

An Improved Algorithm of Rough K-Means Clustering Based on Variable Weighted Distance Measure

Tengfei Zhang^{1*}, Long Chen¹ and Fumin Ma²

¹*College of Automation, Nanjing University of Posts and Telecommunications,
Nanjing China*

²*College of Information Engineering, Nanjing University of Finance and
Economics, Nanjing, China*

**(Corresponding Author) tfzhang@126.com*

Abstract

Rough K-means algorithm has shown that it can provides a reasonable set of lower and upper bounds for a given dataset. With the conceptions of the lower and upper approximate sets, rough k-means clustering and its emerging derivatives become valid algorithms in vague information clustering. However, the most available algorithms ignore the difference of the distances between data objects and cluster centers when computing new mean for each cluster. To solve this issue, an improved algorithm of rough k-means clustering based on variable weighted distance measure is presented in this article. Comparative experimental results of real world data from UCI demonstrate the validity of the proposed algorithm.

Keywords: *Rough k-means, Variable Weighted Distance Measure, Rough Set Theory.*

1. Introduction

Clustering is a process that can partition data objects into different clusters. It requires objects are similar in inner-cluster, and different inter-cluster. Clustering process distinguishes data objects with difference measures, such as distance based method. Clustering methods can be roughly divided into: partitioning method, hierarchical method, density method, grid method and model method. Currently, the clustering algorithm have been applied in a number of areas including data mining, statistics, machine learning, spatial database technology [1].

Traditional k-means clustering partitions a group of objects into a number of non-overlapping sets. Pawan Lingras finds that k-means clustering results often have a vague, rough distinguish unclear boundary in web data mining [2, 3]. He points out that the ‘hard partitioning’ k-means clustering methods are not able to meet the needs of vague data clustering. Pawan Lingras combines k-means clustering algorithm with rough set theory, each cluster is seen as a rough set, each object is either determined to belong to the lower approximation set of one cluster, or belong to the upper approximation sets of multiple clusters [2, 3].

Rough k-means clustering algorithm improves the accuracy of clustering boundary, but its mean formula only considers that the upper and lower approximation set of the cluster is non-empty, or just the boundary set is empty. And traditional rough k-means algorithm only considers the indiscernibility of boundary objects and unified weights of multiple objects, ignoring the differences of the objects in the cluster.

Many improvements to the rough k-means algorithm [4-14, 8-20] are emerging in the past ten years. Georg Peters improves rough k-means algorithm center means iterative formula [5, 6]. Only when the lower approximation or upper approximation

set is empty, rough weights is 1, standing for objects surely belongs to these sets. And Georg Peters using the relative distance as a dissimilarity criteria instead of the absolute distance which the rough k-means, effectively reduces the influence of outliers [5]. Literatures [7, 8] introduced rough fuzzy k-means and fuzzy rough k-means method. These methods combine characteristics of rough k-means and fuzzy k-means, using fuzzy membership to measure the inter-cluster contribution of objects and improve accuracy of clustering boundary based on rough k-means clustering.

The improved algorithms can improve the clustering precision, but still there are some problems. Especially, the same weight was used for all the data objects in a lower or upper approximate set when computing new mean for each cluster. For example, in the lower approximation set, the distance from the mean center of each object is different, indicating distinction of the closeness. The same weight value will certainly results in a greater mobility of the center mean point, thus affecting the clustering precision. To solve this problem, this article proposes a rough k-means clustering based on variable weighted distance measure. Each object, according to the distance from the center, is given different weighted coefficient. The closer the object is to the center, the higher weighted coefficient is, and the higher the contribution is in the center iteration.

The subsequent sections of this paper are organized in the following: Section 2 will introduce some basic theory, including rough set theory, rough k-means clustering, as well as some major improvements. Section 3 gives a new rough k-means clustering based on weighted distance measure and implementation steps. Section 4 will simulate with UCI datasets and compare with the current mainstream of rough k-means clustering algorithm. Section 5 is conclusions.

2. Related k-means Clustering Algorithms

In this section, the related work about classic hard k-means, rough k-means and the improvements will be briefly discussed.

2.1. Classic Hard k-means Algorithm

The classic k-means clustering algorithm is one of the best-known and most popular clustering algorithms used in a variety of domains. The k-means algorithm takes the input parameter, k , and partitions a set of N objects $X = \{x_1, x_2, \dots, x_N\} \subset R^S$ into k clusters. So that the resulting intra-cluster similarity is high but the inter-cluster similarity is low. The cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's center of gravity. Let $C_i \subset R^S$ be the mean vector of cluster U_i . The k-means clustering computes its centers iteratively by minimizing the objective function. Typically, the square-error criterion is used, defined as:

$$J = \sum_{i=1}^c \sum_{x_j \in U_i} \|x_j - C_i\|^2 \quad (1)$$

where J is the sum of the square-error for all objects in the data set, and the criterion tries to make the resulting c cluster as compact and as separate as possible. Compute new mean for each cluster using:

$$C_i = \frac{\sum_{x_j \in U_i} x_j}{|U_i|} \quad (2)$$

where $|U_i|$ is the number of objects in cluster U_i .

The k-means algorithm proceeds as follows. First, it randomly selects c of the objects, each of which initially represents a cluster means or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster. This process iterates until the criterion function converges.

2.2. Rough k-means Algorithm

The rough set theory proposed by Pawlak is an important tool to deal with imprecise, incomplete and inconsistent data. A rough set X is characterized by its lower and upper approximations $\underline{B}X$ and $\overline{B}X$ respectively.

For given a finite set U of objects, each subset $X \subseteq U$ and indiscernibility relation B , the lower and upper approximation set can be separately defined as:

$$\underline{B}X = \nu\{y \in U / B \mid Y \subseteq X\} \quad (3)$$

$$\overline{B}X = \nu\{y \in U / B \mid Y \cap X \neq \phi\} \quad (4)$$

In the Lingras' rough k-means algorithm, the concept of k-means is extended by viewing each cluster as an interval or rough set, with the following properties [2]:

- (1) An object x_j can be part of at most one lower approximation.
- (2) If $x_j \in \underline{B}X$ of cluster X , then simultaneously.
- (3) If x_j is not a part of any lower approximation, then it belongs to two or more upper approximations.

In Rough k-means algorithm, computation of the cluster prototypes is modified in the rough framework, by incorporating the concepts of upper and lower approximations. The centroid C_i of cluster U_i is evaluated as follows [2]:

$$C_i = \begin{cases} w_{low} \times \frac{\sum_{x_j \in \underline{B}U_i} x_j}{|\underline{B}U_i|} + w_{up} \times \frac{\sum_{x_j \in (\overline{B}U_i - \underline{B}U_i)} x_j}{|\overline{B}U_i - \underline{B}U_i|} & \text{if } (\overline{B}U_i - \underline{B}U_i) \neq \phi \\ w_{low} \times \frac{\sum_{x_j \in \underline{B}U_i} x_j}{|\underline{B}U_i|} & \text{otherwise} \end{cases} \quad (5)$$

where the parameters w_{low} and w_{up} correspond to the relative importance of the lower approximation and boundary respectively. $w_{low} + w_{up} = 1$ and usually, $w_{low} > w_{up}$. Here $|\overline{B}U_i - \underline{B}U_i|$ is the number of objects in the rough boundary lying between the two approximations. It is easy to find that the above formula is a generalization of hard k-means. Especially when $|\overline{B}U_i - \underline{B}U_i| = 0$, rough k-means will degenerates to hard k-means.

2.3. Improvements of Rough k-means Algorithm

Georg Peters improves center iteration formula of rough k-means algorithm, considering the case of empty lower approximation set and non-empty boundary. And the center iteration formula removes the weights in both cases. Modified formula is as follows [6]:

$$C_i = \begin{cases} w_{low} \times \frac{\sum_{x_j \in \underline{BU}_i} x_j}{|\underline{BU}_i|} + w_{up} \times \frac{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} x_j}{|\overline{BU}_i - \underline{BU}_i|} & \text{if } \underline{BU}_i \neq \phi \wedge (\overline{BU}_i - \underline{BU}_i) \neq \phi \\ \frac{\sum_{x_j \in \underline{BU}_i} x_j}{|\underline{BU}_i|} & \text{if } \underline{BU}_i \neq \phi \wedge (\overline{BU}_i - \underline{BU}_i) = \phi \\ \frac{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} x_j}{|\overline{BU}_i - \underline{BU}_i|} & \text{if } \underline{BU}_i = \phi \wedge (\overline{BU}_i - \underline{BU}_i) \neq \phi \end{cases} \quad (6)$$

Rough k-means algorithm gives approximate sets and boundary objects the same rough weights, while ignoring the difference between the object and the object. Membership degree is a powerful tool to distinguish objects. It can reflect the association degree of objects by calculating the distance between each object and each cluster. The higher the membership degree is, the closer the association degree is [7-9]. Membership weights are calculated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}} \quad (7)$$

Where $\mu_{ik} \in [0,1]$ and $\sum_{i=1}^c \mu_{ik} = 1$ is the membership degree of object j to cluster i . $m \in (1, \infty)$ is the fuzzifier. And d_{ik} is the Euclidean distance from x_k to center vector C_i , which defines the closeness to the center.

Rough fuzzy k-means clustering integrates fuzzy membership weights into the rough k-means clustering and treats it as a dissimilarity criterion. Its center iterative formula is as following [7, 9]:

$$C_i = \begin{cases} w_{low} \times \frac{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^m x_j}{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^m} + w_{up} \times \frac{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^m x_j}{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^m} & \text{if } \underline{BU}_i \neq \phi \wedge (\overline{BU}_i - \underline{BU}_i) \neq \phi \\ \frac{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^m x_j}{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^m} & \text{if } \underline{BU}_i \neq \phi \wedge (\overline{BU}_i - \underline{BU}_i) = \phi \\ \frac{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^m x_j}{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^m} & \text{if } \underline{BU}_i = \phi \wedge (\overline{BU}_i - \underline{BU}_i) \neq \phi \end{cases} \quad (8)$$

Fuzzy rough k-means algorithm [9, 10] make further improvements on rough fuzzy k-mean membership formula, who makes membership degree of the lower approximate objects equal 1 and considers objects certainly belong to current cluster. The center iteration formula of fuzzy rough k-means clustering is as follows:

$$C_i = \frac{\sum_{j=1}^N (\mu_{ij})^m x_j}{\sum_{j=1}^N (\mu_{ij})^m} \quad (9)$$

3. Rough k-means Based on Variable Weighted Distance Measure

3.1. Variable Weighted Distance Measure

Euclidean distance measure is the best way to present the distribution of objects in clusters. But the distance difference is very small and has a linear distribution after normalizing. It is not ideal to directly use distance with the same weight when computing new center in a cluster. This paper introduces a new weighted distance measure. The formula is as follows:

$$M_{ij} = \frac{2/\pi \times \arctan(-p \times \|x_j - C_i\|^2) + 1}{\sum_{x_j \in U_i} (2/\pi \times \arctan(-p \times \|x_j - C_i\|^2) + 1)} \quad (1)$$

$\|x_j - C_i\|$ is the Euclidean distance from object x_j to center C_i , p is the adjustable parameter, and the function $y = 2/\pi \times \arctan(-p \times x) + 1$ is to redistribute the distance measure.

The response curves of the function $y = 2/\pi \times \arctan(-p \times x) + 1$ with different parameter p are shown in Figure 1.

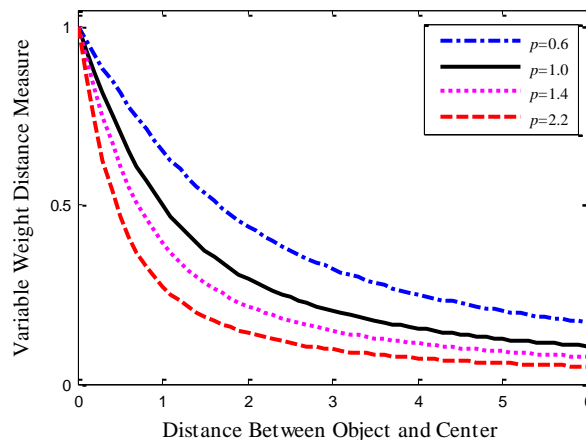


Figure 1. The Curve of Weighted Distance Measure Function

It can be seen from above figure that the smaller the distance between object and a cluster, the greater the contribution for a cluster, and the larger the parameter p , the faster the weighted value falling with the distance increasing.

3.2. Improved Algorithm of Rough k-means Clustering

Combining variable weighted distance measure with rough k-means algorithm, an improved clustering algorithm is designed in this paper. The steps of the algorithm are as follows:

- Step 1.** Initialization: N is the number of the data set, k is the number of clustering objects, $C_i (i=1, \dots, k)$ is the i -th original center, w_{low} is the weight of the lower approximation set, w_{up} is the weight of the boundary set, Δ is the similarity threshold of distance, p is the variable parameter;
- Step 2.** To object $x_j (j=1, \dots, N)$, calculate the Euclidean distance from each center, and divide x_j to the upper approximation set \bar{BU}_i of the nearest center C_i of the cluster U_i .

Step 3. If there is a center C_i' , which makes their difference of distance less than Δ between from x_j to C_i' and from x_j to C_i , then divide x_j to the upper approximation set \overline{BU}_i' of the cluster U_i' ; otherwise, divide it to the lower approximation set \underline{BU}_i of U_i .

Step 4. Calculate M_{ij} according to formula (8), and compute new mean for each cluster using:

$$C_i = \begin{cases} w_{low} \times \sum_{x_j \in \underline{BU}_i} M_{ij} \times x_j + w_{up} \times \sum_{x_j' \in (\overline{BU}_i - \underline{BU}_i)} M_{ij}' \times x_j' & \text{if } \underline{BU}_i \neq \phi \wedge (\overline{BU}_i - \underline{BU}_i) \neq \phi \\ \sum_{x_j \in \underline{BU}_i} M_{ij} \times x_j & \text{if } \underline{BU}_i \neq \phi \wedge (\overline{BU}_i - \underline{BU}_i) = \phi \\ \sum_{x_j' \in (\overline{BU}_i - \underline{BU}_i)} M_{ij}' \times x_j' & \text{if } \underline{BU}_i = \phi \wedge (\overline{BU}_i - \underline{BU}_i) \neq \phi \end{cases} \quad (2)$$

Step 5. Repeat steps 2, 3 and 4, until no new data object.

The algorithm inserts the weighted distance M_{ij} into all three cases of rough k-means clustering. Actually it is improvement for movement of each iteration step.

4. Simulation and Analysis

In order to verify that the treatment effect of the algorithm on the boundary of the clustering results, we select two UCI data sets, which already have clear classification. It will be conducive to the final results of the accuracy analysis. Some characteristics of these two data sets are as follow.

Wine: Wine is about the chemical analysis of Italian wine in the same area. This data contains 178 samples, 13 condition attributes and 1 classification attribute. All samples are divided into 3 classes.

Iris: It is one of the most commonly used UCI data sets. This data contains 150 samples, 4 condition attributes and 1 classification attribute. All samples are divided into 3 classes.

To the clustering of the Wine data and Iris data, we use a uniform initial cluster centers, the specific values of the initial centers are given in Table 1 and 2.

Table 1. The Initial Clustering Centers of Wine Data

The serial number of cluster	The serial number of condition attribute												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	12.96	2.51	2.4	19.78	104	2.14	1.62	0.39	1.52	5.66	0.89	2.4	738
2	13.81	1.89	2.44	16.92	105	2.86	3.02	0.29	1.91	5.76	1.08	3.1	1210
3	12.52	2.47	2.29	20.8	92.44	2.07	1.77	0.39	1.45	4.11	0.94	2.49	459

Table 2. The Initial Clustering Centers of Iris Data

The serial number of cluster	The serial number of condition attribute			
	1	2	3	4
1	6.74	3.04	5.61	2.04
2	5.86	2.75	4.33	1.38
3	5	3.42	1.48	0.25

The classic hard k-means algorithm, rough k-means algorithm [5], rough fuzzy k-means algorithm [7], and the improved rough k-means clustering introduced in this paper are used to analysis the Wine data.

For easy to compare, the several algorithms use the same rough weight value, i.e. the lower approximate weight $w_{low}=0.9$, the boundary weight $w_{up}=0.1$. And the accuracy refers to the percentage of correct clustering objects in the total number of objects compared with the decision of the original data set.

Table 3 shows the contrast of the accuracy with several algorithms on Wine data, where the parameter p is 1. It can be seen that, after adding weighted distance measure, the effect of rough k-means clustering has been greatly improved. Obviously, the simulation results of the Wine data show that rough k-means algorithm based on weighted distance measure provide optimal rough clustering with better boundary.

Table 3. The Compare of Precision of Different Algorithms to Wine Data

Hard k-means Algorithm					
Clutering number $K=3$		Distance threshold $dist=0.01$		Quality function evaluation threshold $e=0.1$	
Accuracy		62.18%			
Rough k-means Algorithm					
Clutering number $K=3$	Distance threshold $dist=0.01$	The lower approximation weight $w_{low}=0.9$	The boundary weight $w_{up}=0.1$	Quality function evaluation threshold $e=0.1$	
Accuracy		64.62%			
Rough Fuzzy k-means Algorithm					
Clutering number $K=3$	Fuzzy Index $m=1.5$	Membership threshold $\mu = 0.01$	The lower approximation weight $w_{low}=0.9$	The boundary weight $w_{up}=0.1$	Quality function evaluation threshold $e=0.1$
Accuracy		72.11%			
Rough k-means Algorithm Based on weighted distance measure ($p=1$)					
Clutering number $K=3$	Distance threshold $dist=0.01$	The lower approximation weight $w_{low}=0.9$	The boundary weight $w_{up}=0.1$	Quality function evaluation threshold $e=0.1$	
Accuracy		72.37%			

In order to directly observe and analyze the results of above algorithms, the high-dimension data is mapped into the two-dimensional space with PCA (Principal Component Analysis) method.

Figure 2-5 are the distribution diagram of the Wine dataset simulation results after dimensionality reduction, where the cluster 1, 2 and 3 are denoted by green points, red circles and blue stars respectively, and the mean of each cluster is denoted by black cross. Bright blue boxes represent the clustering objects in error. Figure 2 is distributed in accordance with the decision attributes of the original Wine data. It can be seen that all clustering boundaries are intertwined, confused, especially for cluster 2 and 3.

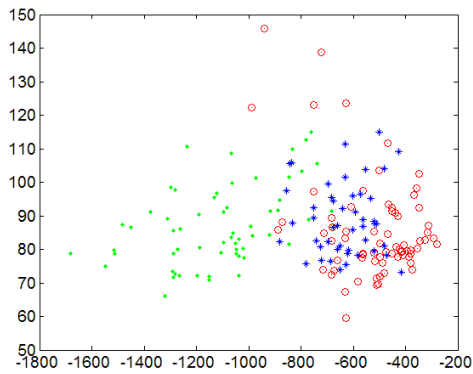


Figure 2. The Distribution of Wine Data

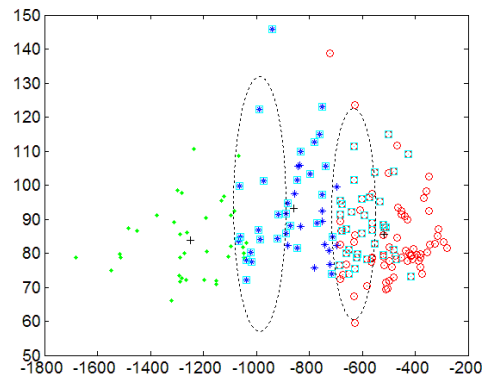


Figure 3. The Clustering Result of Wine Data with RKM

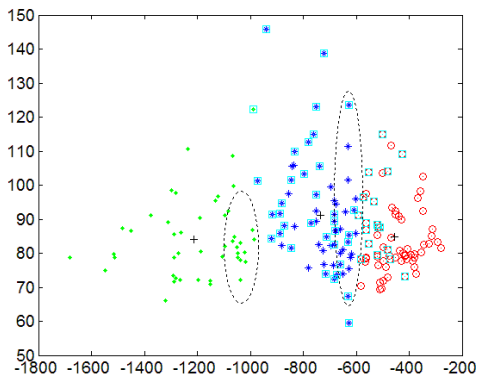


Figure 4. The Clustering Result of Wine Data with RFKM

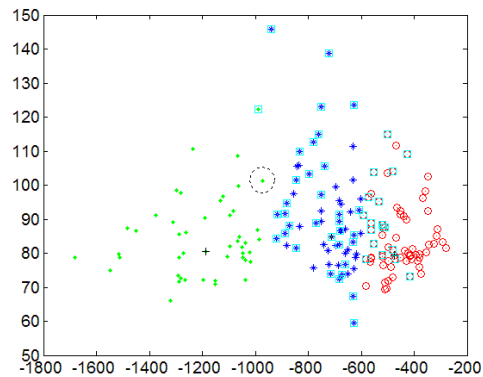


Figure 5. The Clustering Result of Wine Data with the Improved Algorithm

The simulation results show that all points affecting the accuracy of clustering concentrate on the boundary between the cluster 2 and cluster 3. Figure 3 is the result of rough k-means clustering. Black dotted circle reflects some points should belong to cluster 3 but were classified to cluster 2. And cluster 3 contains a large number of the points that do not belong to the cluster, including from cluster 1 and cluster 2. Obviously, unified clustering weight of the clustering object is useless for separating cluster 3. The simulation results indicate the number of correct clustering points of the whole cluster 3 only reached about 30%.

It takes some improvement with weighted distance measure shown in Figure 5. Contrast to Figure 3, the points ringed out of Figure 4 is reclassified from cluster 3 back to the cluster 1 and cluster 2. There are also some parts of cluster 2 which are draw back to cluster 3. The accuracy of cluster 3 increased about 17%, and cluster 2 also increased by nearly 9%. It shows that the difference between the clusters can effectively make the boundary separated.

In order to illustrate the influence of variable parameter p on the clustering results with the improved algorithm, different values of parameter p are used in the simulation with Wine data and Iris data. Table 4 records the clustering results with different values of parameter p .

Table 4. The Computing Time and Precision with Variable Parameter p

The value of parameter p	Convergence steps		Computing time		Precision	
	Wine data	Iris data	Wine data	Iris data	Wine data	Iris data
0.2	82	101	1.406	1.329	62.18%	75.33%
0.6	38	101	0.681	1.339	72.39%	69.33%
1	32	17	0.562	0.292	72.39%	90.67%
1.4	31	101	0.553	1.345	72.39%	76%
1.8	30	101	0.536	1.339	72.39%	79.33%
2.2	30	101	0.534	1.332	72.39%	85.33%
2.6	30	101	0.534	1.299	72.39%	79.33%
3	30	16	0.525	0.255	72.39%	90.67%
5	29	101	0.502	1.314	72.39%	79.33%

The computing time and precision are also selected especially to evaluate the performance of the improved algorithm with different values of parameter p . The test results are shown in Figure 6 and 7.

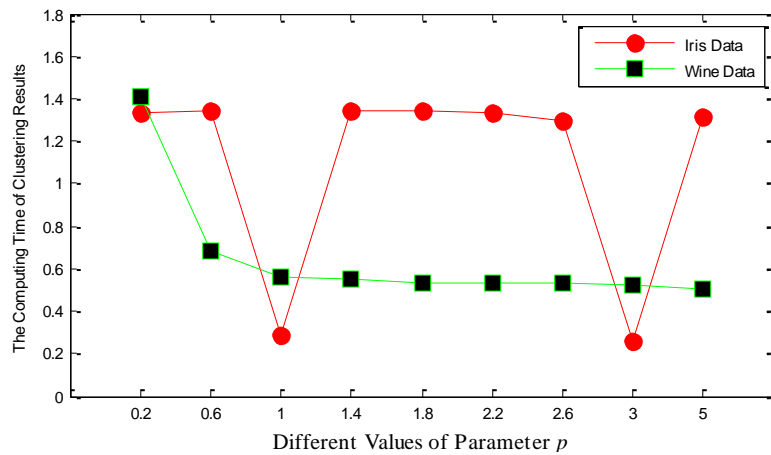


Figure 6. The Computing Time with Different Values of Parameter p

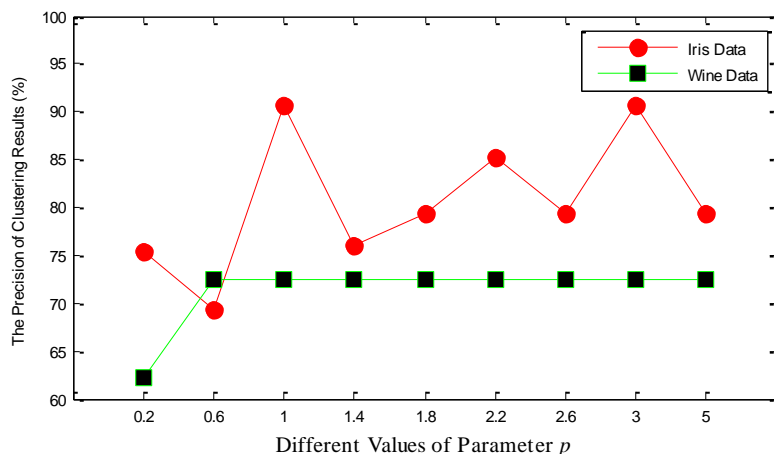


Figure 7. The Precision with Different Values of Parameter p

It can be seen obviously from Table 4 and above Figures that, for Wine data, the larger the parameter p , the faster the clustering convergence with almost the same accuracy. When $p > 1.8$, the convergence steps, the computing time and precision have little change with the increase of the value of parameter p . For Iris data, the monotone decreasing trend of convergence is not presented, but some suitable values of parameter p are found for achieving better clustering result.

5. Conclusion

It is not reasonable for rough k-means clustering algorithm that the same weight was used for all the data objects in a lower or upper approximate set when computing new mean for each cluster. To solve this problem, a variable weighted distance measure method is introduced and an improved algorithm of rough k-means clustering is proposed in this article. Each object, according to the distance from the center, is given different weighted coefficient. The closer the object is to the center, the higher weighted coefficient is, and the higher the contribution is in the center iteration. The influence of variable parameter on clustering results is also analyzed. The simulation results show that the improved algorithm is valid and the suitable value of variable parameter is available.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (61105082), Natural Science Foundation of Jiangsu Province (BK2012470), Qing Lan Project of Jiangsu Province (QL2012), and Jiangsu Government Scholarship for Overseas Studies (JS-2013-219).

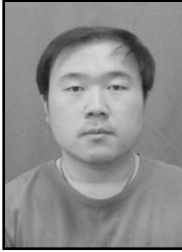
References

- [1] H. J. Wei and K. Micheline, "Data Mining", Concepts and Techniques (Third Edition), Morgan Kaufmann Publishers, (2011); San Francisco, California.
- [2] P. Lingras, "Rough set clustering for Web mining", IEEE International Conference on Fuzzy Systems, vol. 2, (2002).
- [3] P. Lingras and C. West, "Interval set clustering of Web users with rough k-means", Journal of Intelligent Information Systems, vol. 1, (2004).
- [4] P. Lingras, R. Yan and C. West, "Comparison of Conventional and Rough K-Means Clustering", Lecture Notes in Artificial Intelligence, no. 2639, (2003).
- [5] G. Peters, "Outliers in Rough k-Means Clustering", Lecture Notes in Computer Science, (2005).
- [6] G. Peters, "Some refinements of rough k-means clustering", Pattern Recognition, vol. 39, (2006).
- [7] Q. H. Hu and D. Yu, "An Improved Clustering Algorithm for Information Granulation", Lecture Notes in Artificial Intelligence, no. 3613, (2005).
- [8] S. Mitra, H. Banka and W. Pedrycz, "Rough Fuzzy Collaborative Clustering", IEEE Transactions on Systems, Man and Cybernetics—Part B, Cybernetics, vol.4, (2006).
- [9] S. Mitra and H. Banka, "Application of Rough Sets in Pattern Recognition", Transactions on Rough Sets VII, Lecture Notes in Computer Science, no. 4400, (2007).
- [10] G. Peters, F. Crespo, P. Lingras and R. Weber, "Soft clustering - fuzzy and rough approaches and their extensions and derivatives", International Journal of Approximate Reasoning, vol. 54, (2013).
- [11] S. Revathy and Dr. B. Parvathavarthini, "Integrating Rough Clustering with Fuzzy Sets", Second International Conference on Sustainable Energy and Intelligent System, (2011); Chennai, India.
- [12] Q. L. Liu, M. Deng, Y. Shi and J. Q. Wang, "A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity", Computers & Geosciences, vol. 46, (2012).
- [13] A. Azadeh, M. Saberi, M. Anvari, A. Azaron and M. Mohammadi, "An adaptive network based fuzzy inference system-genetic algorithm clustering ensemble algorithm for performance assessment and improvement of conventional power plants", Expert Systems with Applications, vol. 38, (2011).
- [14] L. Bing, X. S. Xiong, Z. Yong and H. X. Dong, "A Sample-Weighted Possibilistic Fuzzy Clustering Algorithm", ACTA Electronica Sinica, vol. 2, (2012).
- [15] P. Lingras and G. Peters, "Rough clustering", Data Mining and Knowledge Discovery, vol. 1, (2011), pp. 64-72, 341-356.
- [16] P. Kaur, A. K. Soni and A. Gosain, "Robust kernelized approach to clustering by incorporating new distance measure", Engineering Applications of Artificial Intelligence, vol. 26, (2013).

Authors



Tengfei Zhang received his bachelor degree from Henan University of China in 2002. And he respectively received master and Ph.D degree from Shanghai Maritime University of China in 2004 and 2007. He is currently a associate professor and supervisor for master student at Nanjing University of Posts and Telecommunications. His research interests include intelligent information processing, pattern recognition and intelligent control.



Long Chen received his Master degree in Pattern Recognition and Intelligent System from Nanjing University of Posts and Telecommunications of China in 2013. His research interests include data mining, rough set theory, pattern recognition and intelligent computation.



Fumin Ma received her bachelor degree from Henan University of China in 2002, Master from Graduate University of Chinese Academy of Sciences in 2005 and the Ph.D degree from Tongji University of China in 2008. Her research interests include System Engineering, Intelligent Manufacturing System and Intelligent Information Processing.

