International Journal of Management, IT & Engineering
(ISSN: 2249-0558)

CONTENTS

Editorial Board

Editorial Advisory Board

Title

THE PROBLEM OF OUTLIERS IN CLUSTERING

Author(s)

Prof. Thatimakula Sudha
Research Supervisor
Sri Padmavathi Women's University
Tirupati

Swapna Sree Reddy.Obili
PhD Research Scholar
Sri Padmavathi Women's University
Tirupati

ABSTRACT:

Clustering has been widely used in many applications including data mining, pattern recognition and machine learning. Noise is a major problem in cluster analysis, which degrades the performance of many existing methods. This paper is aimed at solving noise problems in data clustering.

Many existing clustering algorithms are sensitive to the presence of outliers. In this paper, a new robust operator is developed to attack this problem, namely the modified l_2 norm. There are many merits in using this new measure. No sensitiveuser-defined parameter is needed for this measure and it automatically assigns a small weight to the sample, which is far away from the cluster center. It is robust to outliers and has a theoretical 50% breakdown point. It can be solved without using an exhaustive search and can be extended to more general prototype, for example curve. We have tested this method with four synthetic and three real world datasets. Experiment results show that the method yields better results than other clustering algorithms.

1 Introduction:

The problem of outliers is an important topic in clustering and often appears in many different datasets. Outliers are samples placed far away from the normal samples and make large contributions to the criterion function. Various methods have been proposed to solve the outlier problem. They are classified into two classes: the robust operator and robust weighting function.

1.1 Robust Operator:

In the robust operator approach, a distance measure, which is robust to the presence of outliers, is used in the objective function. For example, in a one-dimensional dataset, the median is a robust operator to the problem of outliers. If a sample in the dataset is moved far away from the others, the median will not be affected. The l_2 norm is one of the widely used robust operators. K-medians and K-medoids methods use it as their distance measures [Chu *et al*. 2002; Kaufman and Rousseeuw 1990]. Although these methods are able to handle the outlier problem,

both measures are not differentiable at zero. In that case, the exhaustive search approach is used to find the solution, which confines the application of these methods. K-medians and K-medoids methods use this exhaustive search approach to find the solution. The solution form for both K-medians and K-medoids methods take the data samples in the dataset as representatives. If there are c groups in the dataset, there will be c data samples drawn from the data as representatives. Their optimization method is to check all the possible combinations of these data samples so that the minimum of the criterion function can be reached. This optimization approach is very time consuming. CLARA and CLARANS are two different extensions for K-medians method. They adopt a stochastic search method to reduce the complexity. However, the computation time is still huge. Apart from this, this optimization approach confines the applications of these methods since the prototypes of these methods must be points. They cannot be line, curve or other kind of prototypes. The l_1 norm and theleast median approaches are other two robust approaches to handle the outlier problem [Kersten 1995; Nasraoui and Rojas 1997]. However, similar to the l_2 normapproach, the objective functions in both approaches are not differentiable and this also confines their applications. They cannot be extended to other prototypes detection including line, curve, etc.

1.2 Robust Weighting Function:

Another widely used class for the outlier problem is the robust weighting function. In this class, the outlier problem is resolved by assigning a small weight to the potential outliers. The weight of the potential outliers is determined by measuring the distance either between the potential outliers and the cluster centers or between the potential outliers and others samples.

There are many methods that assign a small weight to potential outliers, which are distant from cluster centers [Barnett and Lewis 1994; Dave 1991; Melek *et al*. 1999; Miller and Browning 2003; Schneider 2000; Sen and Dave 1998]. One of the well-known examples of these methods is the possibilistic membership [Barni *et al*. 1996; Krishnapuram and Keller 1993; Krishnapuram *et al*. 1995a; Krishnapuram *et al*. 1995b; Masulli and Rovetta 2006; Pal *et al*. 1997; Zhang and Leung 2004]. The weighting function is defined as the exponential function of the negative squared $l_2$norm between the sample and cluster center. If the measured distance is large, the sample is a potential outlier and a small weight will be assigned. Other than this type of assignment, computing the distance between the potential outliers and other samples is also very

common [Chintalapudi and Kam 1998a; Chintalapudi and Kam 1998b]. They measure the distances among samples. If the distance is large, a small weight will be assigned. Although these techniques are able to handle the outlier problem, the concept of a large distance is always controlled by the user. Moreover, there is a lack of theoretical justification to these weighting approaches. We have found it difficult to check the robustness of these approaches to the presence of outliers.

Because of the limitations of these existing methods, there is a need to develop a new technique having the following properties: (i) the new method can be solved without using an exhaustive search, (ii) the proposed objective function is differentiable and is able to extend to more general prototype such as curve. (iii) no sensitive user-defined parameter is needed, and (iv) the robustness of this technique towards outliers can be shown by theoretical justification. In this paper, a new robust operator is developed to deal with the outlier problem having all these properties.

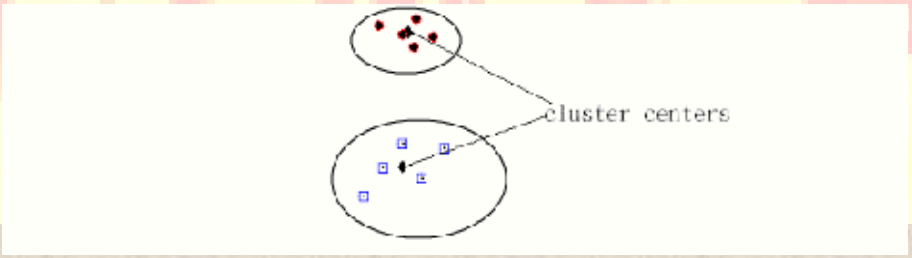
The organization of this paper is given as follows. In Section 2, we show the limitations of existing methods in the presence of outliers. After that, a new robust distance metric, called the modified l_2 norm, is developed to resolve this problem. Byembedding the modified l_2 norm into the fuzzy c-means algorithm, a robust clusteringalgorithm is formed as discussed in Section 3. Experiment results show that the new technique is able to resolve the outlier problem in Section 4. In Section 5, a parameter study of the proposed method is shown. Finally, conclusions will be given in Section 6.

2 Outlier and the Squared Euclidean Distance:

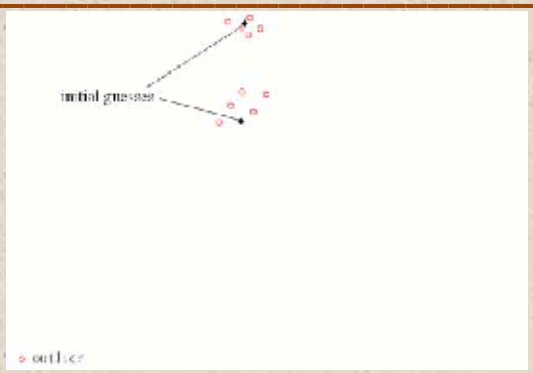
The squared Euclidean distance is one of the widely used distance measures for data clustering. However, the use of this measure can lead to serious problems in outlier problems, which is well known [Barnett and Lewis 1994; Huber 1981]. In this section, we discuss the limitation of this distance measure using the FCM algorithm. Then, a new measure is developed to resolve the outlier problem.

2.1 Limitation of the Squared Euclidean Distance:

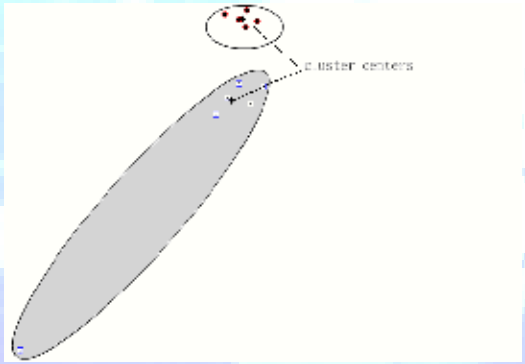
In the FCM algorithm, the classification result is dependent on the placement of cluster centers. In Equation $v_k^p = \frac{\sum_{i=1}^{n} \mu_{ik}^m x_i}{\sum_{i=1}^{n} \mu_{ik}^m}$, there is a direct relationship between x_i and v_k.If v_k is the closest center to the data point x_i, a larger weight will be assigned. If thatpoint is an outlier, the cluster center may shift towards it. An example is given as follows. Figure 2.1(a) shows a dataset with two groups. The clustering result using the FCM algorithm is also shown in Figure 2.1(a). The samples in the same manually drawn circles mean that they are in the same group. Apparently, the two elliptical clusters are classified correctly as two different groups. Next, we add an outlier to this dataset and define two initial clusters (Figure 2.1(b)). The upper cluster center is v_1 while the lower center is v_2. The desired distributions of the twogroups are given by gray and white regions as in Figure 2.1(c) respectively. Since the outlier is closer to the cluster center v_2 than v_1, the contribution of the outlier x_i tov$_2$ will be larger and thus v_2 will be dragged towards this outlier. Figure 2.1(d)shows the clustering result from the FCM algorithm. The two elliptical clusters are classified as one group while the outlier is classified as another group. Next, we move the outlier from the left to the right hand corner. The clustering result using the FCM algorithm is shown in Figure 2.1(e). A similar result is obtained in Figure 2.1(d) for the same reason.



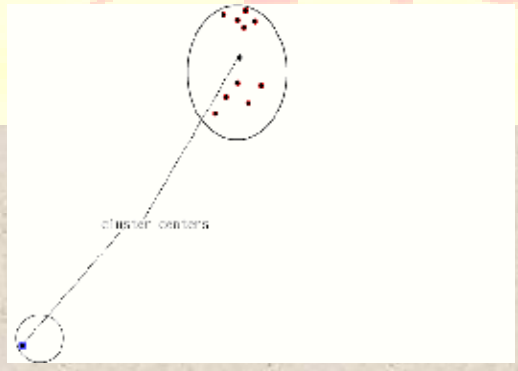
(a) Dataset without any outlier.



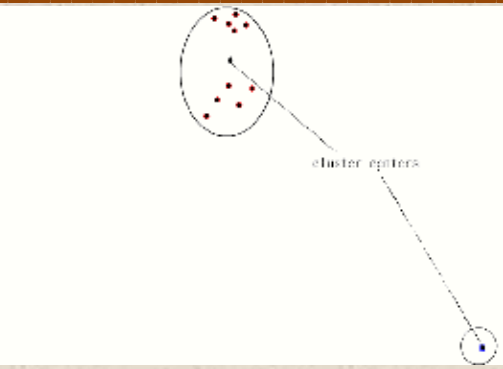
(b) Dataset with an outlier.



(c) Regions of two classes for the data in Figure 2.2.1(b).



(d) Clustering result from the FCM algorithm with an outlier on the left.



(e) Clustering result from the FCM algorithm with an outlier on the right.

(f)

Figure 2.1. Illustration of the influence of outliers on the FCM clustering algorithm. (The samples in the manually drawn circles mean that they are in the same group.)

2.2 The Modified l_2 Norm:

It is well known that the median operator is a robust solver of the outlier problem. However, the median is only uniquely defined in the one-dimensional (1D) space. There are several different definitions for higher dimensional medians, including Oja's Simplex median [Small 1990], halfspace median [Small 1990] and spatial median [Abdous and Theodorescu 1992; Brown 1983; Chakraborty *et al*. 1998; Milasevic and Ducharme 1987]. A widely used high dimensional median is the spatial median, which is the median defined by the l_2 norm. However, the spatial median isnot differentiable at certain points. In this section, we introduce a modified version of a spatial median based on the modified l_2 norm.

The concept of the spatial median is to find a point v such that it has the shortest distance, measured by the l_2 norm, to all data points x_i [Brown 1983]. That is,we have to find v such that the following equation is minimized.

$$E_0(v) = \sum_{i=1}^{n} \|x_i - v\|_2 \quad . \qquad \textbf{(2.1)}$$

In the one-dimensional case, the quantity v has to satisfy the following equation

$$\sum_{i=1}^{n} \text{sign}\ (x_i - v) = 0\ , \qquad (2.2)$$

which is the zero for the first order derivation of Equation (2.1) and sign(y) is the sign of the variable y. Apparently, v is the median of the dataset. However, the $l_2$norm can be equal to zero at $v=x_j$ for some j. We cannot employ such an equation forthe alternative optimization (AO) method in clustering. To resolve this, we modify the $l_2$norm as $|\bullet|_\varepsilon$ where $|u|_\varepsilon = \sqrt{\varepsilon^2 + \|u\|_2^2}$ and ε is a small constant, which is taken to be10^{-2} in our experiments. We call this the modification of the l_2 norm (or in short, themodified l_2 norm). This new measure is not a norm because it is not zero when u = 0. The quantity $|u|_\varepsilon$ is differentiable at all points. It has been widely used in the PartialDifferential Equations-based image processing algorithms and it produces good results [Chan and Shen 2001a; Chan and Shen 2001b; Strong and Chan 2003]. The modified spatial median is then defined as follows:

$$E_\varepsilon(v) = \sum_{i=1}^{n} |x_i - v|_\varepsilon\ , \qquad (2.3)$$

Where $|y|_\varepsilon = \sqrt{\varepsilon^2 + \|y\|_2^2}$. If $\varepsilon = 0$, $E_\varepsilon(v) = E_0(v)$.

Also, in Theorem A1 Appendix A, weshow that the difference between the solutions of $E_\varepsilon(v)$ and $E_0(v)$ is within 2ε.

Theorem A1: Assume that $E_\varepsilon(v)$ is a C^1 function, which is a continuous differentiablefunction. If u and v are the minimum of $E_0(v)$ and $E_\varepsilon(v)$ respectively, then $\|u- v\|_2 \leq 2|\varepsilon|$.

Taking the first order derivative of $E_\varepsilon(v)$ to be zero, we have

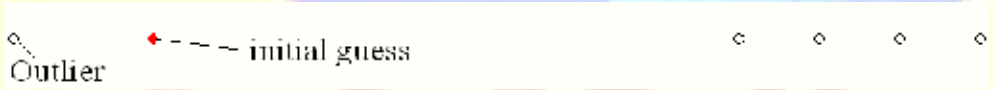
$$\sum_{i=1}^{n} \frac{v - x_i}{|v - x_i|_\varepsilon} = 0\,. \qquad (2.4)$$

As $|y|_\varepsilon \neq 0$ for $\forall y$, , we can find the value v through the fixed-point iteration.

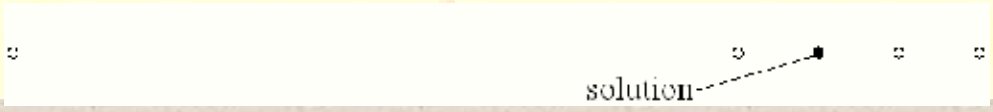
Now, we illustrate this equation through the dataset shown in Figure 2.2(a). It is a one-dimensional data and consists of five points Xu= {-10 -1 0 1 2}. If we ignore the small constant ε, each term in the above equation represents a unit vector. As all five data points are placed on the same horizontal axis, this equation requires a cluster center such that the number of data points on each side is two and the value v is taken as the middle one. That is, there must be two unit vectors pointing to the left and two pointing to the right so that their summation is zero as required by Equation (2.4). For a multi-dimensional problem, the criterion function requires the unit vectors to be summed to zero in all directions. As the equation holds as long as there are two input samples on each side of the cluster center, the solid dot in Figure 2.2(b) is the solution even if the outlier is moved further away. For the multi-dimensional problem, the modified spatial median is still robust to outliers and has a 50% breakdown point, which means the measure is very robust to the presence of outliers. Also, the modified spatial median has a unique solution. The properties are given in the following theorems.

Theorem A2: The estimator in Equation (2.4) for function $E_\varepsilon(v)$ has a 50%breakdown point.

Theorem A3: There is one and only one solution to $E_\varepsilon(v)$ if the dataset X is not on aline.



(a) Five input data points and an initial guess.



(b) The solution given by the modified spatial median.

Figure 2.2. The modified spatial median.

3 The Modified l_2 Norm Based FCM (l_{2m}-FCM) Algorithm:

We alter the FCM model by replacing the standard Euclidean distance $\|x\|_2^2$ with the function $|x|_\varepsilon = \sqrt{\varepsilon^2 + \|x\|_2^2}$.

Good clusters, which are robust to the presence of outliers, are defined by fuzzy partitions and cluster centers that locally minimize the objective function $J_{l2\text{-}FCM}(U,V;X)$: $M_{FCM} \times R^{cd} \rightarrow R^+$ is

$$J_{12-FCM}(U,V;X) = \sum_{i=1}^{n}\sum_{k=1}^{c} \mu_{ik}^{m} |x_i - v_k|_\varepsilon \quad , \qquad (3.1)$$

where $M_{FCM} = \left\{ \mu_{ik} \in R^{nc} \middle| \mu_{ik} \in [0,1]\ \forall i,k; \sum_{k=1}^{c} \mu_{ik} = 1 \forall i; 0 < \sum_{i=1}^{n} \mu_{ik} < n \forall k \right\}$. We call Equation (3.1) the criterion functionfor the modified l_2-norm based FCM (l_{2m}-FCM) algorithm. Here m is the fuzzyparameters and we choose m=1.5 in all the experiments. In the experiments below, we use the same m in both the proposed and the FCM algorithms. We use the AO method to obtain the local optimal solution. The necessary condition for the variable v_k is given asfollowing. Setting the derivatives of Equation (3.1) with respect to the cluster centers to zero, we have

$$\frac{\partial J_{12-FCM}(U,V;X)}{\partial v_k} = \sum_{i=1}^{n} \mu_{ik}^{m} \frac{\partial \sqrt{\varepsilon^2 + \|x_i - v_k\|^2}}{\partial v_k}$$

$$= \sum_{i=1}^{n} \mu_{ik}^{m} \frac{2\,(v_k - x_i)}{2\sqrt{\varepsilon^2 + \|x_i - v_k\|^2}}, \qquad (3.2)$$

$$= \sum_{i=1}^{n} \mu_{ik}^{m} \left(\frac{v_k - x_i}{|v_k - x_i|_\varepsilon} \right)$$

$$= 0.$$

As the necessary condition for vk is a nonlinear equation, we use a fixed pointapproach to estimate the solution. That is, the variable v_k is updated by the equation:

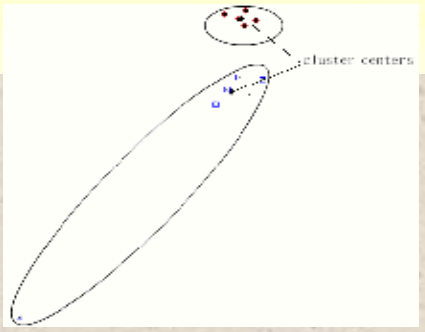
$$v_k^{P} = \frac{\sum_{i=1}^{n} \mu_{ik}^{m} \left(\frac{x_i}{\left| v_k^{p-1} - x_i \right|_{\varepsilon}} \right)}{\sum_{i=1}^{n} \mu_{ik}^{m} \left(\frac{1}{\left| v_k^{p-1} - x_i \right|_{\varepsilon}} \right)}, \qquad (3.3)$$

where $v_k^{(p)}$represents the k-th cluster center at the p-th iteration. In Appendix C, we will show that this method converges. Once this converges, the update equation in Equation (3.3) becomes the necessary condition shown in Equation (3.2). For the fuzzy membership function, Bezdek has given a general formula for updating the membership function [Bezdek 1981] for a general distance function $d(x_i,v_k)$. If wetake $d(x_i,v_k)$ as the modified l_2 norm, the update equation for μ_{ik} will become:

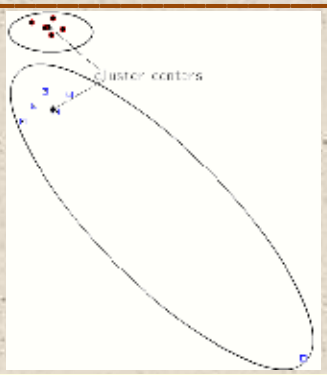
$$\mu_{ik} = \frac{\left| X_i - v_k^{p-1} \right|_{\in}^{-1/\,m-1}}{\sum_{k=1}^{c} \left| X_i - v_k^{p-1} \right|_{\in}^{-1/\,m-1}}, \qquad (3.4)$$

Where $m>1$ and $v_k^{(p)}$ represents the k-th cluster center at the p-th iteration. Since thedenominator is bounded below by ε, we are therefore able to solve this equation and compute v_k using the AO method.

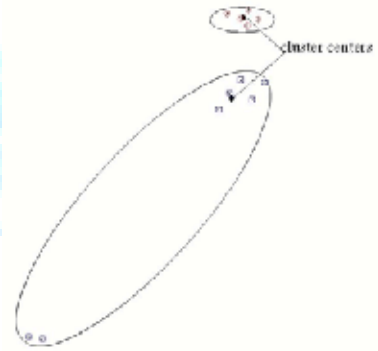
We apply this algorithm to the dataset in Figures 2.1(d) and (e). The clustering results are shown in Figures 3.1(a) and (b) respectively. The results show that the l_{2m}-FCM algorithm is able to handle the outlier problem successfully.



(a) Result from the l_{2m}-FCM algorithm for the dataset in Figure 2.2.1(d).



(b) Result from the l_{2m}-FCM algorithm for the dataset in Figure 2.2.1(e).



(c) Result from l_{2m}-FCM algorithm for the dataset with two outliers.

Figure 3.1. Clustering results obtained using the l_{2m}-FCM algorithm. (Thesamples in the same manually drawn circle mean that they are in the same cluster.)

4 Experiment Results:

In this section, we show the robustness of the l_{2m}-FCM algorithm by: (i) thesynthetic dataset with outliers and (ii) the real world datasets.

4.1 Synthetic Datasets for the Outlier Problem:

The l_{2m}-FCM algorithm is tested by four different synthetic datasets. Outliersare present in each of these datasets.

Example A: Dataset Containing Two Groups (2-gps): This dataset consists of twogroups with each group having 14 samples, which are generated by the Gaussian distributions with two means (11,11) and (11,102) respectively and they share the same covariance matrix

$$\begin{bmatrix} 10 & 0 \\ 0 & 0.8 \end{bmatrix}.$$

Two outliers are added on its upper side. The dataset is shown in Figure 4.1(a).

Example B: DatasetContaining Three Groups (3-gps): This dataset consists of three groups and they are generated by the Gaussian distributions with three means (0,0), (15,0) and (30,0) and each group contains 500 points. They share the same covariance matrix $\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$.

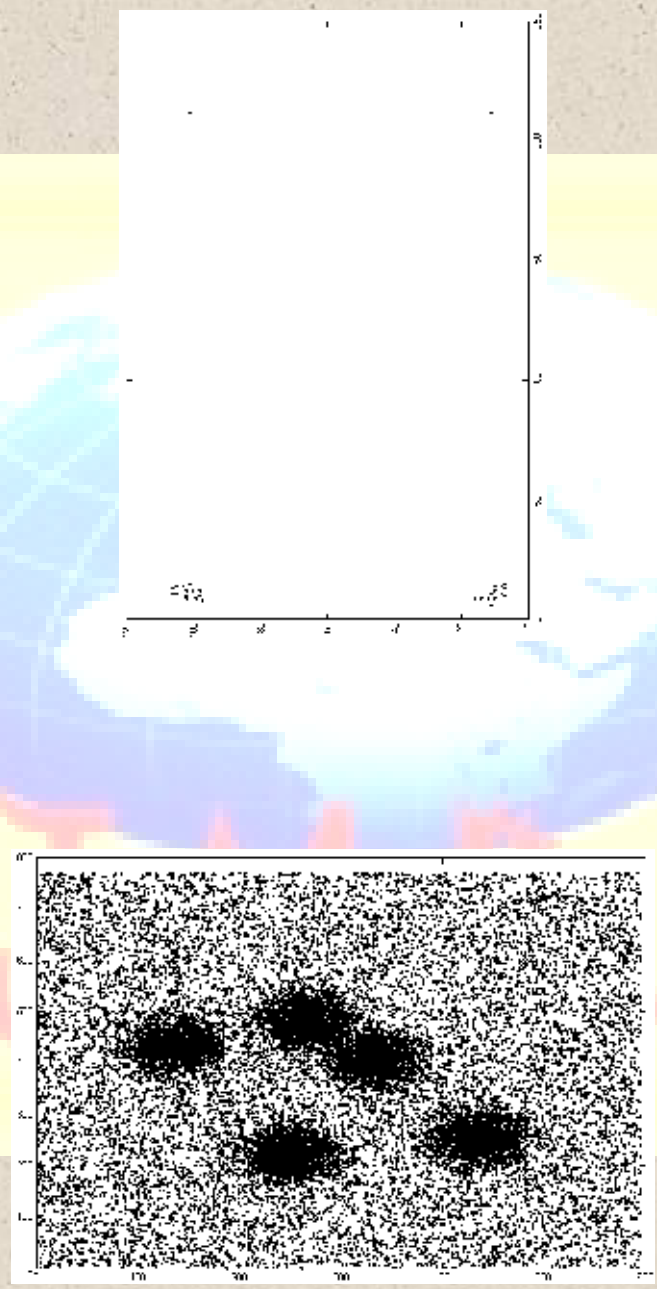
Also, 2500 noise points are added, generated by the uniform distribution. The dataset is shown in Figure 4.1(b). The noise here is also a type of outlier. As by our definition, outliers are caused by contaminants. In this example, the contaminant is caused by the uniform distribution.

Example C: Dataset Containing Groups (5-gps): This dataset consists of five groupswhich are generated by the Gaussian distributions with five means (267.01,479.69), (435.78,251.84), (332.71,405.67) (138.73,432.22) and (250.27,220.58) and each group contains 2900 points. They share the same covariance matrix $\begin{bmatrix} 603 & 5.5 \\ 5.5 & 867 \end{bmatrix}$.

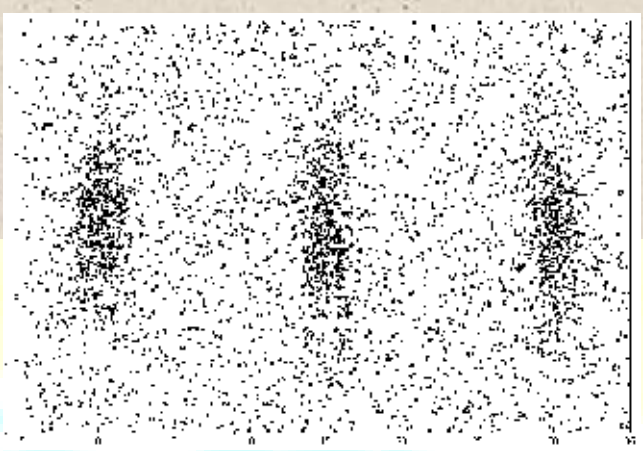
Also, 10700 noise points are added, generated by the uniform distribution. The dataset is shown in Figure 2.4.1(c).

Example D: Dataset Containing Six Groups (6-gps): This dataset consists of sixgroups which are generated by the Gaussian distributions with six means (267.01,479.69), (435.78,251.84), (332.71,405.67) (138.73,432.22), (250.27,220.58) and (507.98,445.35); and each group contains 2900 points. They share the same covariance matrix $\begin{bmatrix} 603 & 5.5 \\ 5.5 & 867 \end{bmatrix}$.

Also, 10700 noise points, generated by the uniform distribution are added. The dataset is shown in Figure 4.1(d).

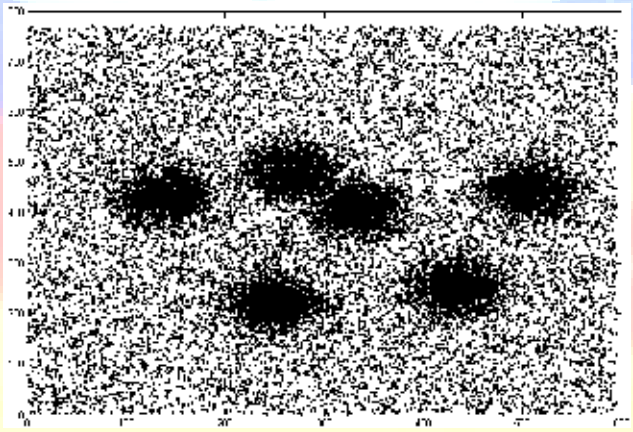


(a) A dataset with two groups with two outliers.



(b) A dataset with three groups with noise.

(c) A dataset with five groups with noise.



(d) A dataset with six groups with noise.

Figure 4.1. Synthetic datasets with outliers.

We compare the performance of the l_{2m}-FCM algorithm with the original FCMand the HCM algorithms. Our evaluation method is given as follows. The method is run 30 times on each of the datasets. In each run, a randomly generated initialization in the convex hull of the dataset is used as an initial guess for a cluster center. The number of clusters 'c' in the algorithm

is given by user, which we set it as the number of groups in the dataset. Then, the mean square error (MSE) between the cluster centers and the ground truth cluster centers is computed. This evaluation measure is used only if the ground truth cluster centers in the dataset are available and this will be used only in the synthetic dataset. As outliers can drive the estimated cluster centers far away from the ground truth, the MSE effectively reflects the sensitivity of the method to the presence of outliers. Then, we compute the mean and standard derivation among these 30 runs. Table 4.1 shows the MSEs using different methods. The number without parenthesis is the average MSE. The number in the parenthesis is the standard derivation of the MSEs for the 30 runs. We can see that the l_{2m}-FCMalgorithm yields the smallest average MSEs compared to the other methods. For the datasets 2-gps and 3-gps, the standard derivations of the MSEs of the l_{2m}-FCMalgorithm are almost zero. This implies that the method produces the same results in the 30 runs. For the 5-gps and 6-gps, other methods have large average MSEs in the 30 runs. That means these noisy datasets seriously degrade their performance. The l_{2m}-FCMalgorithm produces the smallest average MSEs. The minimum MSEs to the 5-gps and 6-gps datasets produced by our method are 7.7362 and 9.1417 respectively. These are very small compared with the Orig-FCM algorithm.

	2-gps	3-gps	5-gps	6-gps
HCM	23.135 (20.55)	4.8033 (1.645)	97.283 (24.33)	103.98 (31.56)
Orig-FCM	60.599 (57.73)	2.3548 (0)	64.058 (0)	63.707 (0)
l_{2m}-FCM	0.548 (0)	0.7568 (0)	19.101 (25.48)	20.311 (22.63)

Table 4.1. The MSEs between the estimated cluster centers and the ground truth.

4.2 Real World Datasets:

We show the robustness of the l_{2m}-FCM algorithm to real world datasets. Theinformation of the datasets is given in Table 4.2. All the real world datasets are downloaded on a website [Blake *et al*. 1998]. As some of the features make much larger contributions than others, we apply the normalization technique. This technique is to normalize each feature of the dataset so that it has zero mean and unit variance.

Name	Full Name	No. of features	Total no. of samples	No. of groups	Normalization
Iris	Iris Plant Database	4	150	3	Yes
Wbcd	Wisconsin Breast Cancer Databases	9	683	2	No
Wine	Wine Recognition	3	178	3	Yes

Table 4.2. Information of datasets.

We adopt cross validation to evaluate the performance of the methods. Each dataset is first randomly divided into two halves. One is for training and the other is for testing. The algorithms are run on the training set and the prototypes are obtained. Then, the testing set is used to evaluate the clustering results based on these prototypes. Each sample in the testing set is assigned to the estimated cluster, which is the closest prototype with respect to this sample. The misclassification rate is computed based on the number of wrongly labeled samples in the testing set. This process is repeated 30 times. Table 4.3 shows the ground truth of the datasets. The ground truth is obtained in the following way. In cross validation, the dataset is split into two halves. One is for testing and one is for training. We use the training set label to obtain the cluster centers, for which we take the mean of each group as a cluster center. This procedure is different from the estimated cluster centers generated by a clustering algorithm, which the goal of using the clustering algorithm is to estimate the label. Using the training set to estimate the ground truth cluster centers is an idea situation. This is because when we apply a clustering algorithm to the training set, the label is not taken into account. So, these ground truth cluster

centers are probably generated the smallest classification error rate in prototype-based approach. Moreover, by measuring the standard derivation of the 30 runs of the misclassification rate using the ground truth cluster centers can reflect the sensitive of the known labels to partition. In Table 4.3, we can see that the datasets wbcd and wine have small standard derivation. This means that the datasets are less sensitive to the partition. For the iris dataset, it has large mean, which shows that an "ideal" cluster centers may not be able to yield an error rate smaller than 14%. Also, it has a large standard derivation, which means the performance is varied much on the partition.

	Mean	Standard Derivation
Iris	13.3778	4.2811
Wbcd	3.6266	0.8131
Wine	3.2197	1.3036

Table 4.3. Error rates using the ground truth for the real world datasets in percentages.

The clustering results using different methods are given in Table 4.4. There are two numbers in the table. The number without parentless is the mean error rate while the one with parentless is the standard derivation. The l_{2m}-FCM algorithmyields the smallest classification error rates compared to the other methods.

	Iris	Wbcd	Wine
HCM	20.36 (7.87)	3.94 (1.21)	14.47 (11.6)
Orig-FCM	15.69 (3.05)	4.37 (0.85)	4.55 (1.98)
l_{2m}-FCM	14.22 (2.56)	3.87 (0.85)	4.46 (1.70)

Table 4.4. Classification error rates for real world datasets using different methods in percentages.

5 Parameter Study of the l_{2m}-FCM Algorithm:

In this section, we test the sensitive of the proposed method and the user-defined parameter e of the proposed method. We use the synthetic datasets given in Section 4.1 for comparison. Again, we apply the proposed method in each of the dataset with 30 runs with different initializations. The user-defined parameter is varied from 10^{-2} to $10^{-0.1}$. The results are given in Table 5.1. The numbers with andwithout parentless are the mean and standard derivation of the error rates, respectively. In this table, we can see that the performance of the proposed method does not change much by varying this parameter.

	2-gps	3-gps	5-gps	6-gps
10^{-2}	0.548 (0)	0.7568 (0)	19.101 (25.48)	20.31 (22.63)
$10^{-1.9}$	0.5287 (0.5332)	0.2359 (0.2379)	7.8028 (16.657)	9.9698 (22.321)
$10^{-1.8}$	0.3525 (0.5013)	0.1573 (0.2237)	7.7800 (21.4707)	5.6945 (15.5635)
$10^{-1.7}$	0.2644 (0.4598)	0.1180 (0.2052)	4.2398 (12.9953)	5.4293 (14.1043)
$10^{-1.6}$	0.2115 (0.4244)	0.0944 (0.1894)	4.9387 (17.3878)	4.5147 (12.7995)
$10^{-1.5}$	0.1762 (0.3952)	0.0787 (0.1764)	2.8802 (12.0196)	4.6982 (15.1867)
$10^{-1.4}$	0.1511 (0.3709)	0.0674 (0.1655)	2.4688 (11.1693)	4.3004 (15.0519)
$10^{-1.3}$	0.1322 (0.3504)	0.0590 (0.1564)	2.7080 (12.5069)	2.9094 (11.6984)

$10^{-1.2}$	0.1175 (0.3329)	0.0524 (0.1486)	2.2209 (10.6452)	3.3151 (12.2574)
$10^{-1.1}$	0.1057 (0.3177)	0.0472 (0.1418)	2.5403 (11.3917)	2.2564 (10.1536)
$10^{-1.0}$	0.0961 (0.3042)	0.0429 (0.1358)	1.4187 (7.6708)	2.0106 (10.2073)
$10^{-0.9}$	0.0880 (0.2921)	0.0393 (0.1306)	1.3005 (7.3538)	2.6006 (10.6275)
$10^{-0.8}$	0.0810 (0.2810)	0.0363 (0.1258)	2.2367 (12.2682)	1.2531 (6.5391)
$10^{-0.7}$	0.0749 (0.2704)	0.0337 (0.1216)	1.3541 (8.9967)	2.1218 (10.4156)
$10^{-0.6}$	0.0694 (0.2601)	0.0314 (0.1177)	1.2209 (7.2829)	1.3535 (7.2466)
$10^{-0.5}$	0.0645 (0.2499)	0.0295 (0.1142)	1.3138 (7.6621)	1.9364 (10.0976)
$10^{-0.4}$	0.0599 (0.2397)	0.0278 (0.1112)	0.6602 (3.9672)	1.3889 (8.9328)
$10^{-0.3}$	0.0556 (0.2295)	0.0263 (0.1086)	0.8850 (6.7699)	1.0289 (6.1376)
$10^{-0.2}$	0.0517 (0.2197)	0.0251 (0.1067)	0.7503 (5.5423)	0.9969 (6.4202)
$10^{-0.1}$	0.0484 (0.2114)	0.0243 (0.1061)	1.4542 (9.9454)	1.0116 (6.4976)

Table 5.1. The MSEs between the estimated cluster centers and the ground truth cluster centers using the l_{2m} -FCM algorithm.

6 Conclusion:

This paper is devoted to study the outlier problem using the fuzzy c-means (FCM) algorithm. The FCM algorithm often produces inaccurate results if the inputdata contain outliers. In this Paper , a new robust distance metric is developed, which is called the modifiedl_2 norm. The merits using this approach compared to existingmethods are that it solves the outlier problem without introducing a sensitive user-defined parameter and this method can be solved without using an exhaustive search. Also, based on the analysis of robust statistics, the modifiedl_2 norm is robust tooutliers and has a 50% breakdown point. By applying this modifiedl_2 norm to theFCM algorithm, a new robust clustering algorithm is developed, called thel_2 norm-based fuzzy c-means clustering (l_{2m}-FCM) algorithm.

REFERENCES:

- [Ball and Hall 1965] - G. H. Ball and D. J. Hall, "A Clustering Technique for Summarizing Multivariate Data_, *Behav. Sci.*, Vol. 12, pp. 153-155, 1967.
- [Barbara *et al*. 2002] - D. Barbara, Y. Li, and J. Couto. Coolcat, "An Entropy Based Algorithm for Categorical Clustering," *In Proceedings of the Eleventh International Conference on In Formation and Knowledge Management*, ACM Press,pp. 582-589, 2002.
- [Barnett and Lewis 1994] - V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3rd edition, John Wiley, 1994.
- [Barni *et al*. 1996] - M. Barni, V. Cappellini and A. Mecocci, "Comments on a Possibilistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems*, Vol. 4, pp. 393-396, 1996.
- [Blake *et al*. 1998] - C. L. Blake, D. J. Newman, S. Hettich and C. J. Merz, UCI repository of Machine Learning Databases, http://www.ics.uci.edu/~mlearn/MLSummary.html , 1998.
- [Bobrowski and Bezdek 1991]- L. Bobrowski and J. Bezdek, "c-Means Clustering with the l_1 and l_∞ Norms," IEEE Trans. SMC, Vol. 21(3), pp. 545-554, 1991.
- [Burden and Faires 1997] - R. Burden and J. Faires, *Numerical Analysis*, 6th edition, Pacific Grove, Calif.: Brooks/Cole Pub, 1997.

- [Chintalapudi and Kam 1998b] - K. K. Chintalapudi and M. Kam, _ The Credibilistic Fuzzy C-means Clustering Algorithm, " *In Proc. IEEE Int. Conf.Systems Man Cybernetics*, pp. 2034–2040, 1998.
- [Dubes and Jain 1979] - R. Dubes and A. Jain, "Validity Studies in Clustering Methodologies," *Pattern Recognition*, Vol. 11, pp. 235-254, 1979.
- [Evans 1998] - L. C. Evans, *Partial Differential Equations*, Providence, R.I.: American Mathematical Society, 1998.
- [Fisher 1987] - D. H. Fisher, "Knowledge Acquisition via Incremental Conceptual Clustering," *Machine Learning*, Vol. 2, pp. 139-172, 1987.
- [Huber 1981] - P. Huber, *Robust Statistics*, Wiley, New York, 1981.
- [Jain *et al*. 1999] - A. K. Jain, M. N. Murty and P. J. Flynn, "Data Clustering: a Review," *ACM Computing Surveys*, Vol. 31, No. 3, pp.264-323, 1999.
- [Jajuga 1991] - K. Jajuga, "L1 Norm-based Fuzzy Clustering," *Fuzzy Sets System*, Vol. 39, pp. 43-50, 1991.
- [Kaplan 1991] - W. Kaplan, *Advanced Calculus*, 4th Edition, Reading, Mass.: Addison-Wesley, 1991.
- [Kaufman and Rousseeuw 1990] - L. Kaufman and P. J. Rousseeuw, *FindingGroups In Data: An Introduction To Cluster analysis*, New York: Wiley, 1990.
- [Kim 2001] - S. Kim, "An O(N) Level Set Method for Eikonal Equations," *SIAMJournal on Scientific Computing*, Vol. 22, pp. 2178-2193, 2001.
- [Krishnapuram and Keller 1993] - R. Krishnapuram and J. M. Keller, "A Possibilistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems*, Vol. 1, pp. 98-110, 1993.
- [Lam and Yan 2004] - B. Lam, and H. Yan, "Robust Clustering Algorithm for Suppression of Outliers," *International Symposium on Intelligent Multimedia, Video& Speech Processing*, accepted, October 20-22, 2004.
- [Liu *et al*. 2000] - B. Liu, Y. Xia, and P. S. Yu., Clustering Through Decision Tree Construction, " *In Proceedings of the Ninth International Conference on Information and Knowledge Management*, pages 20-29. ACM Press, 2000.
- [Melek *et al*. 1999] - W. W. Melek, M. R. Emami, A. A. Goldenberg, "An Improved Robust Fuzzy Clustering Algorithm," *IEEE International Fuzzy Systems ConferenceProceedings*, Vol. 3, pp. 1261 - 1265, 1999.

- [Milligan and Cooper 1985] - G. Milligan and C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika*, Vol. 50, pp. 159-179, 1985.
- [Miyamoto and Agusta 1998] - S. Miyamoto and Y. Agusta, "Algorithms for L1 and Lp Fuzzy C-means and Their Convergence," *in Studies in Classification, DataAnalysis, and Knowledge Organization: Data Science, Classification, and Related Methods*, Japan: Springer-Verlag, pp. 295-302, 1998.
- [Nasraoui and Rojas 2006] -O. Nasraoui, Carlos Rojas, "Robust Clustering for Tracking Noisy Evolving Data Streams," *SDM*, pp. 618-622, 2006.
- [Rudin 1976] - W. Rudin, *Principles of Mathematical Analysis*, 3rd Edition, New York : McGraw-Hill, 1976.
- [Wilson and Watkins 1990] - R. Wilson and J. Watkins, Graphs: *An IntroductoryApproach: a First Course in Discrete Mathematics*, New York: Wiley, 1990.
- *UCLA CAM Report*, 03-09,http://www.math.ucla.edu/applied/cam, 2003.
- [Zou and Yan 1999] - J. Zou and H. Yan, "Extracting Strokes from Static Line Images Based on Selective Searching," *Pattern Recognition*, Vol. 32, pp. 935-946, 1999.