimensional data set with four clusters and 300 noise points containing values in all features.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	95.24	2.501	2.273	0.709
d_{means}	90.86	26.46	27.52	32.15

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	12.81	13.33	6.052	2.797
d_{means}	38.16	1.489	21.15	25.86

duced the most accurate cluster prototypes. Its possibilistic versions PMFCM and PMFCM_HDD also obtained final cluster centers that were close to the actual cluster means. Although FCM with attribute weighting produced the least accurate cluster prototypes, its performance was comparable to the performance of the other methods. Due to noise points, all features contained at least some values different from each other. Therefore, FCM_AW was able to correctly recognize features determining different clusters. In the last experiment, we added 300 noise points that contained values different from zero in all features. As in

Table 5: Comparison between clustering methods on 20-dimensional data set with four clusters distributed in data space.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	27.47	1.333	0.495	0.460
d_{means}	0.0017	67.49	66.64	66.39

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	0.072	0.753	0.959	0.701
d_{means}	65.97	65.34	65.33	65.79

Table 6: Comparison between clustering methods on 20-dimensional data set with four clusters and 150 noise points distributed in data space.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	1.499	1.228	0.767	0.962
d_{means}	64.33	65.88	65.69	66.70

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	0.104	0.665	0.731	0.924
d_{means}	66.00	65.07	65.48	65.37

Table 7: Comparison between clustering methods on 20-dimensional data set with four clusters and 300 noise points distributed in data space.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	3.733	1.728	1.085	0.784
d_{means}	61.34	64.76	65.48	65.74

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	0.072	0.983	0.680	0.787
d_{means}	65.97	64.25	65.34	65.47

Table 8: Comparison between clustering methods on 20-dimensional data set with four clusters distributed in data space and 300 noise points containing values in all features.

	FCM_AW: $m=2, t=2$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=100$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=400$	PFCM_HDD: $m=2, \eta=2, t=2$ $a=0.5, b=1000$
d_{orig}	36.42	12.61	5.016	0.949
d_{means}	127.20	56.79	61.93	65.24

	MFCM: $m=2$	PMFCM: $m=2, \eta=2$ $a=0.5, b=100$	PMFCM: $m=2, \eta=2$ $a=0.5, b=400$	PMFCM: $m=2, \eta=2$ $a=0.5, b=1000$
d_{orig}	12.42	11.21	1.595	0.699
d_{means}	73.79	55.25	62.92	64.89

the case of data set where data points within clusters had real values in the first two dimensions, the most accurate final cluster centers were obtained by PMFCM and PMFCM_HDD for $b = 1000$. MFCM and FCM_AW produced cluster centers that were farther from each other than the original means of clusters.

4 CONCLUSION AND FUTURE WORK

Clustering high dimensional data is still a challenging problem for fuzzy clustering algorithms because of the concentration of distance phenomenon. Noise and outliers in data sets additionally make the partitioning of data difficult because they affect the computation of cluster centers. In this work, we analyzed two fuzzy clustering algorithms for high dimensional data from the literature and two possibilistic versions of the MFCM algorithm in terms of correct determining final cluster prototypes in presence of noise. Our experiments showed that MFCM produced the most accurate cluster centers as long as data items had real values in few features while its possibilistic versions PMFCM and PMFCM_HDD produced quite accurate final cluster centers independently from the number of features in that noise points had real values.

Although the performance results for PMFCM seem to be promising, before applying this method on real data sets, we plan to analyze the performance of fuzzy clustering algorithms in terms of sensitivity to different initializations because usually we do not have any a priori knowledge about the distribution of data in practical applications. In our future work, we also plan to apply other possibilistic clustering models to MFCM to make it less sensitive to outliers. Furthermore, we aim to apply fuzzy clustering algorithms for clustering text and image data and compare their performance with common crisp clustering algorithms.

REFERENCES

Beyer, K. S., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? In *Proceedings of the 7th International Conference on Database Theory*, ICDT '99, pages 217–235, London, UK, UK. Springer-Verlag.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.

Dave, R. N. and Krishnapuram, R. (1997). Robust clustering methods: A unified view. *IEEE Transactions on Fuzzy Systems*, 5(2):270–293.

Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57.

Himmelspach, L. and Conrad, S. (2016). A possibilistic multivariate fuzzy c-means clustering algorithm. In *Proceedings of the 10th International Conference on Scalable Uncertainty Management, SUM 2016*, pages 338–344.

Keller, A. and Klawonn, F. (2000). Fuzzy clustering with weighting of data variables. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(6):735–746.

Klawonn, F. (2013). What can fuzzy cluster analysis contribute to clustering of high-dimensional data? In *Proceedings of the 10th International Workshop on Fuzzy Logic and Applications, WILF2013, Genoa, Italy, November 19–22, 2013.*, pages 1–14.

Klawonn, F., Kruse, R., and Winkler, R. (2015). Fuzzy clustering: More than just fuzzification. *Fuzzy Sets and Systems*, 281:272–279.

Kriegel, H., Kröger, P., Schubert, E., and Zimek, A. (2009). Outlier detection in axis-parallel subspaces of high dimensional data. In *Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, PAKDD 2009*, pages 831–838.

Krishnapuram, R. and Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley. University of California Press.

Pal, N. R., Pal, K., Keller, J. M., and Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 13(4):517–530.

Pimentel, B. A. and de Souza, R. M. C. R. (2013). A multivariate fuzzy c-means method. *Applied Soft Computing*, 13(4):1592–1607.

Rehm, F., Klawonn, F., and Kruse, R. (2007). A novel approach to noise clustering for outlier detection. *Soft Computing*, 11(5):489–494.

Winkler, R., Klawonn, F., and Kruse, R. (2011). Fuzzy c-means in high dimensional spaces. *International Journal of Fuzzy System Applications*, 1(1):1–16.