A Feature Weighted Spectral Clustering Algorithm Based on Knowledge Entropy

Hongjie Jia[1], Shifei Ding[1,2], Hong Zhu[1], Fulin Wu[1], Lina Bao[1]
1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China
Email: jiahongjie@cumt.edu.cn
2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190 China
Email: dingsf@cumt.edu.cn

***Abstract*—Spectral clustering has aroused extensive attention in recent years. It performs well for the data with arbitrary shape and can converge to global optimum. But traditional spectral clustering algorithms set the importance of all attributes to 1 as default, when measuring the similarity of data points. In fact, each attribute contains different information and their contributions to the clustering are also different. In order to make full use of the information contained in each attribute and weaken the interference of noise data or redundant attributes, this paper proposes a feature weighted spectral clustering algorithm based on knowledge entropy (FWKE-SC). This algorithm uses the concept of knowledge entropy in rough set to evaluate the importance of each attribute, which can be used as the attribute weights, and then applies spectral clustering method to cluster the data points. Experiments show that FWKE-SC algorithm deals with high-dimensional data very well and has better robustness and generalization ability.**

***Index Terms*—spectral clustering, rough set, knowledge entropy, attribute importance**

I. INTRODUCTION

Clustering is an important research field in multivariate statistical analysis and pattern recognition. The purpose of clustering is to classify a dataset into several different clusters, so that we can identify the structure of the dataset, and find the internal relations of the data points [1]. Traditional clustering methods, such as k-means algorithm [2,3] and FCM algorithm [4,5], can get good clustering results for the data set of convex structure, but when the sample space is non-convex, these algorithms are easy to fall into local optimum [6].

Due to the good clustering performance and the feature of easy to implement, spectral clustering has aroused more and more attention of academia in recent years [7]. Spectral clustering doesn't make any assumptions on the global structure of the data. It can converge to global optimum and performs well for the sample space of arbitrary shape, especially suitable for non-convex dataset [8]. The idea of spectral clustering is based on spectral graph theory. It uses the data points of the given dataset to construct a weighted graph, in which every edge is weighted by the similarity value between its two vertices. In this way, the data clustering problem is transformed into a graph partitioning problem. There are a variety of traditional graph cut methods, such as Minimum cut [9], Ratio cut [10], Normalized cut [11] and MinMax cut [12]. The optimal clustering results can be obtained by minimizing the objective function of the graph cut methods. However, the eigen-decomposition of Laplacian matrix is usually required during the calculation, in order to gain the global optimal solution of the objective function in the relaxed continuous domain. At present, spectral clustering has been successfully applied to speech separation [13], video indexing [14], character recognition [15] and image processing [16,17], etc. Spectral clustering algorithms provide a new idea to solve the problem of clustering and can effectively deal with many practical problems, so their research has great scientific value and application potential.

However, in spectral clustering algorithms, it is generally assumed that the importance of all data attributes is the same (default as 1) when measuring the similarity of data points. In fact, each attribute contains different information and their contributions to the clustering are also different. Moreover real data sets often contain noise and irrelevant features, which are likely to cause "dimension trap" and interfere with the clustering process [18]. For example: assume a data set has 20 dimensions, in which only 2 attributes are most relevant to clustering, but they may have the farthest distance in the entire attribute space. In this case, it is easily lead to clustering errors, if do not distinguish the attribute roles when calculating the similarity. One of the effective ways to overcome this problem is to add a weight parameter for each attribute, and make different attributes play different roles in the clustering. In Euclidean space, it is elongate the axis corresponding to relevant attributes and shorten the axis corresponding to unrelated attributes [19].

Because of the above reasons, researchers develop a variety of feature weighting methods [20-22]. These methods can be broadly divided into two categories: one is subjective weighting method, in which the weights are

Manuscript received September 1, 2012; revised October 1, 2012; accepted November 1, 2012.
Corresponding author: Shifei Ding, Email: dingsf@cumt.edu.cn

doi:10.4304/jsw.8.5.1101-1108

obtained by the subjective judgment of experts according to their experience; the other is objective weighting method, which automatically generate the weight value based on the characteristics of the data itself. In order to make full use of the information contained in each attribute and eliminate the impact of redundant attributes on the clustering results, this paper proposes a feature weighted spectral clustering algorithm based on knowledge entropy (FWKE-SC).

II. Spectral Clustering

Given a dataset which contains n data points $\chi = \{x_1, x_2, \cdots, x_n\}$, $x_i \in R^l$. Construct an undirected weighted graph $G = (V, E, W)$. Treat each data point as a vertex V in graph G. Each edge E between vertices (x_i, x_j) has a similarity value W. Then the clustering problem can be transformed into a graph partitioning problem on graph G. The optimal partitioning criteria based on graph theory is maximizing the internal similarity of the two divided sub-graphs, and minimizing the similarity between sub-graphs. Most spectral clustering algorithms search clusters utilizing the eigenvectors of similarity matrix [23]. The similarity matrix $W \in R^{n\times n}$ of spectral clustering algorithm is composed of w_{ij}, which is usually represented by Gaussian kernel function:

$$w_{ij} = \exp\left(-\frac{d^2(x_i, x_j)}{2\sigma^2}\right) \tag{1}$$

where $d(x_i, x_j)$ is the Euclidean distance between point x_i and x_j; σ is the scaling parameter which controls how rapidly the similarity w_{ij} falls off with $d(x_i, x_j)$.

In graph G, the sum of the weights of the edges connected to vertex i, is defined as the degree of vertex i, which can be represented by d_i:

$$d_i = \sum_{j=1}^{n} w_{ij} \tag{2}$$

So the degree matrix $D \in R^{n\times n}$ of graph G is a diagonal matrix: the diagonal element is d_i, while the elements outside the diagonal are 0.

In graph cut methods, seeking the optimal solution of the objective function is often NP-hard. With the help of spectral method, the original problem can be solved in polynomial time by relaxing the original discrete optimization problem to the real domain [24]. For graph partitioning, a point can be considered part belonging to subset A and part belonging to subset B, rather than strictly belongs to one cluster. It can be proved that the classification information of vertices is contained in the eigenvalues and eigenvectors of graph Laplacian matrix. And we can get good clustering results, if we make full use of the classification information during the clustering process. This kind of clustering method is essentially based on matrix spectral theory, so it is called spectral clustering. The result of spectral clustering is an approximate optimal solution, which is the relaxation solution of graph cut objective function.

Spectral clustering has many specific implement methods, and these methods are dependent on different graph partitioning criterion and spectral graph theory. According to the *2-way* and *k-way* cut criteria, spectral clustering algorithms can be divided into two categories: iterative spectral methods and multi-way spectral methods. The main steps of spectral clustering algorithms are as follows [25]:

Step 1 Create similarity matrix and Laplacian matrix to describe the samples;

Step 2 Calculate the eigenvalues of the Laplacian matrix, and choose appropriate eigenvalues and their corresponding eigenvectors to create the space R^k;

1) *2-way*: the initial samples are distributed into one-dimensional space (k=1).

2) *k-way*: the initial samples are distributed into space R^k and we can obtain matrix Y which consists of k orthogonal vectors.

Step 3 Treat Y as a new representation space of data samples, and then cluster the data samples according to the new representation space.

1) *2-way*: optimize the objective function in the one-dimensional space, and then repeatedly partition the obtained sub-graphs.

2) *k-way*: use a typical clustering algorithm such as k-means to deal with the eigenvectors in R^k.

These steps form a unified basic framework of spectral clustering algorithms. At present, many spectral clustering algorithms have been proposed and some of them perform very well in specific clustering applications. Their main differences lie in: how to construct the similarity matrix to describe the original dataset and how to select a proper graph cut method. Celikyilmaz [26] uses fuzzy kappa-nearest neighbor approach to identify clusters and develops a new soft-link spectral clustering algorithm. Zhang [27] constructs the similarity matrix based on a random walk approach instead of simply using Gaussian kernel function. Liu [28] introduces Mercer kernel method to spectral clustering, in order to make full use of the local neighborhood information in kernel space. Wang [29] proposes spectral multi-manifold clustering algorithm, which can deal with intersections between clusters very well. For high-dimensional data, Nie [30] develops a spectral embedded clustering framework, which adds a linear regularization into the objective function to reduce the computational complexity of the algorithm. Zhao [31] develops a fuzzy similarity measure for spectral clustering (FSSC), utilizing the prototypes and partition matrix obtained by fuzzy c-means clustering algorithm. Chen [32] proposes parallel spectral clustering, which can deal with large data sets on distributed computers.

III. Attribute Importance Based On Knowledge Entropy

A. Description of Knowledge

Rough set theory is a new mathematical tool to deal with fuzzy and uncertain knowledge. It is based on the boundary line region thought of G. Frege, and was first proposed by Polish mathematician Z. Pawlak in 1982 [33]. In rough set theory, knowledge is an ability to classify objects. The "object" can be anything around, such as kind, state, concept, process, time, etc. Therefore, knowledge must be linked with a specific semantic, which described in mathematical language is domain. Domain is a non-empty finite set, which does not have any subjective assumptions on the characteristic of knowledge. In fact, domain not only provides explicit knowledge about reality, but also has the ability to deduce tacit knowledge from the explicit knowledge.

Definition 1 [34] Knowledge: Suppose U is a non-empty finite set of objects, and then U is called a domain. Any subset X of domain U ($X \subseteq U$), is called a concept or scope of U. The sub-cluster in domain U is called the abstract knowledge about U, or knowledge for short.

Rough set theory focuses on the knowledge that can form partition and overlap in domain U. Classification and knowledge is usually represented with the equivalence relation.

Definition 2 [34] Knowledge base: Given a domain U and a cluster of equivalence relations S of domain U, then tuple *K=(U,S)* is called a knowledge base or approximation space about domain U.

The equivalence relation in domain represents the partition and knowledge. Knowledge base contains a wide range of knowledge which is exported by equivalence relations. The knowledge reflects the classification ability of equivalence relations to the domain. And the various relationships among the knowledge are also implied in the knowledge base.

Definition 3 [34] Indiscernibility relation: Given a domain U and a cluster of equivalence relations S of domain U, if $P \subseteq S$ and $P \neq \Phi$, then the intersection of all equivalence relations in P ($\bigcap P$) is still an equivalence relation of domain U and is called the indiscernibility relation on P, denoted by *IND*(P), which is often abbreviated as P in the case of no confusion.

$$\forall x \in U, [x]_{IND(P)} = [x]_P = \bigcap_{\forall R \in P} [x]_R \quad (3)$$

where $[x]_R$ ($R \in P$) is the set of all the objects that meet the indiscernibility relation with object x, and is called the equivalence class of object x determined by the equivalence relation R.

The collection of all equivalence classes exported by *IND*(P) is denoted by *U/IND(P)*, which constitutes a classification of U and is called P-basic knowledge of domain U. *U/IND(P)* can be abbreviated as *U/P*.

B. Measure Knowledge Entropy

Rough set theory defines knowledge as the classification mode of domain and the roughness of knowledge is described by equivalence relations and inclusion relations in algebra. Specifically, domain U is divided into several equivalence classes by knowledge. The more equivalence classes, the less the samples in each class, and the smaller the knowledge roughness; conversely the less equivalence classes, the more the samples in each class, and the greater the knowledge roughness. Miao DuoQian et al. [35] study rough set from the view of information theory and establish the relationship between knowledge roughness and information entropy. They treat the knowledge in rough set as random variables. The smaller the information entropy of knowledge, the greater the amount of information contained in knowledge, and the smaller the uncertainty or randomness of knowledge.

Definition 4 [36] Knowledge representation system: The tuple *KRS=(U,A,V,f)* can be called a knowledge representation system, where U is a non-empty finite set of objects, called domain; A is a non-empty finite set of attributes, containing condition attribute C and decision attribute D, $A = (C \cup D)$, $C \cap D = \Phi$; V is the range of all attributes, $V = \bigcup_{a \in A} V_a$, V_a is the range of attribute $a \in A$; f represents a mapping $U \times A \to V$, called information function.

Knowledge representation systems can be divided into two categories: one is information system (information table), in which decision attribute set $D = \Phi$, namely the knowledge representation system that does not contain decision attributes; the other is decision system (decision table), in which decision attribute set $D \neq \Phi$, namely the knowledge representation system that contains decision attributes.

Definition 5 [37] Information entropy: Assume *KRS=(U,A,V,f)* is an information system, and $P \subseteq A$ is the set of equivalence relations in U. The partition exported from U by P is *U/P=U/IND(P)* $= \{X_1, X_2, \cdots, X_n\}$, then the information entropy of knowledge *U/P* is defined as

$$H(U / P) = -\sum_{i=1}^{n} p(X_i) \log p(X_i) \quad (4)$$

where $p(X_i) = \frac{|X_i|}{|U|}$ represents the probability of equivalence class *Xi* in U.

Theorem 1 [38] Assume *KRS=(U,A,V,f)* is an information system, and equivalence relation set $P, Q \subseteq A$. If $IND(Q) \subseteq IND(P)$, then $H(U / P) \leq H(U / Q)$.

Theorem 1 shows that the more partitions the domain is divided into, the greater the information entropy of knowledge. The range of the information entropy obtained by this method is (1, +∞), which is not suitable for use as attribute weight. Therefore, scholars proposed another method to measure knowledge entropy (see Definition 6).

Definition 6 [39] Knowledge entropy: Assume $Z = \{X_1, X_2, \cdots, X_n\}$ is a partition of domain U, then the knowledge entropy of Z is defined as

$$H(Z) = \sum_{i=1}^{n} p(X_i) \times p(X_i^c) \tag{5}$$

where $p(X_i^c) = 1 - p(X_i)$.

$H(Z)$ describes the amount of information contained in knowledge Z and it has the following properties:

(1) $0 \le H(Z) \le 1 - \frac{1}{|Z|}$;

(2) $H(Z) = 0$, if and only if $Z = \{U\}$;

(3) $H(Z) = 1 - \frac{1}{|Z|}$, if and only if $|X_i| = |X_j|$, $(i, j = 1,2,\cdots,n)$.

The above method of measuring knowledge entropy is easy to understand, and the calculation is simple. From property (1) we can see that if $|Z| = n$, then $0 \le H(Z) \le 1 - \frac{1}{n}$, which is more in line with the requirements of weight values; property (2) indicates that if there is no partition, uncertainty will not exist, so the knowledge entropy is 0; property (3) shows that when domain is divided evenly, knowledge entropy reaches the maximum. In this case, uncertainty is the greatest and it is difficult for people to make a choice, so more knowledge is needed to eliminate the uncertainty.

C. Calculate Attribute Importance

Using knowledge entropy to measure the importance of sample attributes can objectively reflect the partition ability of the attributes to sample data, without the need for any prior information [40]. Here are the concrete steps of calculating attribute importance based on knowledge entropy.

(1) Data preprocessing. Rough set method can only handle discrete data. If the data is continuous, it should be discretized first. Namely, selecting appropriate points to divide the range of continuous attribute values into a number of discrete intervals, and then use different integers to represent the attribute values of each subinterval.

(2) Calculate the knowledge entropy of attributes. Based on the discretization of attribute values, determine the partition of sample data corresponding to the current attribute. The partition corresponding to attribute a_j is

$$\begin{aligned} U / IND(a_j) &= \{ \bigcup_{a_j(x)=v_k} x \mid x \in U, v_k \in V \} \\ &= (X_1, X_2, \cdots, X_n) \end{aligned} \tag{6}$$

Calculate its knowledge entropy according to Definition 6:

$$\begin{aligned} H(a_j) &= \sum_{i=1}^{n} p(X_i) \times p(X_i^c) \\ &= \sum_{i=1}^{n} p(X_i) \times [1 - p(X_i)] \\ &= 1 - \sum_{i=1}^{n} [p(X_i)]^2 \end{aligned} \tag{7}$$

(3) Determine the importance of attributes. Normalize the knowledge entropy of each attribute, and treat the proportion of knowledge entropy as attribute importance. The formula is as follows:

$$w_j = H(a_j) / \sum_{j=1}^{n} H(a_j) \tag{8}$$

IV. Feature Weighted Spectral Clustering Algorithm Based On Knowledge Entropy

Real world datasets may contain redundant and irrelevant attributes. In view of the fact that different attributes make different contributions to the clustering, this paper proposes a feature weighted spectral clustering algorithm based on knowledge entropy (FWKE-SC). The basic idea of this algorithm is: first calculate the importance of each attribute according to the concept of knowledge entropy in rough set; then weight the attributes with the corresponding attribute importance; next measuring the similarities between data points to construct similarity matrix and Laplacian matrix; select the eigenvectors corresponding to the largest k eigenvalues of Laplacian matrix and map the points of original dataset to a R^k space; finally, using k-means or other traditional clustering algorithm to cluster the data points in R^k space. The detailed steps of FWKE-SC algorithm are as follows.

Input: Dataset $\chi = \{x_1, x_2, \cdots, x_n\}$ $(x_i \in R^l)$, Class number k.

Output: k clusters (independent connected components).

Step 1 Preprocess dataset χ. According to specific problems of the study, transform the attribute values of dataset into the data format suitable for rough set approach, such as discretizing continuous attribute values.

Step 2 Calculate the importance of each attribute. Using formula (7) to calculate the knowledge entropy of each attribute and then confirm the importance of each attribute according to formula (8).

Step 3 Weight the attributes with the corresponding attribute importance and obtain the feature weighted dataset χ'.

Step 4 Based on χ', using formula (1) to construct similarity matrix $W \in R^{n \times n}$ and using formula (2) to construct degree matrix $D \in R^{n \times n}$.

Step 5 Form the Laplacian matrix L_{sym}: $L_{sym} = D^{-1/2}(D - W)D^{-1/2}$, according to the similarity matrix W and degree matrix D.

Step 6 Calculate the eigenvectors $u_1, \cdots, u_k$ corresponding to the largest k eigenvalues of matrix L_{sym}, and form the matrix U by stacking these eigenvectors in columns: $U = [u_1 \vdots \cdots \vdots u_k] \in R^{n \times k}$.

Step 7 Form the matrix Y by normalizing each row of matrix U to have unit length: $y_{ij} = u_{ij} / \left[\sum_{j=1}^{n} u_{ij}^2\right]^{1/2}$.

Step 8 Treat each row of matrix Y as a point in space R^k and classify them into k clusters via k-means or any other algorithm.

Step 9 Assign the original point x_i to cluster j, if and only if row i of matrix Y is assigned to cluster j.

The implementation process of FWKE-SC algorithm is described in detail by the above nine steps. Figure 1 shows the flowchart of FWKE-SC algorithm, so that people can make a more intuitive understanding of this algorithm.

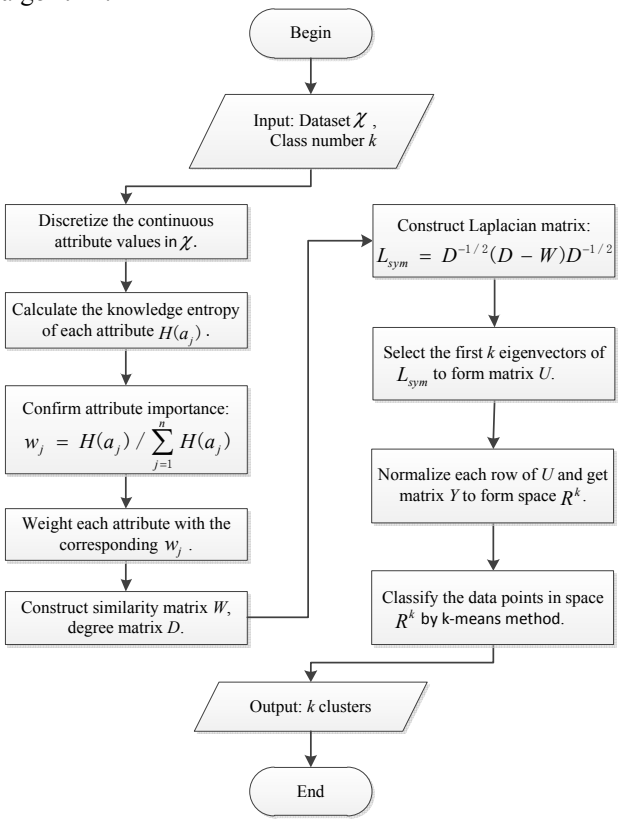


Figure 1. The flowchart of FWKE-SC algorithm.

V. EXPERIMENTAL EVALUATION

A. Clustering Evaluation Criteria

In order to verify the effectiveness of FWKE-SC algorithm, we select four real world datasets from UCI machine learning database and their characteristics are shown in Table I.

TABLE I.
THE CHARACTERISTICS OF THE DATASETS

Dataset	Data Characteristic		
	Sample number	Attribute number	Class number
Dermatology	366	34	6
Hepatitis	155	19	2
Iris	150	4	3
Soybean (Small)	47	35	4

Evaluating the merits of clustering results needs appropriate clustering evaluation criteria. F-measure is an effective clustering evaluation method coming from the field of information retrieval. It uses precision and recall to get a comprehensive evaluation index F-score, which objectively reflects the quality of the generated clusters [41]. Assume that the original dataset consists of k classes, in which class i corresponds to the class i^* of clustering results. Then the precision P, recall R and F-score of class i are defined as:

$$P(i) = N_{ii^*} / N_{i^*} \tag{9}$$

$$R(i) = N_{ii^*} / N_i \tag{10}$$

$$F(i) = \frac{2 \times P(i) \times R(i)}{P(i) + R(i)} \tag{11}$$

where N_{ii^*} is the number of elements in the intersection of class i and class i^*, N_i is the number of elements in class i, N_{i^*} is the number of elements in class i^*.

For the entire clustering results, the overall index F of F-measure is the weighted average F-score of each class:

$$F = \frac{1}{n} \sum_{i=1}^{k} [N_i \times F(i)] \tag{12}$$

where n is the total number of elements in dataset, k is the number of clusters, N_i is the number of elements in class i. $F \in [0,1]$, the larger the F value, means the better the clustering.

B. Experimental Results and Analysis

In the experiment, we compare FWKE-SC algorithm with NJW algorithm [42] and k-means algorithm [43]. NJW algorithm is a typical spectral clustering algorithm. It conducts the k-way partitioning using multiple eigenvectors, which is easy to implement and has low time complexity. k-means algorithm is a partition-based clustering algorithm. It requires setting cluster number or cluster centers in advance, and then iteratively calculates the distances between data points to get the optimal clustering partition. The comparison results of the three algorithms on each dataset are shown from Figure 2 to Figure 5. In these figures, the abscissa represents the class labels of clustering results; the ordinate represents the F-score of each class.

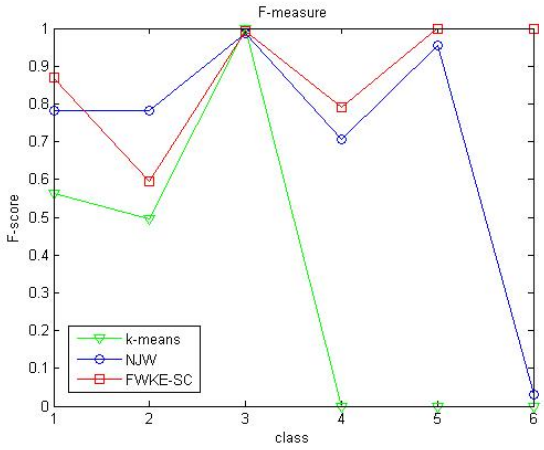


Figure 2. The experimental results on dataset Dermatology.

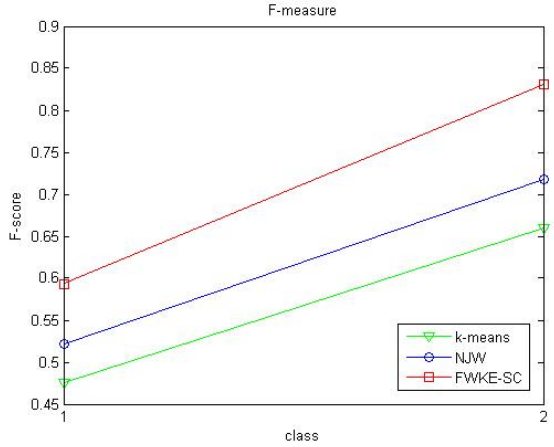


Figure 3. The experimental results on dataset Hepatitis.

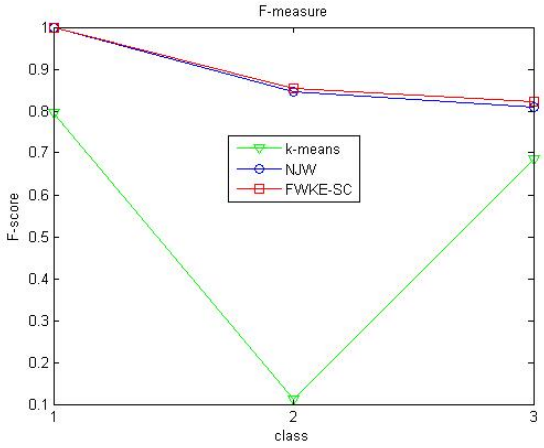


Figure 4. The experimental results on dataset Iris.

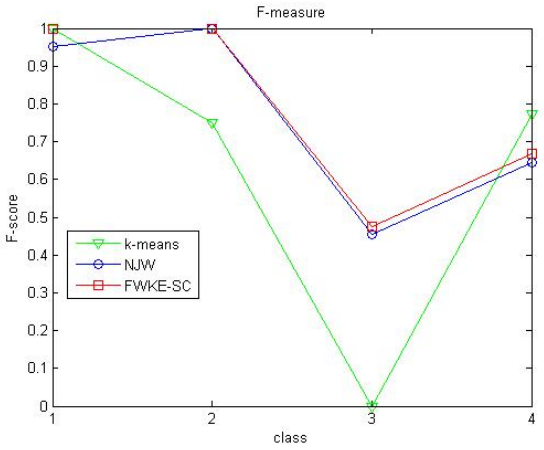


Figure 5. The experimental results on dataset Soybean(Small).

In the above experiment figures, k-means algorithm has the lowest indicators, because this algorithm is very sensitive to the initialization of cluster centers and easy to fall into local minimum, resulting in poor clustering results. FWKE-SC algorithm and NJW algorithm are all spectral clustering algorithms. By constructing Laplacian matrix and spectral decomposition, the objective function can converge to global optimum and obtain more balanced clusters. However, the curve of NJW algorithm is slightly lower than FWKE-SC algorithm, because NJW algorithm treats all the attributes of dataset equally without discrimination. In fact, each attribute contains different information and their contributions to the clustering are also different. FWKE-SC algorithm takes into account the importance of attributes. The feature weighted method fully taps the information contained in each attribute, which can help the algorithm to generate better clustering results. The overall index *F* of FWKE-SC algorithm, NJW algorithm and k-means algorithm on each dataset are given below, as shown in Table II.

TABLE II.
THE OVERALL INDEX *F* OF THE THREE ALGORITHMS ON EACH DATASET

Algorithm	Index *F*			
	Dermatology	Hepatitis	Iris	Soybean (Small)
FWKE-SC	0.8626	0.7820	0.8923	0.7680
NJW	0.7952	0.6774	0.8853	0.7455
k-means	0.4520	0.6216	0.5309	0.6518

From the above experimental results, we can see that FWKE-SC algorithm performs better than NJW algorithm and k-means algorithm on real world datasets. For different datasets, FWKE-SC algorithm can improve the clustering performance to some extent and raise the accuracy of clustering results. This algorithm uses the concept of knowledge entropy in rough set to determine the importance of each attribute. Attribute importance strengthen the role of important attributes and weaken the negative impact of noise data and redundant attributes on the clustering. So when measuring the similarities between data points, the distribution of sample data can

be better described, which can improve the performance of spectral clustering and enhance the robustness and generalization ability of the algorithm.

VI. Conclusion

Traditional spectral clustering algorithms ignore attribute importance when measuring the similarities between data points and are susceptible to the interference of noise data and irrelevant attributes. In order to solve these problems, this paper proposes a feature weighted spectral clustering algorithm based on knowledge entropy (FWKE-SC). This algorithm adds a weight parameter for each attribute, so that different attributes are able to play different roles in the clustering. The weight parameters are obtained through knowledge entropy in rough set theory. Using knowledge entropy to measure attribute importance can objectively reflect the partition ability of each attribute on the sample data, without the need for any prior information. Experiments show that FWKE-SC algorithm can better handle high-dimensional data and has strong robustness and generalization ability. However, in FWKE-SC algorithm, the number of clusters needs to be given in advance. How to automatically determine the class number of dataset and improve the self-adaptability of algorithm, still need further research.

Acknowledgment

This work is supported by "the China University of Mining and Technology Special Funds for the Graduate Practice and Research Innovation" (No.GSF122112).

References

[1] J. G. Sun, J. Liu, L. Y. Zhao, "Clustering Algorithms Research," *Journal of Software*, Vol. 19, No. 1, pp. 48-61, 2008.

[2] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281-297, 1967.

[3] J. Y. Xie, S. Jiang, W. X. Xie, et al, "An Efficient Global K-means Clustering Algorithm," *Journal of Computers*, Vol. 6, No. 2, pp. 271-279, 2011.

[4] S. D. Li, X. Q. Lv, C. L. Zhan, et al, "Regional Division of police patrols Based on Adaptive FCM Clustering and MMAS," *Journal of Computers*, Vol. 6, No. 2, pp. 313-320, 2011.

[5] T. Y. Li, Y. Chen, J. S. Zhang, "Logistics Service Provider Segmentation Based on Improved FCM Clustering for Mixed Data," *Journal of Computers*, Vol. 7, No. 11, pp. 2629-2633, 2012.

[6] Z. Huang, "A fast clustering algorithm to cluster very large categorical data sets in data mining," *Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, pp. 146-151, 1997.

[7] M.C.V. Nascimento, A.C.P.L.F. de Carvalho, "Spectral methods for graph clustering - A survey," *European Journal of Operational Research*, Vol. 211, No. 2, pp. 221-231, 2011.

[8] X. Y. Cai, G. Z. Dai, L. B. Yang, "Survey on Spectral Clustering Algorithms," *Computer Science*, Vol. 35, No. 7, pp. 14-18, 2008.

[9] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, Vol. 23, No. 2, pp. 298-305, 1973.

[10] L. Hagen, A. B. Kahng, "New spectral methods for radio cut partitioning and clustering," *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, Vol. 11, No. 9, pp. 1074-1085, 1992.

[11] J. Shi, J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 888-905, 2000.

[12] C. H. Q. Ding, X. He, H. Zha, et al, "A min-max cut algorithm for graph partitioning and data clustering," *Proceedings of IEEE International Conference on Data Mining*, pp. 107-114, 2001.

[13] F. R. Bach, M. I. Jordan, "Learning spectral clustering, with application to speech separation," *Journal of Machine Learning Research*, Vol. 7, pp. 1963-2001, 2006.

[14] V. T. Chasanis, A. C. Likas, N. P. Galatsanos, "Scene Detection in Videos Using Shot Clustering and Sequence Alignment," *IEEE Transactions on Multimedia*, Vol. 11, No. 1, pp. 89-100, 2009.

[15] S. Zeng, N. Sang, X. J. Tong, "Hand-written Numeral Recognition Based on Spectrum Clustering," *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 8004, 2011.

[16] H. Q. Liu, L. C. Jiao, F. Zhao, "Non-local spatial spectral clustering for image segmentation," *Neurocomputing*, Vol. 74, No. 1-3, pp. 461-471, 2010.

[17] F. Tung, A. Wong, D. A. Clausi, "Enabling scalable spectral clustering for image segmentation," *Pattern Recognition*, Vol. 43, No. 12, pp. 4069-4076, 2010.

[18] S. F. Ding, H. Ding, F. X. Jin, "An Improved Algorithm for Attribute Reduction Based on Database Technology," *Information-An International Interdisciplinary Journal*, Vol. 14, No. 10, pp. 3489-3498, 2011.

[19] X. Z. Wang, Y. D. Wang, Y. Zhan, et al, "Optimization of K-means Clustering by Feature Weight Learning," *Journal of Computer Research and Development*, Vol. 40, No. 6, pp. 869-873, 2003.

[20] S. F. Ding, H. Zhu, W. K. Jia, C. Y. Su, "A survey on feature extraction for pattern recognition," *Artificial Intelligence Review*, Vol. 37, No. 3, pp. 169-180, 2012.

[21] L. G. S. Giraldo, G. C. Domínguez, "Weighted feature extraction with a functional data extension," *Neurocomputing*, Vol. 73, No. 10-12, pp. 1760-1773, 2010.

[22] L. Edwin, "On-line incremental feature weighting in evolving fuzzy classifiers," *Fuzzy Sets and Systems*, Vol. 163, No. 1, pp. 1-23, 2011.

[23] S. F. Ding, B. J. Qi, H. J. Jia, et al, "Research of Semi-supervised Spectral Clustering Based on Constraints Expansion," *Neural Computing and Applications*, 2012 (DOI: 10.1007/s00521-012-0911-8).

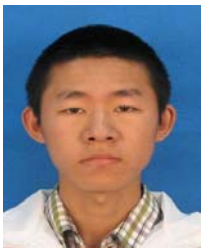
[24] D. Hamad, P. Biela, "Introduction to spectral clustering," *Proceedings of the International Conference on Information and Communication Technologies from Theory to Applications - ICTTA'08*, pp. 634-639, 2008.

[25] S. F. Ding, H. J. Jia, L. W. Zhang, et al, "Research of semi-supervised spectral clustering algorithm based on pairwise constraints," *Neural Computing & Applications*, 2012 (DOI: 10.1007/s00521-012-1207-8).

[26] A. Celikyilmaz, "Soft-Link Spectral Clustering for Information Extraction," *Proceedings of the 2009 IEEE International Conference on Semantic Computing (ICSC)*, pp. 434-441, 2009.

[27] X. C. Zhang, Q. Z. You, "An improved spectral clustering algorithm based on random walk," *Frontiers of Computer Science in China*, Vol. 5, No. 3, pp. 268-278, 2011.

[28] X. Y. Liu, X. Yong, H. F. Lin, "An Improved Spectral Clustering Algorithm Based on Local Neighbors in Kernel Space," *Computer Science and Information Systems*, Vol. 8, No. 4, pp. 1143-1157, 2011.
[29] Y. Wang, Y. Jiang, Y. Wu, et al, "Spectral Clustering on Multiple Manifolds," *IEEE Transactions on Neural Networks*, Vol. 22, No. 7, pp. 1149-1161, 2011.
[30] F. P. Nie, Z. N. Zeng, I. W. Tsang, et al, "Spectral Embedded Clustering: A Framework for In-Sample and Out-of-Sample Spectral Clustering," *IEEE Transactions on Neural Networks*, Vol. 22, No. 11, pp. 1796-1808, 2011.
[31] F. Zhao, H. Q. Liu, L. C. Jiao, "Spectral clustering with fuzzy similarity measure," *Digital Signal Processing*, Vol. 21, No. 6, pp. 701-709, 2011.
[32] W. Y. Chen, Y. Q. Song, H. J. Bai, et al, "Parallel Spectral Clustering in Distributed Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 3, pp. 568-586, 2011.
[33] Z. Pawlak "Rough sets," *International Journal of Computer and Information Sciences*, Vol. 11, No. 5, pp. 341-356, 1982.
[34] G. Y. Wang, Y. Y. Yao, H. Yu, "A Survey on Rough Set Theory and Applications," *Chinese Journal of Computers*, Vol. 32, No. 7, pp. 1229-1246, 2009.
[35] D. Q. Miao, J. Wang, "On the Relationships Between Information Entropy and Roughness of Knowledge in Rough Set Theory," *Pattern Recognition and Artificial Intelligence*, Vol. 11, No. 1, pp. 34-40, 1998.
[36] S. F. Ding, "Artificial Intelligence," Beijing: *Tsinghua University Press*, 2011.
[37] I. Duntsch, G. Gediga, "Uncertainty measures of rough set prediction," *Artificial Intelligence*, Vol. 106, No. 1, pp. 106-109, 1998.
[38] M. J. Wierman, "Measuring uncertainty in rough set theory," *International Journal of General Systems*, Vol. 28, No. 4, pp. 283-297, 1999.
[39] W. X. Zhang, G. F. Chou, "Uncertain Decision Based on Rough Set," Beijing: *Tsinghua University Press*, 2005.
[40] G. Ma, S. F. Ding, Z. Z. Shi, "Rough RBF Neural Network Based on Extreme Learning," *Microelectronics & Computer*, Vol. 29, No. 8, pp. 9-14, 2012.
[41] X. M. Tao, S. Y. Song, P. D. Cao, et al, "A Spectral Clustering Algorithm Based on Manifold Distance Kernel," *Information and Control*, Vol. 41, No. 3, pp. 307-313, 2012.
[42] A. Y. Ng, M. I. Jordan, Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in Neural Information Processing Systems*, Vol. 14, pp. 849-856, 2002.
[43] X. F. Lei, K. Q. Xie, F. Lin, et al, "An Efficient Clustering Algorithm Based on Local Optimality of K-Means," *Journal of Software*, Vol. 19, No. 7, pp. 1683-1692, 2008.



Hongjie Jia is currently a graduate student now studying at School of Computer Science and Technology, China University of Mining and Technology, and his supervisor is Prof. Shifei Ding. He received his B.Sc. degree in computer science from China University of Mining and Technology in 2011. His research interests include pattern recognition, machine learning, and spectral clustering et al.

Shifei Ding is a professor and Ph.D. supervisor at China University of Mining and Technology. His research interests include intelligent information processing, pattern recognition, machine learning, data mining, and granular computing et al. He has published 3 books, and more than 100 research papers in journals and international conferences. He received his B.Sc., M.Sc. degree in mathematics and computer science in 1987, 1998 respectively from Qufu Normal University, received Ph.D. degree in computer science from Shandong University of Science and Technology in 2004, and received post Ph.D. degree in computer science from Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, and Chinese Academy of Sciences in 2006.

In addition, prof. Ding is a senior member of China Computer Federation (CCF), and China Association for Artificial Intelligence (CAAI). He is a member of professional committee of Artificial Intelligence and Pattern Recognition, CCF, professional committee of distributed intelligence and knowledge engineering, CAAI, professional committee of machine learning, CAAI, and professional committee of rough set and soft computing, CAAI. He is an associate Editor-in-Chief for International Journal of Digital Content Technology and its Applications (JDCTA), acts as an editor for Journal of Convergence Information Technology (JCIT), International Journal of Digital Content Technology and its Applications (JDCTA) et al.

Hong Zhu is an associate professor at Xuzhou Medical College. Since 2009, she has been a Ph.D. candidate in computer science and technology from the China University of Mining and Technology, and her supervisor is Prof. Shifei Ding. She received her M.Sc. degree in computer science and technology from China University of Mining and Technology in 2007. Her research interests includes granule computing, clustering, parallel computing et al.

Fulin Wu is currently a graduate student now studying in School of Computer Science and Technology, China University of Mining and Technology, and his supervisor is Prof. Shifei Ding. He received her B.Sc. degree in computer science from China University of Mining and Technology in 2012. His research interests include cloud computing, feature selection, pattern recognition, machine learning et al.

Lina Bao is currently a graduate student now studying in School of Computer Science and Technology, China University of Mining and Technology, and her supervisor is Prof. Shifei Ding. She received her B.Sc. degree in computer science from China University of Mining and Technology in 2012. His research interests include cloud computing, feature selection, pattern recognition, machine learning et al.