

Research Article

Cluster Analysis Based on Bipartite Network

Dawei Zhang,¹ Fuding Xie,² Dapeng Wang,¹ Yong Zhang,¹ and Yan Sun³

¹ School of Computer Science and Technology, Liaoning Normal University, Dalian Liaoning 116081, China

² School of Urban and Environmental Science, Liaoning Normal University, Dalian Liaoning 116029, China

³ School of Psychology, Liaoning Normal University, Dalian Liaoning 116029, China

Correspondence should be addressed to Yan Sun; sunyan@lnnu.edu.cn

Received 11 November 2013; Accepted 27 December 2013; Published 10 February 2014

Academic Editor: Fuzhong Nian

Copyright © 2014 Dawei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Clustering data has a wide range of applications and has attracted considerable attention in data mining and artificial intelligence. However it is difficult to find a set of clusters that best fits natural partitions without any class information. In this paper, a method for detecting the optimal cluster number is proposed. The optimal cluster number can be obtained by the proposal, while partitioning the data into clusters by FCM (Fuzzy c -means) algorithm. It overcomes the drawback of FCM algorithm which needs to define the cluster number c in advance. The method works by converting the fuzzy cluster result into a weighted bipartite network and then the optimal cluster number can be detected by the improved bipartite modularity. The experimental results on artificial and real data sets show the validity of the proposed method.

1. Introduction

Clustering is the unsupervised classification of data points into groups or clusters, such that samples in the same group are similar to each other, while patterns in different groups are dissimilar. In past decades, a number of clustering algorithms have been proposed, attempting to classify the given data set into groups by using different similarities. Generally, these algorithms can be classified into three cases: hard (crisp) cluster, soft (fuzzy) cluster, and possibilistic clustering [1–3]. The classical clustering algorithms should be c -means and fuzzy c -means and their various improved versions. However, these algorithms require the user to specify the number c , of clusters which the user does not usually know in advance or may not want to specify. As a consequence, the clustering result often depends on the choice of parameter. Therefore a very challenging problem in cluster analysis is the cluster validity problem in the literatures, which consists of finding the optimal value for c .

In fact, clustering validation is a technique to find a set of clusters that best fits natural partitions (number of clusters) without any class information. Based on FCM algorithm and bipartite network, in this paper, a method for finding the optimal cluster number c is suggested. A weighted bipartite network is first constructed by using of membership matrix

and centers of each clusters obtained by FCM algorithm. Then the bipartite modularity of the bipartite network is extended to the case of the weighted bipartite network. Last, the optimal cluster number c is obtained by optimal community structures. To validate the effectiveness of the proposed method, one has done the experiments on seven artificial data sets and seven real data sets. As a result, the optimal cluster number c has been detected for most of data sets.

2. Related Works

2.1. Some Cluster Validity Indices. The purpose of clustering algorithm is to divide the given data set $X = \{x_1, x_2, \dots, x_n\}$ into groups X_1, X_2, \dots, X_c . The partition matrix U of size $c \times n$ may be represented as $U = [u_{ij}]$, $i = 1, 2, \dots, c$ and $j = 1, 2, \dots, n$, where u_{ij} is the membership of sample x_j to cluster X_i . In the case of crisp partitioning, the following condition should be satisfied: $u_{ij} = 1$ if $x_j \in X_i$; otherwise, $u_{ij} = 0$ [4].

Formally, the goal of clustering analysis can be represented as follows:

$$\begin{aligned} \bigcup_{i=1}^c X_i &= X, \\ X_i &\neq \emptyset, \quad i = 1, 2, \dots, c, \\ X_i \cap X_j &= \emptyset, \quad i = 1, 2, \dots, c, \quad j = 1, 2, \dots, n. \end{aligned} \quad (1)$$

In the case of fuzzy clustering, the membership u_{ij} is in interval $[0, 1]$. The crisp clustering may be viewed as a special case of fuzzy clustering.

No matter what, we take crisp partitioning method or fuzzy clustering algorithm to separate data set into groups, the generated partitions may not reflect the desired clustering of the data because of inappropriate choice of algorithmic parameters. Thus, it is a key problem how to determine the cluster number c , especially for real data sets. Clustering is an unsupervised classification process and therefore it has no a priori information of a data set. That is to say that user has no prior knowledge about the number of classes. For two-dimensional data or data set with small size, it is possible to visually verify the validity of the results. However, the visualization of the data set in high dimensional space or with huge size would be difficult. Thus it is difficult to partition the given data set into clusters which best fit the natural distribution of the data set.

To find the optimal cluster number c , the researchers proposed some validity indices based on fuzzy cluster result obtained by FCM algorithm. Once the partition is obtained by a clustering method, the validity function can help us to validate whether it accurately presents the structure of the data set or not. In Table 1, some popular validity indices are listed. The earliest proposed fuzzy cluster validity functions associated with FCM are the partition coefficient (PC) and partition entropy (PE) [5, 6]. NPC and NPE, modification of the PC and PE, proposed by Dunn and Roubens [7, 8] in succession. These indices only use the membership values and have the advantage of being easy to compute. Considering the compactness and separation of data set, Fukuyama and Sugeno [9] and Xie and Beni [10] proposed FS and XB indices, respectively. Kwon extended Xie and Beni's index to eliminate its tendency to monotonically decrease by introducing a punishing function in Xie and Beni's original validity index [11]. Based on the concepts of hypervolume and density, Gath and Geva [12] put forward the fuzzy hypervolume validity FHV. The introduction of the concepts of fuzzy compactness and fuzzy separation to the traditional validity indices allows Zahid et al. to propose a novel index to evaluate the fitness of partitions produced by fuzzy clustering algorithms [13]. Wu and Yang use the factors from a normalized partition coefficient and an exponential separation measure for each cluster and then pool these two factors to create the PCAES validity index [14]. They also discussed the problem that the validity indexes face in a noisy environment. By solving the maximum value of PBMF index, Pakhira et al. [15] measured the goodness of clustering on different partitions of a data set.

In general, an optimal cluster number c can be found by solving maximum or minimum of these indices, respectively, to produce the best clustering performance for the data set X .

2.2. The Bipartite Network. Networks have attracted a burst of attention in the last decade, with applications to natural, social, and technological networks. To understand the nature of a complex system, a common approach is to map the interconnected objects in the complex system to a complex

network and study the structure of the complex network. Many real-world networks display natural bipartite structure. Examples include the actors-films network and the papers-scientists network [16, 17]. Because one-mode projection of bipartite networks will bring some drawbacks and affect the properties of the networks, it is an important problem to directly analyze the original bipartite networks. In bipartite networks, there are two nonoverlapping sets of nodes called top nodes and bottom nodes. The edges only connect a pair of vertices which belong to different sets.

One of the common characteristics of many bipartite networks is their community structures which are seen as groups of nodes within which connections are dense and between which connections are sparse. Although the notion of community structure is straightforward, construction of an efficient algorithm for identification of the community structure in a bipartite network is highly nontrivial. A number of algorithms for detecting the communities have been developed, each of them attempts to uncover a reasonable good partition of the network. To measure the division quality of a network, various modularity functions have been suggested. Using the modularity, the quality of any assignment of vertices to modules can be assessed. Guimerà et al. [18] proposed a projection based method. It transforms the bipartite network into an one-mode network and uses method for one-mode network to discover communities. They strikingly demonstrated that the analysis of a projection can give incorrect results in a model network. It will affect the properties including the community structures of the networks. Barber [19] extended the Newmans modularity to bipartite network and proposed a method searching communities by minimizing the modularity. The algorithm is based on the idea that the modules in the two parts of the network are dependent, with each part mutually being used to induce the vertices for the other part into the modules. To resolve communities in bipartite networks, Xu et al. [20] proposed an MDL 21 criterion for identifying a good partition of a bipartite network. Recently, a new measurement for community extraction from bipartite networks is proposed which is a straightforward generalization of Newman's modularity [21].

3. The Fuzzy c -Means Clustering Algorithm

The most commonly used fuzzy clustering method should be fuzzy c -means algorithm. The purpose of fuzzy clustering is to divide the data set into c distinct clusters. The fuzzy c -means (FCM) algorithm is proposed by Dunn [22] and then extended by Bezdek [23]. Formally, the FCM algorithm can be expressed as follows.

Let $X = \{x_1, \dots, x_n\}$ be a data set with n points in P -dimensional feature space R^P , $X \subset R^P$. The FCM clustering algorithm partitions X into $1 < c < n$ fuzzy groups by minimizing objective function J_m which is the weighted sum of squared errors within groups and is defined as follows:

$$J_m(U, X) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - v_i\|_A^2, \quad 1 < m < \infty, \quad (2)$$

TABLE 1: Some validity indices based on FCM algorithm.

Name and reference	Index
Bezdek [5]	$PC = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^C u_{ij}^2$
Bezdek [6]	$PE = -\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^C u_{ij} \log(u_{ij})$
Dunn [7]	$NPC = \frac{C \times PC - 1}{C - 1} = \frac{(c/n) \sum_{j=1}^n \sum_{i=1}^C u_{ij}^2 - 1}{C - 1}$
Roubens [8]	$NPE = \frac{n \times PE}{n - C} = \frac{-\sum_{j=1}^n \sum_{i=1}^C u_{ij} \log(u_{ij})}{n - C}$
Fukuyama and Sugeno [9]	$FS = J_m - \sum_{j=1}^n \sum_{i=1}^C u_{ij}^m \ c_i - \bar{c}\ ^2$
Gath and Geva [12]	$FHV = \sum_{i=1}^C \sqrt{\det \left(\frac{\sum_{j=1}^n u_{ij}^m (x_j - c_i)(x_j - c_i)^T}{\sum_{j=1}^n u_{ij}^m} \right)}$
Xie and Beni [10]	$XB = \frac{J_m}{n \min_{i \neq j} \ c_i - c_j\ ^2}$
Bensaid et al. [24]	$BSC = \sum_{i=1}^C \left(\frac{\sum_{j=1}^n u_{ij} \ x_j - c_i\ ^2}{\sum_{j=1}^n u_{ij} \sum_{k=1}^C \ c_i - c_k\ ^2} \right)$
Kwon [11]	$K = \frac{J_m + (1/c) \sum_{i=1}^C \ c_i - \bar{c}\ ^2}{\min_{i \neq j} \ c_i - c_j\ ^2}$
Wu and Yang [14]	$PCAES = \sum_{i=1}^C \left(\frac{\sum_{j=1}^n u_{ij}^2}{\min_{1 \leq i \leq C} \sum_{j=1}^n u_{ij}^2} - \exp \left(\frac{-C \min_{k \neq i} \ c_i - c_k\ ^2}{\sum_{l=1}^C \ c_l - \bar{c}\ ^2} \right) \right)$
Zahid et al. [13]	$ZSC = \frac{\sum_{i=1}^C \ c_i - \bar{c}\ ^2 / C}{\sum_{i=1}^C \left(\sum_{j=1}^n u_{ij}^m \ x_j - c_i\ ^2 / \sum_{j=1}^n u_{ij} \right)} - \frac{\sum_{i=1}^{C-1} \sum_{l=i+1}^C \left(\sum_{j=1}^n (\min(u_{ij}, u_{lj}))^2 / \sum_{j=1}^n \min(u_{ij}, u_{lj}) \right)}{\sum_{j=1}^n \left(\max_{1 \leq i \leq C} u_{ij} \right)^2 / \sum_{j=1}^n \max_{1 \leq i \leq C} u_{ij}}$
Pakhira et al. [15]	$PBMF = \left(\frac{1}{C} \times \frac{\sum_{j=1}^n u_{ij} \ x_j - \bar{c}\ }{\sum_{j=1}^n \sum_{i=1}^C (u_{ij})^m \ x_j - c_i\ } \times \frac{C}{\max_{i,j=1}^C \ c_i - c_j\ } \right)^2$
Chen and Linkens [25]	$P = \frac{1}{n} \sum_{j=1}^n \max_i (u_{ij}) - \frac{\sum_{i=1}^{C-1} \sum_{k=i+1}^C \left((1/n) \sum_{j=1}^n \min(u_{ij}, u_{kj}) \right)}{\sum_{i=1}^{C-1} i}$
Bouguessa and Wang [26]	$SCG = \frac{\text{trace} \left(\sum_{i=1}^C \sum_{j=1}^n u_{ij}^m (c_i - \bar{c})(c_i - \bar{c})^T \right)}{\sum_{i=1}^C \text{trace} \left(\sum_{j=1}^n u_{ij}^m (x_j - c_i)(x_j - c_i)^T / \sum_{j=1}^n u_{ij}^m \right)}$

where $\bar{c} = (1/n) \sum_{j=1}^n x_j$ in above indices.

and subject to

$$u_{ij} \in [0, 1], \quad \sum_{i=1}^c u_{ij} = 1, \quad (3)$$

where $U = [u_{ij}]_{c \times n}$ is a fuzzy partition matrix composed of the degree of membership of data point x_j to i th cluster; (v_1, v_2, \dots, v_c) is a vector of unknown cluster prototype (centers), $v_i \in R^P$. A norm matrix A defines a measure of similarity between a data point and the cluster prototypes. The parameter m controls the fuzziness of membership of each datum. The cluster centroids and the respective membership functions, which are solutions of the constrained optimization problem in (2) are given by the following equations:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\|x_j - v_i\|_A / \|x_j - v_k\|_A \right)^{2/(m-1)}}, \quad (4)$$

$$1 \leq i \leq c, \quad 1 \leq j \leq n,$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, \quad 1 \leq i \leq c. \quad (5)$$

Equations (4) and (5) constitute an iterative optimization procedure.

The FCM algorithm is executed in the following steps

Step 1. Given a preselected c clustering centroids (v_1, v_2, \dots, v_c) and fuzzy factor m ($m > 1$), initialize the fuzzy partition matrix U .

Step 2. Calculate the fuzzy clustering centroid v_i using (5).

Step 3. Use (4) to update the fuzzy membership matrix U .

Step 4. If the improvement in $J_m(U, X)$ is less than a certain threshold (ϵ), then stop; otherwise, go to Step 2.

4. The Improved Bipartite Modularity and the Proposed Method

4.1. The Bipartite Modularity. Very recently, Murata [21] introduced a bipartite modularity to assess the partitioning quality of a given bipartite network. Murata's bipartite modularity can be depicted as follows.

Suppose G is an unweighted bipartite network, whose vertices V is divided into two disjoint sets, V^t and V^b , such that every edge in the network connects a node from one set from the other. Here, a bipartite network is denoted as $G = (V^t, V^b, E)$ and top nodes and bottom nodes are called as V^t and V^b , respectively. Conveniently, we denote the top nodes by X -vertices $\{x_1, \dots, x_n\}$, and Y -vertices the bottom nodes $\{y_1, \dots, y_p\}$.

Let us suppose that M is the number of edges in a bipartite network. Consider a particular division of the bipartite network into X -vertex communities and Y -vertex communities, and the numbers of the communities are L^t and L^b , respectively. V_i^t and V_j^b are the individual community that belong to the sets $V^t = \{V_1^t, V_2^t, \dots, V_{L^t}^t\}$ and $V^b = \{V_1^b,$

$V_2^b, \dots, V_{L^b}^b\}$, respectively. $A(i, j)$ is an adjacency matrix of the network whose element (i, j) is equal to 1 if vertices i and j are connected and is equal to 0 otherwise.

Under the condition that the vertices of V_l and V_s are different types (which means $(V_l \in V^t \wedge V_s \in V^b) \vee (V_l \in V^b \wedge V_s \in V^t)$), e_{lm} (the fraction of all edges that connect vertices in V_l to vertices in V_s) and a_i (its row sums) are defined as follows:

$$e_{lm} = \frac{1}{2M} \sum_{i \in V_l} \sum_{j \in V_s} A(i, j),$$

$$a_i = \sum_j e_{ij} = \frac{1}{2M} \sum_{i \in V_l} \sum_{j \in V_s} A(i, j). \quad (6)$$

The bipartite modularity Q_B is defined as

$$Q_B = \sum_i (e_{ij} - a_i a_j), \quad j = \max_k (e_{ik}). \quad (7)$$

High Q_B value indicates strong community structure in a bipartite network.

4.2. The Proposed Method. For the given data set, the FCM algorithm with a proper distance norm is employed to partition it into clusters. After we have obtained the membership degree matrix, one can construct a weighted bipartite network whose top nodes consist of all clustering centers and bottom nodes are made up of all samples. The membership degrees can be thought as its weighted edges. Obviously, it will produce a fully connected bipartite network.

Example 1. If one partitions the given data set described in Figure 1(a) by FCM algorithm for $c = 3$, a membership degree matrix is obtained. In terms of the idea mentioned above, a weighted bipartite network will easily be constructed. The top nodes consist of three clustering centers and bottom nodes are made up of ten samples. The numbers in Figure 1(b) stands for the weights on each edge in network.

In this case, the bipartite modularity defined in (7) can not be used to measure the quality of partitioning network because the Q_B in (7) is only useful to the unweighted bipartite network. Therefore, it is necessary to improve Q_B so as to deal with weighted bipartite network.

If we define

$$L^t = L^b = C, \quad V_i^t = \{v_i\},$$

$$V_j^b = \left\{ x_i \mid u_{ji} = \max_k (u_{ki}) \right\}, \quad (8)$$

$$M = C \sum_{i=1}^n \sum_{j=1}^n u_{ji},$$

one can successfully extend the bipartite modularity Q_B to the weighted bipartite network which is constructed by FCM algorithm.

In what follows, an algorithm on how to partition a given data into ideal groups is described.

Input: data set S ; threshold ϵ (used in FCM algorithm);

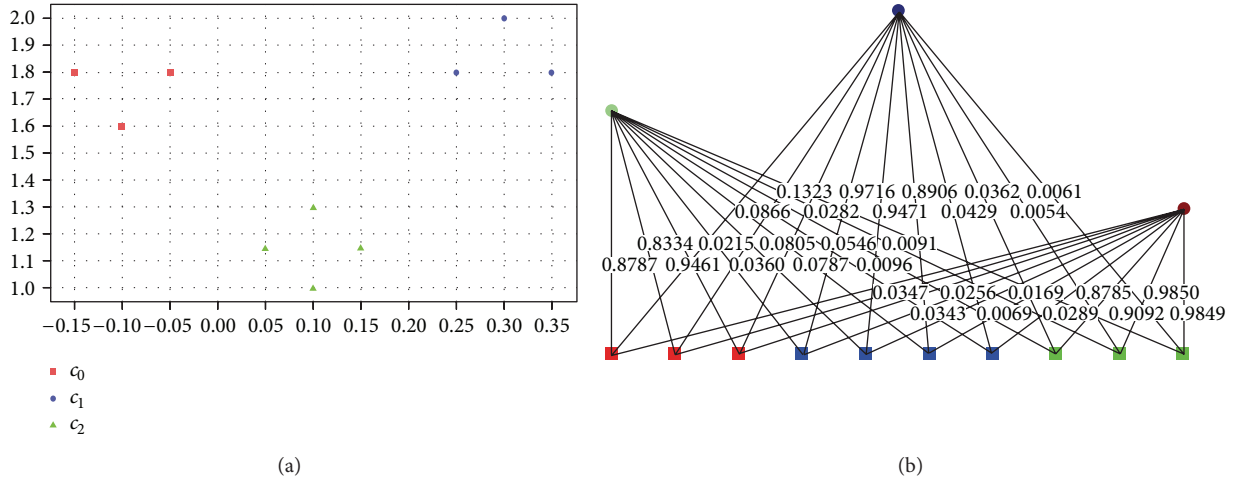


FIGURE 1: The data set and a weighted bipartite network.

TABLE 2: Fourteen artificial and real data sets.

(a) Seven artificial data sets			
Dataset	Clusters	Attributes	Nodes
elliptical_2_10	10	2	500
spherical_2_5	5	2	250
spherical_2_6	6	2	300
spherical_3_4	4	3	400
st900_2_9	9	2	900
y_10_14	14	10	480
y_2_3	3	2	300
(b) Seven real data sets			
Dataset	Clusters	Attributes	Nodes
bupa_7_2	2	7	345
glass_9_6	6	9	214
ionosphere_32_2	2	32	351
iris_4_3	3	4	150
pima_8_2	2	8	768
wall-following_4_4	4	4	5456
wbcd_8_2	2	8	699

TABLE 3: Indices comparable for data sets.

DataSet	Clusters	Q_B	PC	NPC	PE	NPE
elliptical_2_10	10	10	2	10	2	2
spherical_2_5	5	5	2	4	2	2
spherical_2_6	6	6	6	6	6	6
spherical_3_4	4	4	4	4	4	4
st900_2_9	9	9	2	9	2	2
y_10_14	14	15	12	12	15	15
y_2_3	3	3	2	3	2	2
bupa_6_2	2	2	2	2	2	2
glass_9_6	6	4	2	2	2	2
ionosphere_32_2	2	2	2	2	2	2
iris_4_3	3	3	2	2	2	2
pima_8_2	2	3	2	2	2	3
wall-following_4_4	4	4	2	4	2	2
wbcd_8_2	2	2	2	2	2	2

Output: the optimal clustering result.

Step 1: execute FCM algorithm for the given data set;

Step 2: construct a weighted bipartite network by matrix of membership degree U and clustering centroids v_i according to (8).

Step 3: calculate the improved bipartite modularity Q_B .

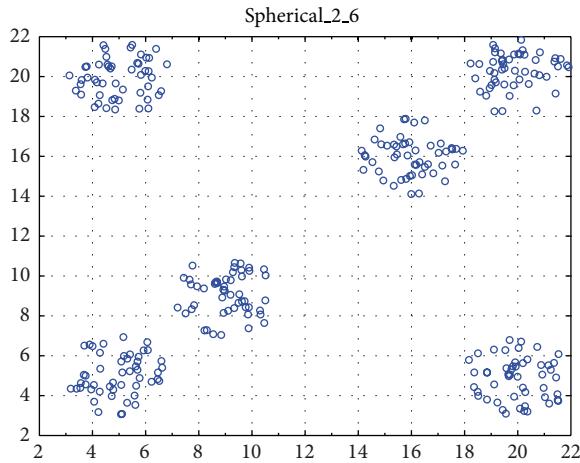
From $c = 2$ to c_{\max} , repeat the above process. Among these values, it is easy to find the maximum value of Q_B . Generally, C_{\max} is unknown. Conventionally, one can take $C_{\max} = \sqrt{n}$ which is widely used in clustering analysis.

5. Experiments

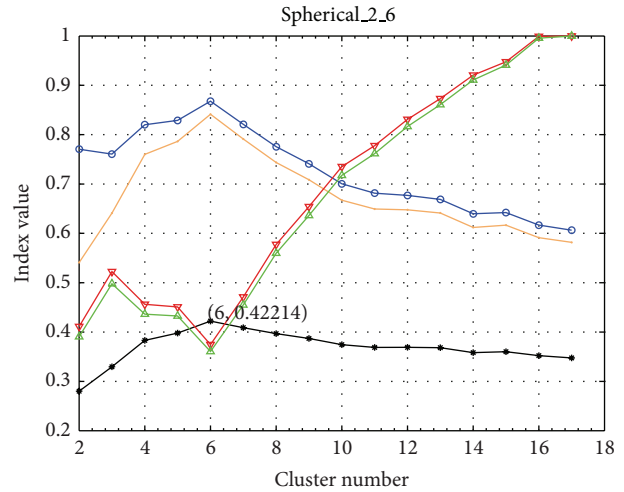
To validate the introduced method, we would like to test it on seven artificial data sets [27–29] and seven well-known real

data sets [30, 31] which are widely used to test the validity of various indices and clustering algorithm. Meanwhile, we also compare the results obtained by the proposal with those achieved by a number of popular validation indices on these data sets. The experimental platform is based on Windows 7 with Intel Core 2 Duo CPU T9550 @ 2.66 GHz 2.67 GHz and 4.00 GB memory. The clustering parameters m and the default minimum amount of improvement in FCM are taken 2.0 and $e - 5$, respectively.

The names of seven artificial data sets imply the dimension of the samples and the actual cluster numbers in the data sets. For example, in the elliptical_2_10 data set, the samples with 10 classes are in 2-dimensional space. The information about seven artificial data sets and seven well-known real data sets are listed in Table 2. The authors can also browse the web page <http://www.isical.ac.in/~sanghami/data.html> and <http://archive.igbmc.fr/projets/fcm/#datasets> to see them in detail.



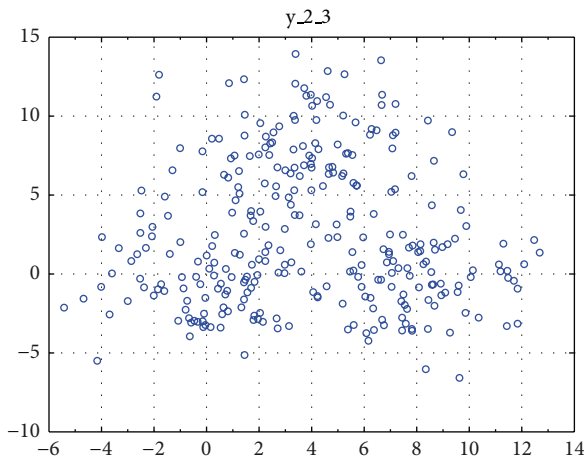
(a) Spherical_2.6



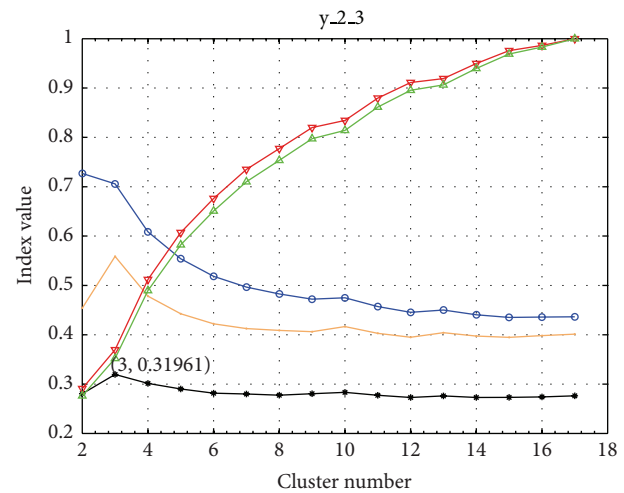
QB (black line with circles)
 PC (blue line with circles)
 NPC (orange line with circles)
 PE (red line with inverted triangles)
 NPE (green line with triangles)

(b) The line charts of five indices on spherical_2.6

FIGURE 2: The scatter diagram of spherical_2.6 and the comparison of the proposal with the other four indices.



(a) y_2.3



QB (black line with circles)
 PC (blue line with circles)
 NPC (orange line with circles)
 PE (red line with inverted triangles)
 NPE (green line with triangles)

(b) The change of values of five indices with cluster number for y_2.3

FIGURE 3: The scatter diagram of y_2.3 and the comparison of the proposal with the other four indices.

The optimal cluster numbers detected by our method and the other four indices are listed in Table 3. As it can be seen that 11 ideal optimal cluster numbers on 14 data sets can be found. The data set spherical_2.6 and spherical_3.4 are well separated so that the practical class numbers 6 and 4 can be found by all indices. The fact that there are no apparent borders among clusters for the rest of the artificial data sets leads to produce the improper cluster numbers, especially for y_2.3 and spherical_2.5. From Figure 2, one can find that the distinct change of values of five indices occur when $c =$

6. This case exactly corresponds to the natural distribution of samples in spherical_2.6. The scatter diagram of y_2.3 in Figure 3(a) shows that three clusters overlap each other. That is to say that there is no clear borderline among these three classes. Thus it is difficult to partition them into three clusters. The correct cluster number 3 is discovered by our method and index NPC.

The selected seven real data sets are typical data sets which are usually employed to test the validity of clustering/classifying algorithms or cluster index. The dimensions

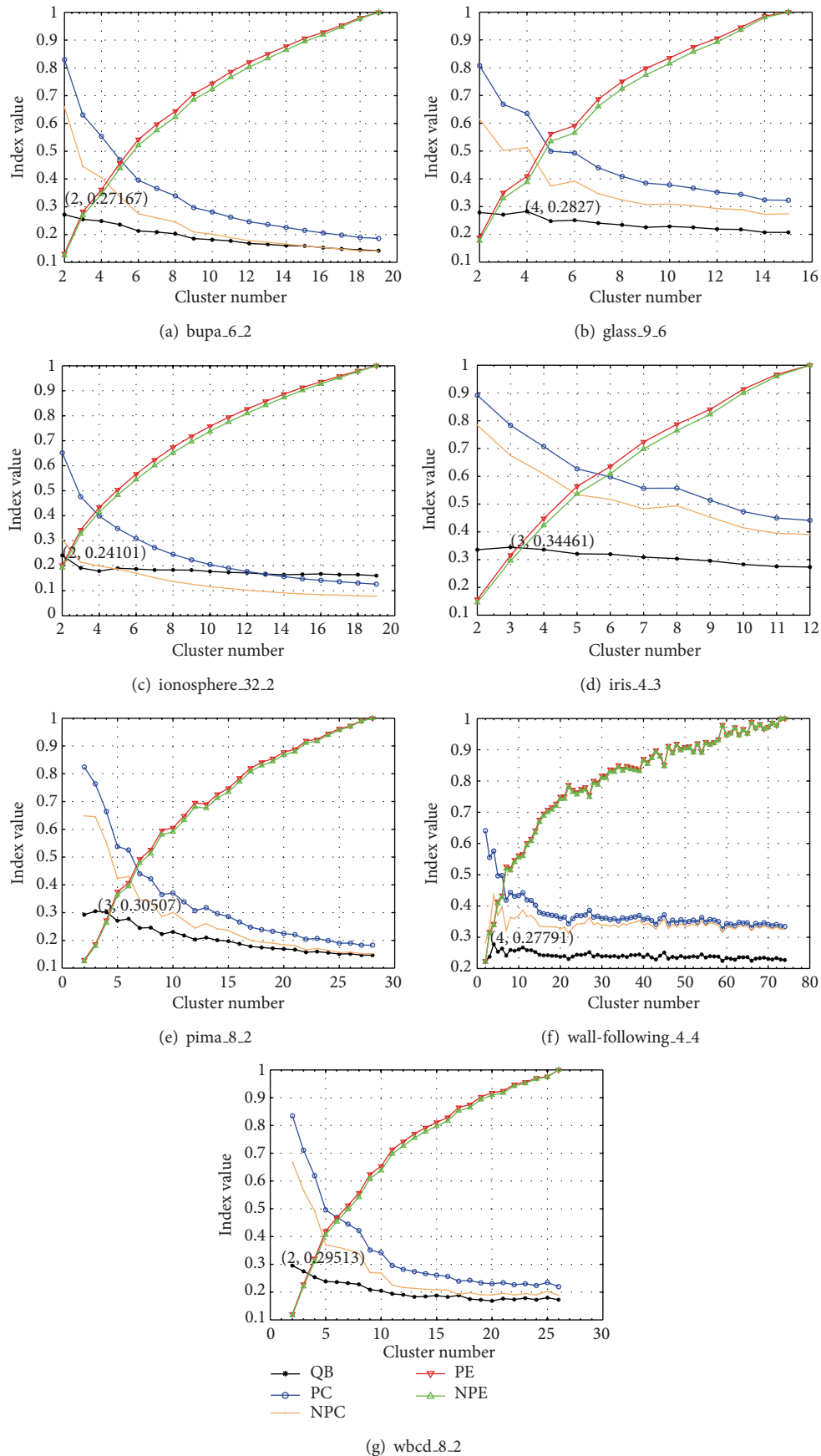


FIGURE 4: The comparison of the proposal with the other four indices on real data sets.

of these data sets range from 4 to 32. That is to say that it is impossible to estimate whether the obtained cluster results are good or not by visualizing method. Furthermore, the clusters of some data sets are heavily overlapped. Therefore, it is difficult to detect the optimal cluster number of these seven data sets by the validity index. More information about these data sets can be found in [31].

In Figure 4, the comparison results are obtained by the proposed method and other four indices on these real data sets. It is easily to see that there exists the distinct difference of Q_B value between the maximum value and others on bupa_6_2, ionosphere_32_2, wbc8_8_2, and wall-following_4_4. Thus the optimal cluster numbers are unanimously detected by the proposed method. This is also true for the other four indices except wall-following_4_4. The Q_B value on iris_4_3, glass_9_6, and pima_8_2 are close to each other when $c = 2$, $c = 3$, and $c = 4$. This implies that it is reasonable, while one partitions these three data sets into 2 or 3 or 4 clusters. Iris data set is a typical example where there are three classes, but looking at the data without the class information, one could also argue that there are two and not three clusters. Fortunately, the optimal cluster number 3 is found for Iris data set. These examples also show that the proposal is superior to other four indices.

6. Conclusions

In this paper, a method is proposed to detect an optimal cluster number for a given data set without any class information. The key idea of this paper is to convert the problem of detecting the optimal cluster number on a data set into the one of finding the optimum community structure in a weighted bipartite network. One of the advantage of the introduced method is to implement the automatic learning process of unsupervised classifying algorithm. The experimental results show the validity of the proposed method. The comparison results indicate that our proposal is superior to the other four indices which are only related to membership obtained by FCM algorithm. However, nonspherical cluster of data set and overlap structure will still affect the cluster results. Therefore it is worthwhile to try to construct or propose new method or index to find the optimal class structure of data set.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the Education Fund of Liaoning Province (no. L2012381), Liaoning province education science “twelfth-five-year” plan project (no. JG12CB114) and NSFC (under Grant no. 61373127).

References

- [1] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, Wiley Interscience, New York, NY, USA, 1973.
- [2] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, John Wiley, New York, NY, USA, 1999.
- [3] R. Krishnapuram and J. M. Keller, “Possibilistic approach to clustering,” *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.
- [4] W. N. Wang and Y. J. Zhang, “On fuzzy cluster validity indices,” *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2095–2117, 2007.
- [5] J. C. Bezdek, “Numerical taxonomy with fuzzy sets,” *Journal of Mathematical Biology*, vol. 1, no. 1, pp. 57–71, 1974/75.
- [6] J. C. Bezdek, “Cluster validity with fuzzy sets,” *Journal of Cybernetics*, vol. 3, no. 3, pp. 58–73, 1973.
- [7] J. C. Dunn, “Indices of partition fuzziness and the detection of clusters in large data sets,” in *Fuzzy Automata and Decision Processes*, Elsevier, New York, NY, USA, 1977.
- [8] M. Roubens, “Pattern classification problems and fuzzy sets,” *Fuzzy Sets and Systems*, vol. 1, no. 4, pp. 239–253, 1978.
- [9] Y. Fukuyama and M. Sugeno, “A new method of choosing the number of clusters for the fuzzy c -means method,” in *Proceedings of 5th Fuzzy Systems Symposium*, pp. 247–250, 1989.
- [10] X. L. Xie and G. Beni, “A validity measure for fuzzy clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841–847, 1991.
- [11] S. H. Kwon, “Cluster validity index for fuzzy clustering,” *Electronics Letters*, vol. 34, no. 22, pp. 2176–2177, 1998.
- [12] I. Gath and A. B. Geva, “Unsupervised optimal fuzzy clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 773–780, 1989.
- [13] N. Zahid, M. Limouri, and A. Essaid, “A new cluster-validity for fuzzy clustering,” *Pattern Recognition*, vol. 32, no. 7, pp. 1089–1097, 1999.
- [14] K. L. Wu and M.-S. Yang, “A cluster validity index for fuzzy clustering,” *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1275–1291, 2005.
- [15] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, “Validity index for crisp and fuzzy clusters,” *Pattern Recognition*, vol. 37, no. 3, pp. 487–501, 2004.
- [16] D. J. Watts and S. H. Strogatz, “Collective dynamics of “small-world” networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [17] M. E. J. Newman, “Scientific collaboration networks. I. Network construction and fundamental results,” *Physical Review E*, vol. 64, no. 1, 8 pages, 2001.
- [18] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, “Module identification in bipartite and directed networks,” *Physical Review E*, vol. 76, 2007.
- [19] M. J. Barber, “Modularity and community detection in bipartite networks,” *Physical Review E*, vol. 76, no. 6, p. 9, 2007.
- [20] K. K. Xu, C. J. Tang, C. Li et al., *Database Systems For Advanced Applications*, vol. 5981 of *Lecture Notes in Computer Science*, 2010.
- [21] T. Murata, “Modularity for bipartite networks,” in *Data Mining for Social Network Data*, *Annals of Information Systems*, N. Memon, J. J. Xu, D. L. Hicks, and H. Chen, Eds., vol. 12, Springer, Berlin, Germany, 2010.
- [22] J. C. Dunn, “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,” *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

- [23] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY, USA, 1981.
- [24] A. M. Bensaid, L. O. Hall, J. C. Bezdek et al., "Validity-guided (re)clustering with applications to image segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 112–123, 1996.
- [25] M. Y. Chen and D. A. Linkens, "Rule-base self-generation and simplification for data-driven fuzzy models," *Fuzzy Sets and Systems*, vol. 142, no. 2, pp. 243–265, 2004.
- [26] M. Bouguessa and S. R. Wang, "A new efficient validity index for fuzzy clustering," in *Proceedings of 3rd International Conference on Machine Learning and Cybernetics*, pp. 26–29, Shanghai, China, August 2004.
- [27] S. Bandyopadhyay and U. Maulik, "Nonparametric genetic clustering: comparison of validity indices," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 31, no. 1, pp. 120–125, 2001.
- [28] S. Bandyopadhyay and U. Maulik, "Genetic clustering for automatic evolution of clusters and application to image classification," *Pattern Recognition*, vol. 35, no. 6, pp. 1197–1208, 2002.
- [29] D. Dembèle and P. Kastner, "Fuzzy C-means method for clustering microarray data," *Bioinformatics*, vol. 19, no. 8, pp. 973–980, 2003.
- [30] O. L. Mangasarian and W. H. Wolberg, "Cancer diagnosis via linear programming," *SIAM News*, vol. 23, pp. 1–18, 1990.
- [31] "UCI data sets," <http://www.ics.uci.edu/mllearn/MLRepository.html>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

