A new Ensemble Clustering Algorithm using a Reconstructed Mapping Coefficient

Tuoqia Cao[1,2,3], Dongxia Chang[1,2,3*] and Yao Zhao[1,2,3]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing 100044 – China

[2]Bijing Key Laboratory of Advanced Information Science and Network Technology

Beijing 100044 - China

[3]School of Computer and Information Technology, Beijing Jiaotong University

Beijing 100044 - China

[e-mail: dxchang@bjtu.edu.cn]

*Corresponding author: Dongxia Chang

Received February 5, 2020; revised May 6, 2020; accepted May 30, 2020;

published July 31, 2020

Abstract

Ensemble clustering commonly integrates multiple basic partitions to obtain a more accurate clustering result than a single partition. Specifically, it exists an inevitable problem that the incomplete transformation from the original space to the integrated space. In this paper, a novel ensemble clustering algorithm using a newly reconstructed mapping coefficient (ECRMC) is proposed. In the algorithm, a newly reconstructed mapping coefficient between objects and micro-clusters is designed based on the principle of increasing information entropy to enhance effective information. This can reduce the information loss in the transformation from micro-clusters to the original space. Then the correlation of the micro-clusters is creatively calculated by the Spearman coefficient. Therefore, the revised co-association graph between objects can be built more accurately because the supplementary information can well ensure the completeness of the whole conversion process. Experiment results demonstrate that the ECRMC clustering algorithm has high performance, effectiveness, and feasibility.

Keywords: Ensemble clustering, supplementary information, reconstructed mapping coefficient, information entropy, Spearman coefficient.

This work was supported in part by National Natural Science Foundation of China (No. 61532005)

http://doi.org/10.3837/tiis.2020.07.013 **ISSN : 1976-7277**

1. Introduction

Clustering, as an important algorithm in data mining field, is used to identify groups of similar characteristics. More specifically, in clustering, a set of data is grouped into clusters that data in the same sense. In recent years, a large number of improved clustering algorithms based on traditional algorithms have made progress to a large extent by using various techniques. Generally, these algorithms can be divided into the improvement of k-means [1-2], modification based on mean shift [3-4], graph-based partition optimization [5-7]. However, no matter how to improve the algorithm cannot avoid the disadvantages such as poor visibility, poor stability and high sensitivity to the initialization of a single clustering algorithm [8-9]. In order to overcome the shortcomings mentioned above, the research of ensemble clustering algorithm has become a hot topic in recent years. Ensemble clustering, emerged as a powerful tool, aims at combining multiple different clustering results into a probably better and more robust consensus clustering [10]. Many ensemble clustering algorithms [11-13] have been proposed in recent years. These algorithms can be broadly divided into these categories: based on the optimization of the utility function, based on the weights assigned to the base clustering and based on enhancing effective information.

Firstly, the representative algorithms based on the utility function are KCC [14], ECC [15], SIVID [16] and so on. Utility function, which can be optimized by calculating the consensus partition from multiple base partitions, is defined to supervise the ensemble clustering process. In [14], the consensus clustering problem is transformed into a k-means clustering problem. And the utility function used in KCC is calculated by the Shannon entropy which makes the algorithm more robust. In order to improve the efficiency, ECC transfers the above model to a modified one. And according to efficiently optimize the objective function that fuses many basic partitions to a consensus one. A binary matrix gained from each basic partition via 1-of-K encoding and a modified distance function are derived, and they are beneficial to optimize the utility function. The experiments indicate that ECC is more suitable for incomplete multi-view data. However, it is considered that the traditional ensemble clustering focuses more on measuring the validity and diversity on base partitions, while ignoring the structure of original objects. In order to address this problem, a new category utility function is proposed in SIVID. The effectiveness and robustness have been greatly improved.

Secondly, some algorithms are based on the weights assigned to the base clusterings. For the clustering ensemble, the quality and diversity measures are considered as two critical factors for the selection from the base clustering to be the ensemble. Then in [17], a generalized validity function is presented for evaluating the base clustering results of categorical data. And a normalization method based on the obtained bounds is proposed, which purpose is to reduce the effects of data characteristics on the performance of the base

clustering algorithm. The clustering validity indices are adapted to assess the quality of the candidate base clustering and to select the base partitions. In LWGP [18], it believes that not every base clustering is a correct and useful classification. Moreover, the wrong existence of some base clustering that does not contain any useful information can even interfere with the final consensus process. Thus, LWGP creatively presents an ensemble-driven cluster validity index (ECI) to evaluate the result of ensemble clusters. In this process, a local weighting scheme is presented to extend the conventional co-association matrix into the LWCA matrix via the ECI measure. In Ref. [20], the weights are globally assigned to the base clusterings according to their clustering quality.

Finally, algorithms to enhance useful information have been increasingly important in recent years. From this point of view, many excellent algorithms have taken this as a starting point to make the whole ensemble clustering more effective and accurate. With aggregating several basic clustering to generate a single output clustering, most of them generate the final solution based on incomplete information of a cluster ensemble [19]. SEC [21] adds the idea of spectral clustering to it when constructing the similarity matrix between objects. In this algorithm, a formula for calculating the weighting coefficient is designed. This operation adds a lot of object-level information in the subsequent consensus process and makes the algorithm more effective. PTGP [22] aims to use the local structure and the size of the local quantity. In the process of calculating the intersection of the micro-clusters, each intersection that is not treated equally avoids the structure information lost. It is considered that carrying out the straightforward mapping from the base clustering to the object-wise co-association matrix lacks some information. This means different clusters are independent of each other. In fact, the potentially rich information hidden in the relationship between different clusters is lost. In order to overcome this problem, ECPCS [23] uses the random walk to extract the rich information contained in the co-association matrix. In this algorithm, the random walk [24] process is performed on the associated graph recovering some useful information that has been lost. It is a dynamic process using the idea of a probability transfer matrix that transits from the object to one of its neighbors at each step with a certain probability.

In order to make full use of the useful information in the process of the mapping between objects and micro-clusters, a novel ensemble clustering algorithm using the reconstructed mapping coefficient (ECRMC) is proposed in this paper. Inspired by the principle of increasing information entropy to supplementary information, a new reconstructed mapping coefficient between the objects and the micro-clusters is introduced which makes less information loss. Furthermore, the similarity between the micro-clusters is calculated using the Spearman's correlation coefficient [25] which generates more accurate similarity. Finally, the edge weights between the whole objects in the correlation graph are captured simultaneously, and the hierarchical clustering is proceeding among the object-level correlation graph to get the ensemble result.

The remainder of the paper is organized as follows. Section 2 provides the development of

the ensemble clustering based on enhancing effective information more specifically. Then a detail of our ECRMC clustering algorithm is described in Section 3. The experimental results are given in Section 4. Finally, Section 5 offers conclusions.

2. Related Work

In recent years, more and more ensemble clustering algorithms have been proposed to solve the shortages of the single clustering algorithm. The process of ensemble clustering is given in **Fig. 1**. It is obviously that the mapping matrix, the micro-cluster' similarity matrix and the co-association matrix all carry some information. Generally, the more information each matrix holds, the more helpful for completing the consensus cluster. However, the errors in the calculation of the mapping relationship can make some valid information contained in the mapping matrix lost. And due to the sensitivity of the similarity measure function to data distribution, some relevant information in the micro-clusters' similarity matrix is not fully expressed. Both of them are the important factors that make the information of ensemble clustering incomplete. Moreover, many types of information are involved in the process of clustering, including feature information, context information, relevance information and so on. However, the redundancy of information may lead to the inaccurate characteristics of the clustering process, noise interference and other phenomena. NSCR [26] adopts the nonnegative spectral analysis to select the most discriminative features information. And for learning the best representations information, a novel robust structured NMF learning framework is proposed in [27]. DEC [28] is to collaboratively explore the rich context information of social images.

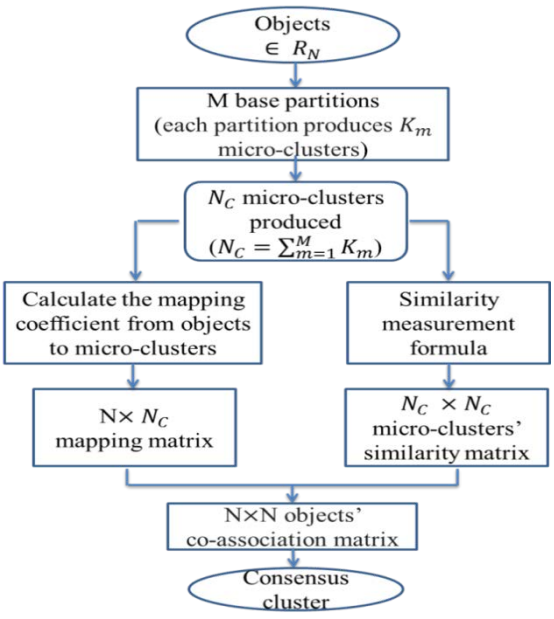


Fig. 1. The flow chart of the ensemble clustering algorithm.

In the following, some ensemble clustering algorithms improved using enhancing useful information will be introduced in detail. From the supplementary information of the objects' co-association matrix, two ensemble clustering algorithms, NegMM [29] and GSOE [30], were introduced. NegMM is deduced from the formula that the useful information is increased relatively when the redundant information can be reduced under one specific condition. Different thresholds are set and the similarity connections smaller than the thresholds are abandoned. This operation helps obtain lots of remodeled objects' co-association matrices which are applied in the consensus process. GSOE is also a typical example of supplying more effective information using the optimizing co-association matrix between the objects. After ranking the collection, it allows inserting effective ranking in any selection function. Map the rankings onto a graph structure and the ensemble cluster onto a mining problem in a graph. All of these are innovative ways to improve the co-association matrix. From using the supplementary information of the micro-cluster' similarity matrix, PTGP [22] and ECPCS [23] use the random walk to construct the similarity graph. They think any vertex not only has a connectivity relationship with its nearest neighbor, but also has weights between its k-top neighbors. Therefore, more effective information can be used in the micro-cluster' similarity matrix and this can improve the performance of the ensemble clustering. In our ECRMC, the information of the mapping matrix is increased by the reference set based on the principle of increasing information entropy. And the Spearman correlation coefficient is used to build the micro-clusters' similarity matrix more accurately. Both of them are helpful for forming a more precise co-association matrix between objects and then beneficial to the consensus cluster.

3. Reconstructed Mapping Coefficient Ensemble Clustering by Increasing Information Entropy

Let $X = \{x_1, x_2, ..., x_N\}$ be a finite subset of a N-dimensional vector space, the goal of the ensemble clustering is to assign the relatively authentic label to each object. In this section, our new ensemble algorithm based on the reconstructing mapping coefficient will be described in detail. Here, the k-means algorithms with different numbers of clusters were used to obtain the initial partitions firstly. And for each k-means, the number of clusters K_m is a random number from $(K, \min(\sqrt{N}, 100))$ where K is the number of categories in the ground truth and N is the number of objects in the datasets. After M initial partitions, micro-clusters $\{Q_1^{K_1}, Q_2^{K_1} ..., Q_{K_1}^{K_1}, Q_1^{K_2}, Q_2^{K_2} ..., Q_{K_2}^{K_2}, ..., Q_1^{K_M}, Q_2^{K_M} ..., Q_{K_M}^{K_M}\}$ are produced, and let $N_C = \sum_{m=1}^{M} K_m$. The aim of our algorithm is to combine the multiple k-means operations with different cluster numbers into a better result. In fact, this can be viewed as an information conversion that conveys the information from the multiple k-means operations to the ensemble one. Obviously, the mapping coefficient matrix and the co-association matrix can be improved to contain more information. The reconstructed mapping matrix and the revised co-association matrix we proposed are beneficial to enhance the information. Moreover, the

Spearman coefficient which is used to calculate the similarity between micro-clusters helps the micro-clusters' matrix more accurately.

3.1 Reconstructed Mapping Coefficient Matrix

Mapping coefficient matrix commendably reflects the degree of intimacy from objects to micro-clusters. Then we introduce that how to reconstruct the mapping coefficients based on the reference sets, and use the principle of increasing information entropy to make this matrix contains more information.

Definition 1. Entropy is uncertainty. It is also the average information in an ensemble (or event) [31]. Let the information $X=\{x_1,x_2,...,x_i,...,x_N\}$ is constituted of N symbols, and the probability that each symbol appears is $P=\{p_1,p_2,...,p_i,...,p_N\}$, then the information entropy is

$$H(X)=-\sum_{i=1}^{N}p(x_i)\log_b p(x_i) \tag{1}$$

Theorem 1. According to the property of additivity of information entropy, the following conclusion is drawn,

$$H(p_{11},...,p_{1r_1},p_{21},...,p_{2r_2},...,p_{N1},...,p_{Nr_N})= H(p_1,p_2,...,p_N)+\sum_{i=1}^{N}p_i H(\frac{p_{i1}}{p_i},\frac{p_{i2}}{p_i},...,\frac{p_{ir}}{p_i}) \tag{2}$$

where $p_1+p_2+...+p_N=1$, $p_i=\sum_{j=1}^{r_i}p_{ij}$, and r_i denotes the number of dimension that p_i can be decomposed .

Theorem 2. The principle of increasing information entropy. As this principle described, when the low dimensional distribution is decomposed into the high dimensional distribution, the information entropy increases gradually.

$$H(p_{11},...,p_{1r_1},p_{21},...,p_{2r_2},...,p_{N1},...,p_{Nr_N})\geq H(p_1,p_2,...,p_N) \tag{3}$$

Definition 2. Reference set (R_S). Reference set is defined as a collection of neighbors that can provide reference opinions for central object. When determining whether an object is related to the micro-cluster, the function of the reference set element is to provide a reference opinion. And the reference R_S is defined as follows, r is used to control the number of elements in reference set.

$$R_S(x_i)=\{x_j \mid dist(x_i,x_j)<r\} \tag{4}$$

Assume each object can be represented by a D-dimensional vector. The neighbors of the element obtained according to the Euclidean distance constitute the collection of reference set. The distance expression is as follows,

$$\text{dist}(x_i,x_j)=\sqrt{\sum_{d=1}^{D}(x_{id}-x_{jd})^2} \tag{5}$$

and the weight of each element in the reference set is calculated as follows,

$$\lambda_j = 1 - \frac{\mathrm{dist}(x_i, x_j)}{\sum_{x_j \in R_s(x_i)} \mathrm{dist}(x_i, x_j)} \tag{6}$$

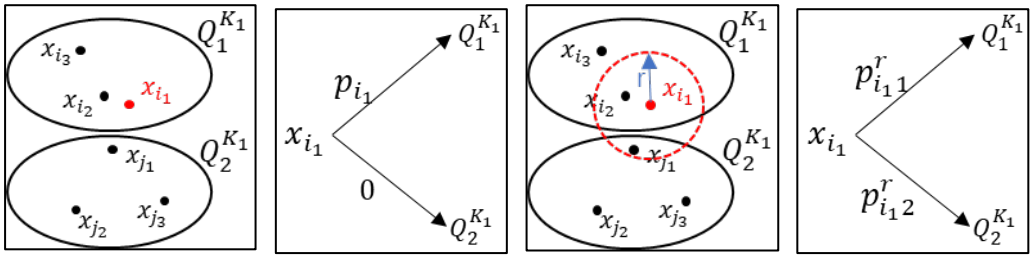
In fact, some useful information has been lost in the process of the mapping from objects to micro-clusters. The original mapping matrix $\boldsymbol{P}$ is sparse and is defined as:

$$\boldsymbol{P} = \{p_{il}\}_{N \times N_C} \tag{7}$$

$$p_{il} = \begin{cases} 1, & \text{if } \ i \in Q_l \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where Q_l denotes the $l\text{-}th$ micro-cluster, p_{il} denotes the original mapping coefficient from $i\text{-}th$ object to $l\text{-}th$ micro-cluster. And the previously proposed algorithms directly used this sparse matrix to calculate the objects' co-association matrix. Obviously, objects' co-association matrix is the core part of the ensemble clustering. However, this matrix contains so little information that it is of little guiding significance for the ensemble process. In fact, each object is not only considered to have a mapping relationship within the cluster divided by a single basis partition, but also has a similarity relationship with some points outside the cluster. Therefore, in our new algorithm, the information entropy and the reference set are used to increase the information. The elements in the reference sets are taken into account while calculating the mapping coefficient. The mapping matrix reconstructed by the principle of increasing information entropy is more convincing than the original one.

According to Definition 2, using the idea of reference sets, we think that if one object which has not mapping relationship to the micro-cluster but its reference neighborhood has mapping relationship to the micro-cluster, then we will add the possibility of this object to the micro-clusters. Here, we use the example shown in **Fig. 2** to describe this process. Assume two micro-clusters $\{Q_1^{K_1}, Q_2^{K_1}\}$ ($K_1 = 2$) can be obtained by an initialization partition shown in **Fig. 2(a)** and the traditional mapping relationship from x_{i_1} to the micro-clusters is shown in **Fig. 2 (b)**. In order to retain more opinions when making the final decision of ensemble clustering, $R_S(x_{i_1})$ is used to represent the reference set of x_{i_1} which is shown using the red circle shown in **Fig. 2 (c)**. Because the $x_{j_1} \in R_S(x_{i_1})$ having the mapping relationship with $Q_2^{K_1}$, then x_{i_1} also builds mapping relationship with it. By considering the opinions in reference set, the reconstructed mapping relationship is shown in **Fig. 2 (d)**. The reconstructed mapping coefficients from x_{i_1} to $Q_1^{K_1}$ and $Q_2^{K_1}$ are $p_{i_1 1}^r$ and $p_{i_1 2}^r$, and satisfy $p_{i_1} = p_{i_1 1}^r + p_{i_1 2}^r$. And specific calculation expressions are shown next in detail.



(a) (b) (c) (d)

Fig. 2. (a)The initialization partition, (b) The traditional mapping relationship, (c) The partition with reference set, (d) The reconstructed mapping relationship.

And then the reconstructed mapping coefficient matrix $\boldsymbol{P}^{\boldsymbol{r}}$ can be written as

$$\boldsymbol{P}^{\boldsymbol{r}} = \{p_{il}^r\}_{N \times N_C} \tag{9}$$

$$\rho_{il} = \sum_{x_j \in R_s(x_i)} \lambda_j p_{il} \tag{10}$$

$$p_{il}^r = \frac{\rho_{il}}{\sum_{l'=1}^{K_m} \rho_{il'}} \tag{11}$$

Where p_{il} is the traditional mapping coefficient from $i\text{-}th$ object to $l\text{-}th$ micro-cluster, p_{il}^r is the reconstructed mapping coefficient from $i\text{-}th$ object to $l\text{-}th$ micro-cluster using the idea of reference set, K_m is the number of micro-clusters in each base partition. Because the base partitions are hard cluster, there exists only one mapping relationship from one object to the K_m micro-clusters obtained by one base partition, and the mapping coefficient is 1. On the basis of Eq. (10), the original mapping relationship of non-1-or-0 has become a probabilistic mapping. And by the normalized operation of Eq. (11), the original mapping coefficient can be decomposed into the sum of reconstructed mapping coefficients. In this way, the probability dimensions of the mapping from an object to the micro-clusters increases. According to the principle of increasing information entropy in Theorem 2, increasing the dimension of probability distribution can increase the information entropy, then the information contained in the reconstructed mapping matrix increases. In the next step, we will use this reconstructed mapping matrix to revise the co-association matrix to make the objects' similarity more accurate.

3.2 Revised Co-association Matrix

In section 3.1, the reconstructed mapping coefficient can be obtained from the object mapping to the micro-cluster and it reflects the degree of the intimate relationship between objects and micro-clusters. After the new mapping matrix from objects to micro-clusters finished, it can be used to revise the co-association matrix. The co-association matrix gives the similarity between the objects, and this revised co-association matrix will be used in the subsequent consensus processes. The traditional co-association matrix depicts the frequency of a pair objects appearing in the same micro-cluster shown in Eq. (12).

$$\boldsymbol{C} = \{c_{ij}\}_{N \times N} \tag{12}$$

$$c_{ij} = \frac{\sum_{l=1}^{N_C} T(i,j,Q_l)}{M} \tag{13}$$

$$T(i,j,Q_l) = \begin{cases} 1 & x_i \in Q_l, x_j \in Q_l \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

N_C denotes the number of the micro-clusters. c_{ij} denotes the similarity between $i\text{-}th$ object and $j\text{-}th$ one. $x_i \in Q_l$ means that x_i has been assigned to the $l\text{-}th$ micro-cluster in one original k-means operations. However, this co-association matrix ignores the interrelationships between different micro-clusters. In order to overcome this problem, a revised co-association matrix is proposed which can simultaneously capture the micro-clusters' information and objects' information.

Firstly, the similarity between the $l\text{-}th$ micro-cluster and $h\text{-}th$ micro-cluster, $S(Q_l, Q_h)$, is given. Generally, the similarity can be calculated by the Pearson correlation coefficient and Spearman correlation coefficient. Pearson correlation coefficient is more applicable to measure the degree of linear correlation between the data samples which obey the normal distribution and Spearman correlation coefficients is a rank correlation coefficient which is also suitable for nonlinear correlation of any distributed data samples. For the generalization of our algorithm, the Spearman correlation coefficient is adopted in our algorithm. The reconstructed mapping coefficients for the $l\text{-}th$ and the $h\text{-}th$ micro-clusters are $Q_l = \{p_{1l}^r, p_{2l}^r, ..., p_{Nl}^r\}$ and $Q_h = \{p_{1h}^r, p_{2h}^r, ..., p_{Nh}^r\}$ respectively. The similarity $S(Q_l, Q_h)$ between $l-th$ and $h-th$ micro-cluster can be written as,

$$S(Q_l, Q_h) = \frac{\sum_{i=1}^{N} (\mu_i - \overline{\mu})(\nu_i - \overline{\nu})}{\sqrt{\sum_{i=1}^{N} (\mu_i - \overline{\mu})^2 \sum_{i=1}^{N} (\nu_i - \overline{\nu})^2}} \tag{15}$$

where μ and ν denote the ranked serial numbers for Q_l and Q_h. $\overline{\mu}$ and $\overline{\nu}$ denote the mean of μ and ν. Here, we use the following example to show how to construct μ and ν. For $Q_l = \{170, 150, 210, 180, 160\}$ and $Q_h = \{180, 165, 190, 168, 172\}$, sort their elements in descending order, then the rank of Q_l and Q_h are $\mu = \{3, 1, 5, 4, 2\}$ and $\nu = \{4, 1, 5, 2, 3\}$ respectively. Therefore, the similarity between Q_l and Q_h is $S(Q_l, Q_h) = 0.7$ which is calculated using Eq. (15).

Then the revised co-association matrix $\boldsymbol{C^r}$ can be calculated as

$$\boldsymbol{C^r} = \{c_{ij}^r\}_{N \times N} \tag{16}$$

$$c_{ij}^r = \frac{\sum_{l=1}^{N_C} \sum_{h=1}^{N_C} T(i, Q_l) T(i, Q_h) S(Q_l, Q_h)}{M} \tag{17}$$

$$T(i, Q_l) = \begin{cases} 1 & x_i \in Q_l \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

$$T(j, Q_h) = \begin{cases} 1 & x_j \in Q_h \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

This revised co-association matrix is calculated based on the similarity of the micro-clusters that objects-pairs divided into. After the revised co-association matrix is obtained, the consensus process is proceeded using this revised co-association matrix.

3.3 Consensus Process

Consensus process is the highlight task of ensemble clustering. Generally, hierarchical clustering [32] is adopted. Hierarchical clustering uses a hierarchical nested cluster tree which is created by calculating the similarity between data points of different category. Let R^0 be the initial clusters, and the number of clusters is initialized as N. That is,

$$R^0 = \{R_1^0, R_2^0, ..., R_{|R^0|}^0\}, |R^0| = N \tag{20}$$

Then merge the two most similar points as a new cluster and the number of clusters is reduced by one. Simply speaking, the merging algorithm of hierarchical clustering is to calculate the mini-distance between data points of each category. The method for calculating the distance between two combined data points is Complete Linkage. Complete Linkage takes the mean of all the distances as the distance between two combined data points. The expression is,

$$sim_{ij} = \frac{\sum_{k \in R_i, l \in R_j} c_{kl}^r}{|R_i||R_j|} \tag{21}$$

where sim_{ij} denotes the average linkage between $i\text{-}th$ and $j\text{-}th$ regions. R_i and R_j denote the elements' number in $i\text{-}th$ and $j\text{-}th$ regions. The true label numbers has been given in advance, through iteration, when the number of regions is equal to the number of true label numbers, the iteration stops. That's the process of consensus process. For clarity, the overall algorithm of ECRMC is summarized in **Table 1**.

Table 1. The ensemble clustering algorithm using a new reconstructed mapping coefficient (ECRMC)

Input: Dataset X Number of the objects in dataset N Number of clusters K Number of the basic clusterings M
· for $m = 1:M$ K_m is equal to a random number from $(K, \min(\sqrt{N}, 100))$; $\{Q_1^{K_m}, Q_2^{K_m}, ..., Q_{K_m}^{K_m}\} = \text{kmeans}(X, K_m)$ End Obtain $N_C = \sum_{m=1}^{M} K_m$ micro-clusters, and initializes the mapping matrix.
· Find the reference sets for each object by Eq. (4).

- Reconstructed the mapping coefficient by Eq. (10)(11) taking the advice of the elements in reference set.
- Calculate the similarity between micro-clusters by Eq. (15).
- Gain the Revised Co-association Matrix by Eq. (16)(17).
- Consensus process by Eq. (21), and achieve the optimal result.

Output: The final clustering result.

3.4 Complexity Analysis

In our ECRMC, the computational cost includes these parts, the initializing k-means operations, the reconstructed mapping matrix, the similarity between micro-clusters and revised co-association matrix. The complexity of the initialized mapping matrix is $O(\sum_{m=1}^{M} l \cdot K_m ND)$ which is obtained by M k-means operations. Here, N is the number of the objects in the datasets, D is the feature dimension, l is the iteration number of the k-means and K_m is the number of clusters. For getting the reconstructed mapping matrix, each object gets its distance from other objects firstly by Eq.(5), its computational complexity is $O(D)$. And then the complexity of getting the mapping value is $O(T \cdot N \cdot D + T \cdot N)$, where T is the number of objects in the reference set, and $T \ll N$. Next, the computational complexity of calculating the micro-clusters similarity is $O(N^2)$. The computational complexity of obtaining the revised co-association matrix is $O(N^2 \cdot {N_C}^2)$, N_C is the number of micro-clusters. Finally, the computational complexity of the consensus process is $O(N^2 \cdot D)$. Therefore, the whole computational complexity of our ECRMC is $O(\sum_{m=1}^{M} lK_m ND + TND + TN + N^2 \cdot {N_C}^2 + N^2 D)$.

4. Experiments

In order to validate the performance of the ECRMC clustering algorithm, a set of experiments are conducted on a variety of benchmark datasets. The performances of the ECRMC, k-means, KCC [14], ECC [15], SEC [21], PTGP [22], ECPCS-HC [23] and LWGP [18] are compared through the experiments. The results show that our ECRMC clustering algorithm has high performance and flexibility.

In the experiments, eleven benchmark datasets are used, Breast Cancer (BC), Letter Recognition (LR), MNIST, Pen Digits (PD), COIL20, USPS, Iris, semeion, Image Segmentation (IS), steel plates faults (SPF) and vehicle silhouettes (VS). BC dataset consists of 683 objects, which can be divided into two categories: diseased cells and normal cells, and each object is 9 dimensions. LR dataset consisted of 20000 English letter objects with 16 dimensions can be divided into 26 categories. MNIST, PD, Semeion and USPS are the handwritten digital datasets with 10 categories, and the characteristic dimension of the separation is 784, 16, 256 and 256. COIL20 contains 20 categories, each rotated 360 degrees

horizontally, taking a picture every five degrees. Iris dataset is also known as iris flower dataset, which contains 150 data samples, divided into 3 categories, 50 for each category. IS dataset consisted of 2100 objects can be divided into 7 categories. SPF dataset is steel plates faults dataset whose characteristic dimension is 27, and can be divided into 7 categories. And VS dataset consisted of 846 objects can be divided into 4 categories. Except MNIST and USPS datasets are from [33], others are from the UCI machine learning repository [34]. For convenience, we summarize the eleven sets in **Table 2** with the characteristics of the data sets. The three columns show the number of objects N, the number of classes k and the dimension of the feature space d.

Table 2. Eleven data sets used in our experiments

datasets	N	d	k
BC	683	9	2
LR	20000	16	26
MNIST	5000	784	10
PD	10992	16	10
COIL20	1440	4096	20
USPS	11000	256	10
Iris	150	4	3
semeion	1593	256	10
IS	2100	19	7
SPF	1941	27	7
VS	846	18	4

4.1 Evaluation Index

In order to evaluate the results of clustering more accurately, three evaluation indicators (Normalized mutual information, Adjusted Rand index and Accuracy rate) are used. All of which are calculated over 20 runs for the eight clustering algorithms.

Normalized mutual information (NMI) [35] Let π^* be the final clustering result obtained by the clustering algorithm, π^g be the ground-truth labels. Then the NMI is calculated as follows,

$$\mathrm{NMI}(\pi^*,\pi^g)=\frac{\sum_{i=1}^{n^*}\sum_{j=1}^{n^g}n_{ij}\log\frac{n_{ij}n}{n_i^* n_j^g}}{\sqrt{\sum_{i=1}^{n^*}n_i^*\log\frac{n_i^*}{n}\sum_{j=1}^{n^g}n_j^g\log\frac{n_j^g}{n}}} \tag{22}$$

where, $n*$ and n^g is the cluster numbers of π^* and π^g respectively. n_i^* and n_j^g is the number of objects in the $i\text{-}th$ and $j\text{-}th$ cluster of π^* and π^g.

Adjusted Rand index (ARI) [36] The ARI measures the agreement of the clustering result with the true cluster structure. ARI is computed by,

$$\mathrm{ARI}(\pi^*,\pi^g)=\frac{2(N_{00}N_{11}-N_{01}N_{10})}{(N_{00}+N_{01})(N_{01}+N_{11})+(N_{00}+N_{10})(N_{10}+N_{11})} \tag{23}$$

where N_{11} is the number of object-pairs that belong to the same cluster in both π^* and π^g, N_{00} is the number of object pairs that belong to different clusters in both π^* and π^g, N_{10} is the number of object-pairs that belong to the same cluster in π^* while belonging to different clusters in π^g, and N_{01} is the number of object-pairs that belong to different clusters in π^* while belong to the same cluster in π^g.

Accuracy rate (AC) The accuracy rate calculates the proportion of label stations in each category. It is given by,

$$\mathrm{AC}(\pi^*,\pi^g)=\frac{N_S}{N} \tag{24}$$

where N is the total number of the objects, and N_S denotes the number of object-pairs that both belong to the same cluster set in π^* and π^g.

4.2 Results

In the experiments, all the algorithms run for 20 times in order to eliminate the effects of randomness embedded in the algorithms. At each run, the integration of 20 basic clustering as a whole and the number of clusters randomly selected in the range of $(K,\min(\sqrt{N},100))$ in each basic clustering, where K is the number of clusters and N is the number of objects in the datasets. The mean and standard deviation of the NMI, the ARI and the AC are given in **Table 4**, **Table 5** and **Table 6**, respectively. From the results of **Table 4**, it may be realized that our ECRMC clustering algorithm performs superior to k-means, KCC, ECC, SEC, PTGP, ECPCS-HC and LWGP algorithm under the NMI scores. Except 4 out of 11 datasets are slightly worse than LWGP, others are all the best scores in the remained algorithms. From the results of **Table 5**, it can be seen the performance of our ECRMC clustering algorithm is better than other seven algorithms. On 7 out of 11 datasets, ECRMC performs better than other algorithms and on the remaining datasets ECRMC exhibits better performance. LWGP ranks first on the LR and IS datasets, and USPS ranks the first on the USPS dataset. **Table 6** shows that our ECRMC method is ranked in the first position with a high AC index for the BC, PD, MNIST, LR, semeion, and SPF datasets respectively. It is a pity that our ECRMC method ranks the third place under the coil20 dataset and LWGP gets

the best grade. In fact, from **Table 4** to **Table 6**, one may observe that our algorithm outperforms other seven algorithms in the average sense.

Table 4. Mean and standard deviation of the NMI(%) produced by the eight algorithms.

datasets	kmeans	KCC	ECC	SEC	PTGP	ECPCS-HC	LWGP	ECRMC
BC	74.191±0.16	77.868±10.22	79.362±1.86	59.413±26.26	74.705±3.01	79.460±1.06	78.057±1.87	**81.395**±2.48
PD	66.667±1.57	67.438±3.74	69.938±2.67	54.449±7.58	73.971±5.69	74.910±5.35	76.903±3.27	**77.875**±4.19
MNIST	50.468±1.37	56.088±2.77	57.446±2.08	53.171±4.32	57.587±1.26	60.359±1.58	63.251±2.26	**63.354**±1.87
LR	34.867±0.59	37.04±0.95	34.881±0.83	33.420±1.44	38.761±1.20	39.230±0.81	**40.902**±1.73	39.775±1.15
COIL20	72.160±2.43	75.534±1.18	75.274±2.27	74.553±2.98	68.392±1.97	76.844±1.19	79.468±1.34	**80.791**±0.92
USPS	44.119±1.22	52.771±3.32	57.009±3.01	48.357±4.36	56.006±5.09	56.951±4.07	**61.399**±3.19	57.552±2.89
Iris	72.253±0.77	76.918±3.57	75.611±2.79	58.502±22.12	77.546±5.02	77.999±0.96	74.567±3.56	**78.359**±1.11
semeion	42.744±2.28	47.443±2.54	55.610±1.64	55.201±2.86	64.087±1.84	60.366±2.24	64.203±1.88	**65.092**±1.43
SPF	8.032±2.90	4.361±1.71	9.899±1.93	7.1533±2.18	13.125±2.29	11.032±1.63	**13.876**±2.67	13.151±1.43
IS	47.452±4.61	52.414±0.91	51.471±2.01	44.909±5.72	54.315±4.85	55.109±2.79	**62.174**±2.91	59.343±3.16
VS	16.805±1.08	16.086±2.93	18.682±1.07	15.863±4.97	18.123±2.09	18.291±1.42	13.356±1.16	**19.624**±0.44
average	46.869	51.269	52.967	45.908	54.147	55.495	57.105	**57.983**

Table 5. Mean and standard deviation of the ARI (%) produced by the eight algorithms.

datasets	kmeans	KCC	ECC	SEC	PTGP	ECPCS-HC	LWGP	ECRMC
BC	84.651±0.12	84.088±14.33	87.980±1.27	46.090±35.22	84.599±3.72	87.717±2.25	87.051±2.05	**88.661**±2.57
PD	54.033±3.81	54.003±6.21	52.574±5.11	30.687±9.99	63.875±4.87	63.711±5.22	67.05±4.69	**67.739**±4.96
MNIST	40.027±2.01	44.515±4.11	45.056±3.73	38.657±6.46	46.383±1.41	50.158±2.30	51.239±1.77	**53.087**±1.58
LR	12.852±0.61	14.828±1.04	13.375±0.66	12.463±1.79	16.166±1.29	14.398±0.73	**16.213**±1.27	15.568±1.45
COIL20	51.487±4.41	56.971±2.73	53.055±2.48	54.752±6.59	45.517±3.07	58.253±2.06	62.748±1.89	**64.791**±1.38
USPS	28.901±1.93	36.828±4.88	37.151±4.58	31.655±4.90	41.677±4.33	**46.385**±1.65	46.147±2.41	44.658±2.97
Iris	69.051±9.25	73.050±2.06	73.120±2.97	50.226±16.26	73.999±7.27	73.643±4.05	70.161±3.41	**74.282**±3.67
semeion	36.307±2.71	36.702±3.21	37.519±3.49	39.397±4.30	52.441±2.48	48.443±1.10	52.073±1.63	**52.892**±1.33
SPF	4.060±1.70	2.899±1.83	4.600±0.48	3.361±1.61	6.508±2.89	6.146±1.19	7.327±2.45	**7.838**±2.12
IS	34.771±5.19	38.749±2.94	37.604±3.29	30.625±6.19	40.118±6.68	45.529±2.89	**52.108**±2.72	48.969±3.15
VS	12.269±0.61	12.993±2.49	**15.050**±0.72	10.911±4.87	12.317±1.25	12.968±1.42	11.693±1.46	13.731±1.34
average	36.219	41.369	41.462	33.439	43.964	46.127	47.619	**48.484**

Table 6. Mean and standard deviation of the AC(%) produced by the eight algorithms.

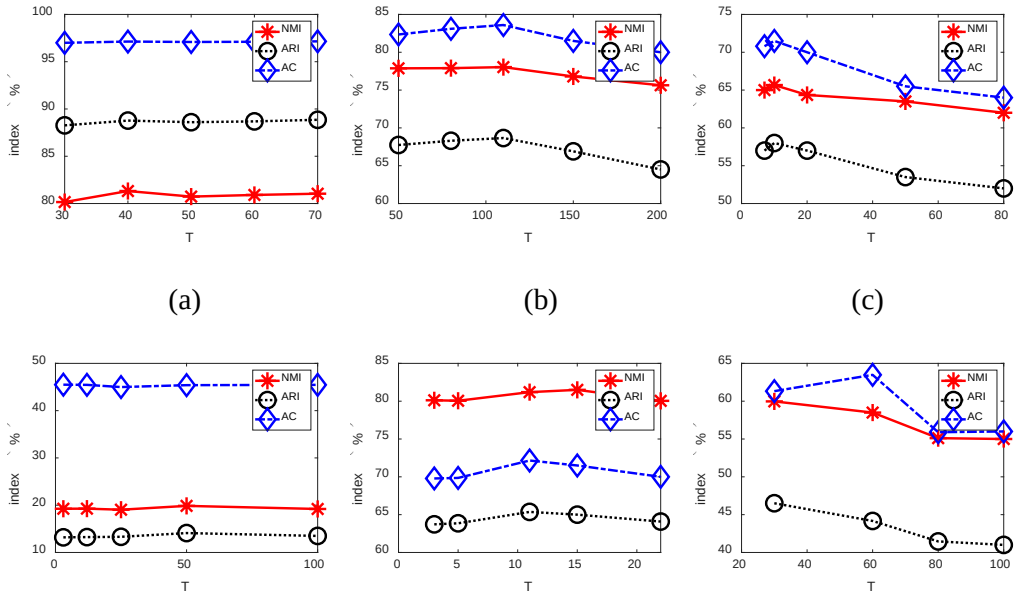
datasets	kmeans	KCC	ECC	SEC	PTGP	ECPCS-HC	LWGP	ECRMC
BC	96.054±0.03	96.691±1.32	96.775±1.96	89.375±10.74	96.025±2.90	96.867±2.52	96.276±2.44	**97.042**±2.84
PD	71.855±3.09	69.857±4.97	69.915±4.38	57.385±14.26	80.085±5.01	77.047±6.16	81.045±4.87	**82.351**±5.73
MNIST	59.849±1.85	62.048±4.11	62.645±3.27	57.582±5.45	66.658±4.11	60.851±3.27	67.096±3.75	**67.482**±3.39
LR	28.275±1.07	28.846±0.96	26.339±1.08	26.443±1.26	31.416±1.36	31.154±1.15	32.987±1.22	**32.991**±0.86
COIL20	61.941±4.39	64.694±3.07	62.633±4.42	63.687±3.49	75.417±1.54	63.674±1.85	**75.563**±2.06	71.465±1.93

USPS	46.633±1.99	54.216±4.23	54.057±3.82	52.102±3.84	**61.887**±2.45	58.665±3.43	60.465±3.10	58.126±4.56
Iris	86.733±6.87	89.233±0.95	89.335±1.56	70.533±15.18	**90.367**±3.23	89.500±2.67	88.032±2.22	89.867±1.85
semeion	50.433±3.17	45.527±3.50	60.830±3.32	59.561±2.62	73.731±1.80	65.725±2.00	72.718±1.96	**73.944**±3.32
SPF	39.879±2.29	20.649±2.12	41.621±1.93	39.196±2.36	36.906±3.92	43.635±3.57	39.356±2.67	**44.742**±3.76
IS	52.514±4.26	58.131±0.81	55.249±2.72	52.929±3.59	64.143±5.67	58.936±4.13	**66.361**±2.87	65.031±4.45
VS	29.291±4.83	43.363±2.80	45.100±1.14	43.085±5.77	48.292±3.71	44.539±1.09	**49.298**±1.43	45.508±1.52
average	60.98	58.023	60.409	55.625	65.902	62.814	66.291	**66.322**

4.3 Effect of the Parameters

4.3.1 Analysis of Parameter T

The parameter T represents the number of samples in reference sets we take when the reconstructed the mapping coefficient being build. In the experiment, T is derived from the values range from 5 percent to 20 percent of the true number of each category. The results are averaged over 20 runs for each value of T. The mean of the NMI, the ARI and the AC are shown in **Fig. 3**. The horizontal coordinate is the parameter T, and the longitudinal coordinate is three indexes with NMI, ARI and AC. The figures show that, as expected, as the T is increased, the NMI, the ARI and the AC obtained by our ECRMC clustering algorithm keep stable.



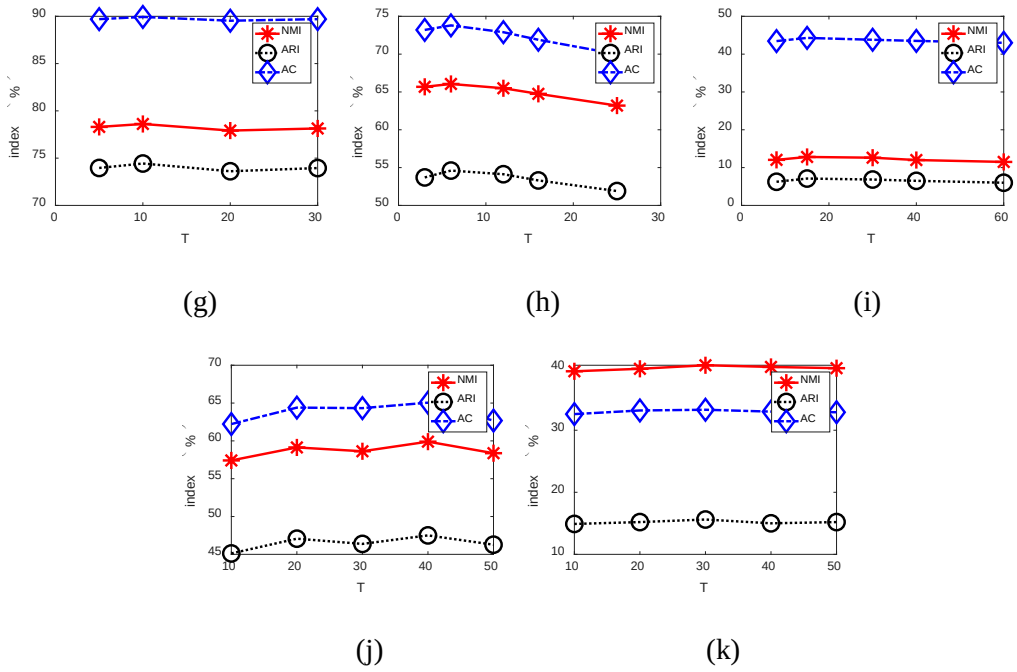


Fig. 3. The mean of the NMI, the ARI and the AC obtained with different T by our ECRMC clustering algorithm for the (a)BC, (b)PD, (c) MNIST, (d) VS, (e)COIL20, (f)USPS, (g)Iris, (h) semeion, (i)SPF, (j)IS, (k)LR.

4.3.2 Analysis of Parameter M

The parameter M denotes the times we do the k-means, and the clustering results with different M obtained by the six ensemble clustering algorithms are shown in **Fig. 4**, **Fig. 5** and **Fig. 6**. We can clearly observe that in the process of M changing from 10 to 50, the scores of the three evaluation indexes fluctuate in a small range. Firstly, we analyze **Fig. 4**, under BC, PD, MNIST, Letters and semeion datasets, both in terms of stability and in terms of scoring, our ensemble clustering is slightly better. Under COIL20, USPS, iris datasets we are behind the PTGP on both counts, and under VS datasets, although we scored poorly, we were the least affected by M. Under SPF dataset, our ensemble clustering basically keeps level with ECPCS-HC. In the case of IS dataset, when M is equal to 10, PTGP algorithm gains an advantage. When M is equal to 20, 30, 40 and 50, our ensemble algorithm has a little advantage. And in **Fig. 5**, with the exception of the VS and Letters dataset, which do not get the best results, the other nine datasets were as good as ever. As shown in **Fig. 6**, our ensemble clustering generally obtains the best performance among the whole ten datasets, and it's basically not affected by M very much. From the overall influence of the ensemble size on the algorithm score comes to see, our algorithm is indeed very little affected.

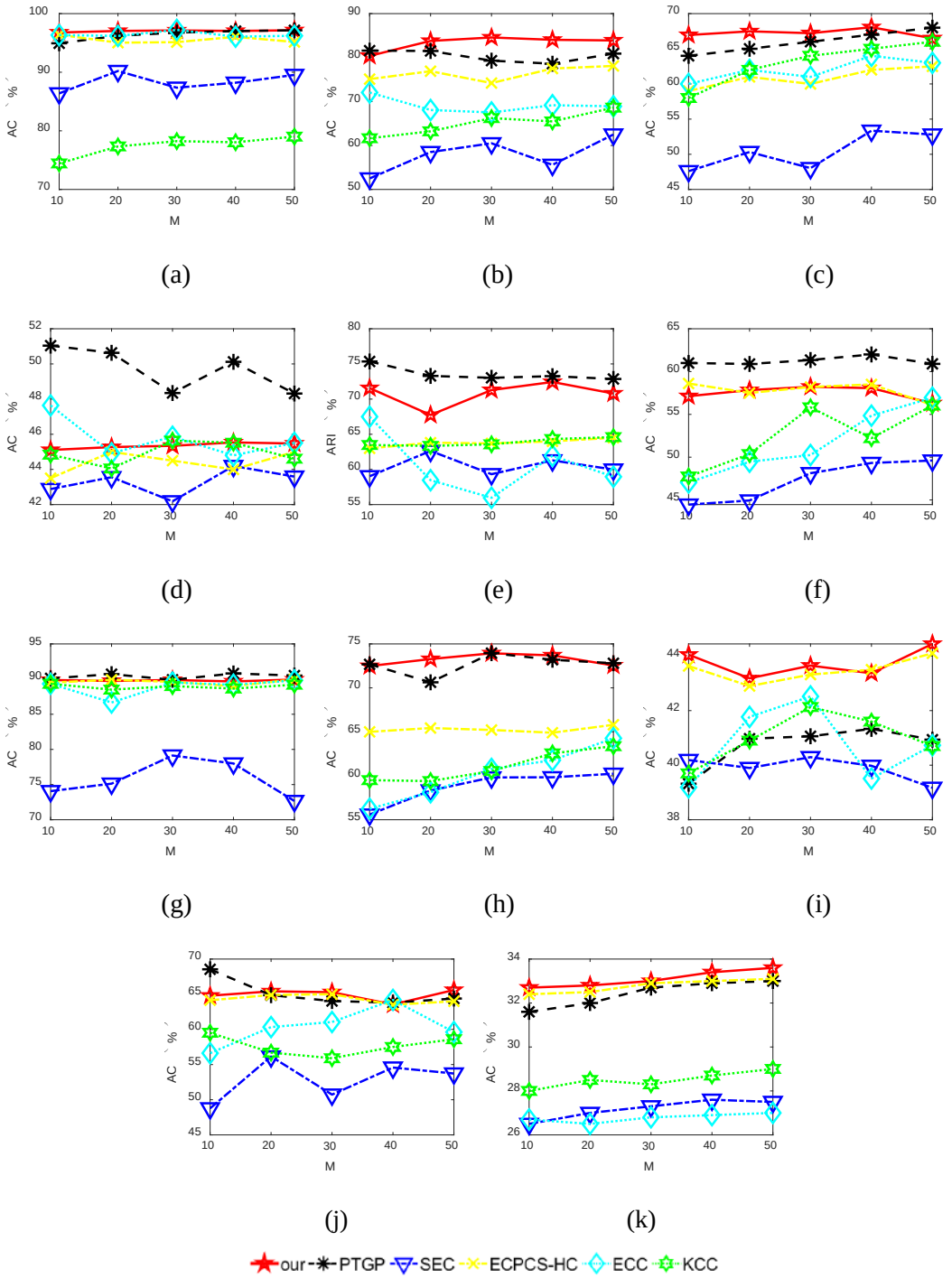


Fig. 4. The mean AC index obtained by the five ensemble algorithms with different M for (a)BC, (b)COIL20, (c) Iris, (d) IS, (e)MNIST, (f)PD, (g)semeion, (h)SPF, (i)USPS, (j)VS and (k)LR.

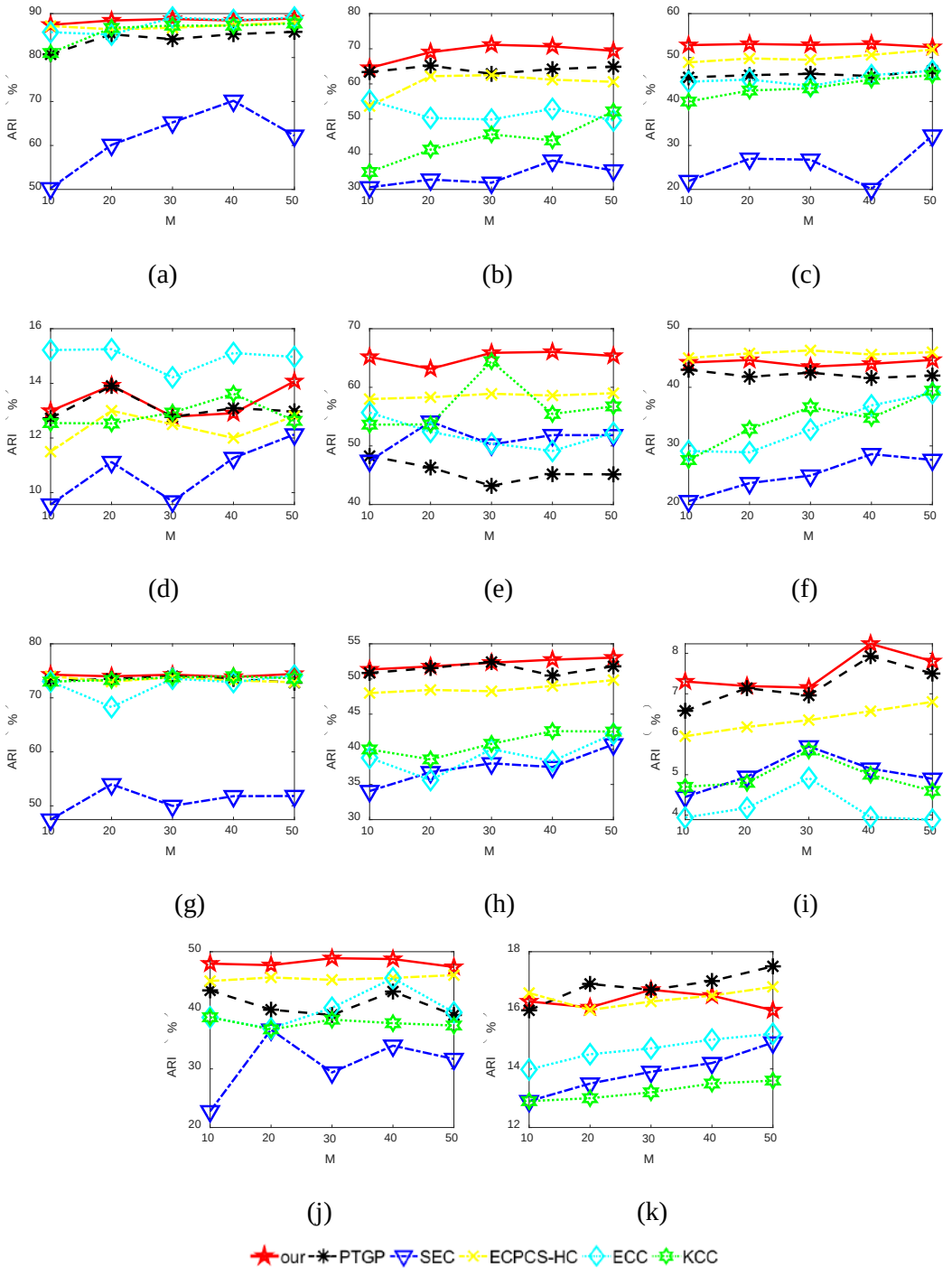


Fig. 5. The mean of the ARI index obtained by the five ensemble algorithms with different M for (a)BC, (b)PD, (c) MNIST, (d) VS, (e)COIL20, (f)USPS, (g)Iris, (h) semeion, (i)SPF, (j)IS and (k)LR.

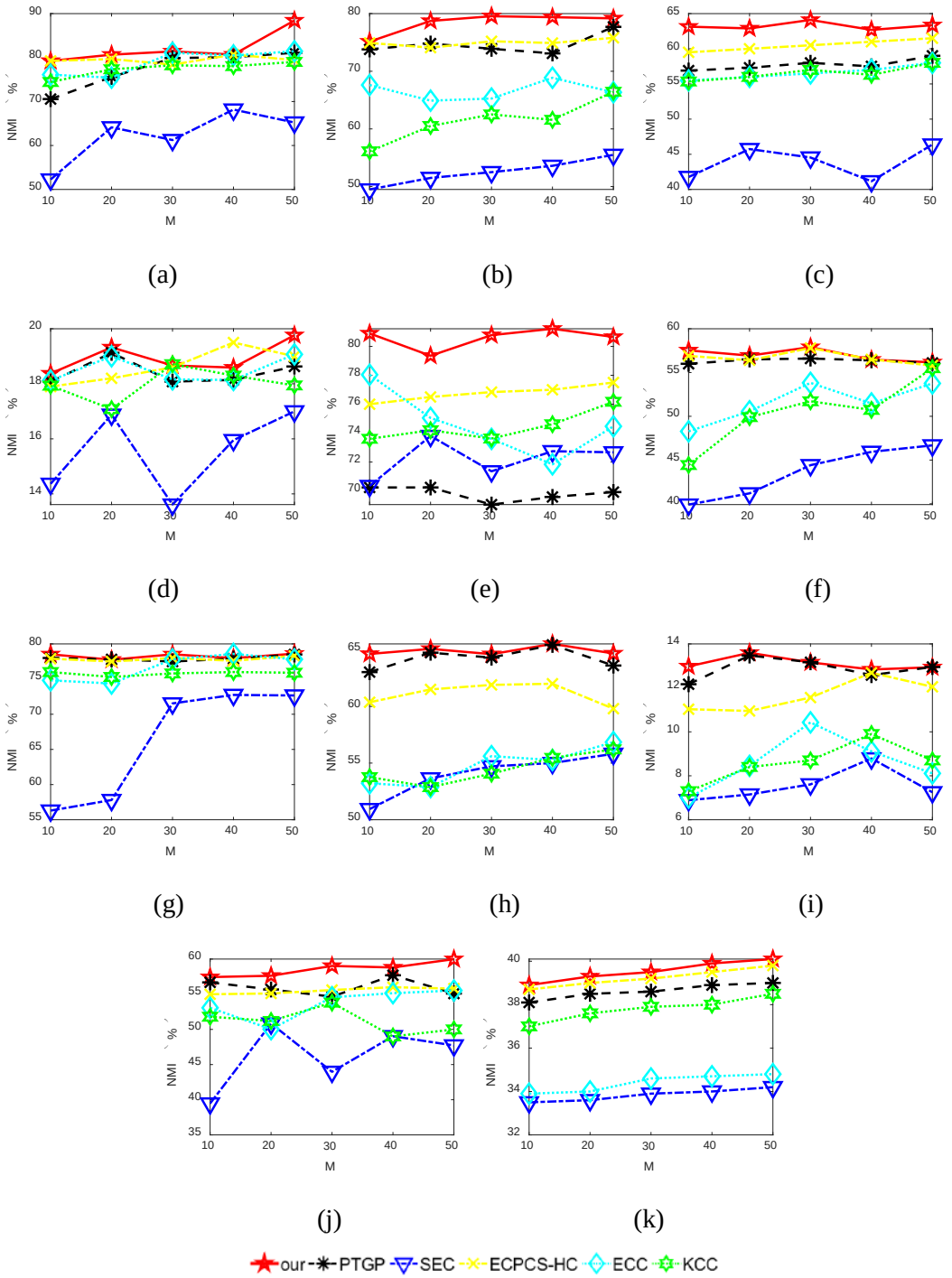
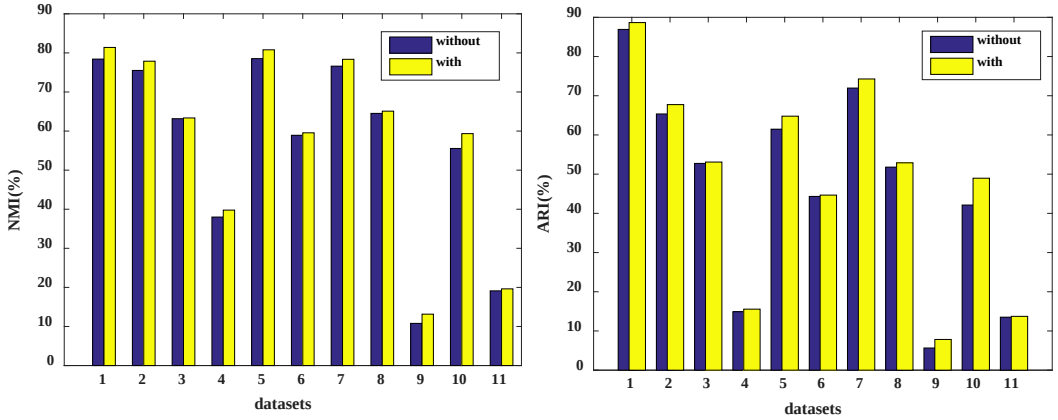


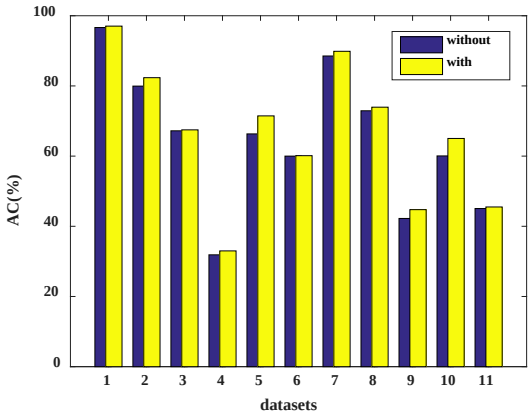
Fig. 6. The mean of the NMI index obtained by the five ensemble algorithms with different M for (a)BC, (b)PD, (c) MNIST, (d) VS, (e)COIL20, (f)USPS, (g)Iris, (h) semeion, (i)SPF, (j)IS, (k)LR.

4.4 The effect of the reconstructed mapping coefficient matrix

In this section, we will test the effect of the proposed reconstructed mapping coefficient matrix. In the experiments, the performance of the algorithm using the reconstructed mapping coefficient matrix and without it are compared. The results which are calculated over 20 runs are shown in the **Fig. 7**. In the figure, the yellow bars indicate the results obtained using the reconstructed mapping coefficient matrix while the blue show the results without the reconstructed mapping coefficient matrix. Generally, as shown in **Fig. 7**, the proposed reconstructed mapping coefficient matrix make the clustering algorithm getting better results.



(a) (b)



(c)

Fig. 7. The NMI, ARI and AC obtained using the reconstructed mapping coefficient matrix and without it: (a) NMI, (b) ARI, (c) AC. Here, the numbers in the horizontal coordinate represent the 11 datasets used in our experiments. And 1 to 11 represent BC, PD, MNIST, LR, coil20, USPS, Iris, semeion, SPF, IS and VS, respectively.

5. Conclusion

In this paper, a new ensemble clustering algorithm using a reconstructed mapping coefficient (ECRMC) is proposed. As we know, there is often a loss of information in the process of information transmission. And the loss of some useful information will have a bad effect on the ensemble clustering. In order to reduce information loss, a reconstructed mapping coefficient matrix is given which makes use of the key information of the neighborhood to add the supplementary information. Moreover, the Spearman's correlation coefficient is used to measure the similarity between micro-cluster sets. It's very friendly to any distribution of data, such as the non-normally distributed. Finally, the co-association matrix is revised and the consensus process is applied to obtain the clustering result. The superiority of the ECRMC clustering algorithm over k-means, KCC, ECC, SEC, PTGP, ECPCS-HC and LWGP algorithm has demonstrated by the experiments. All the experimental results described in this paper have shown that our algorithm is effective.

Acknowledgments

The authors would like to thank the Editor and the anonymous referees for their helpful comments and suggestions to improve the quality of the paper. This paper was supported by the National Natural Science Foundation of China (No. 61532005) and the Fundamental Research Funds for the Central Universities of China (2018JBZ001).

References

[1] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition. Lett.*, vol. 31, no. 8, pp. 651–666, 2010. Article (CrossRef Link)

[2] Zhen-Feng H E, Fan-Lun X, “A Constrained Partition Model and K-Means Algorithm,” *Journal of Software*, pp. 799-809, 2005.

[3] O. Tuzel, F. Porikli, and P. Meer, “Kernel methods for weakly supervised mean shift clustering,” in *Proc. of ICCV, Kyoto, Japan*, pp. 48–55, 2009. Article (CrossRef Link)

[4] R. Collins, “Mean shift blob tracking through scale space,” in *Proc. of CVPR*, vol. 2, pp. 234–240, 2003. Article (CrossRef Link)

[5] Z. Lu and M. A. Carreira-Perpiñán, “Constrained spectral clustering through affinity propagation,” in *Proc. of CVPR, Anchorage, AK, USA*, pp.1-8, 2008. Article (CrossRef Link)

[6] B. Kulis, S. Basu, I. Dhillon, and R. Mooney, “Semi-supervised graph clustering: A kernel approach,” *Machine Learning*, vol. 74, pp. 1–22, 2009. Article (CrossRef Link)

S. Mimaroglu and E. Erdil, “Combining multiple clusterings using similarity graph,” *Pattern Recognition*, vol. 44, no. 3, pp. 694–703, 2011. Article (CrossRef Link)

[7] C.-L. Liu, W.-H. Hsaio, C.-H. Lee, and F.-S. Gou, “Semi-supervised linear discriminant clustering,” *IEEE Transactions on Cybernetics*, vol. 44, no. 7, pp. 989–1000, Jul. 2014. Article (CrossRef Link)

[8] A. Adolfsson, M. Ackerman, N. Brownstein, “To cluster, or not to cluster, An analysis of cluster ability methods,” *Pattern Recognition*, vol 88, pp. 13-26, 2019. Article (CrossRef Link)

[9] A. L. N. Fred and A. K. Jain, “Combining multiple clusterings using evidence accumulation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 835–850, 2005. Article (CrossRef Link)

[10] A. Topchy, A. K. Jain, and W. Punch, “Clustering ensembles: models of consensus and weak partitions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1866–1881, 2005. Article (CrossRef Link)

[11] D. Huang, J. Lai, and C.-D. Wang, “Ensemble clustering using factor graph,” *Pattern Recognition*, vol. 50, pp. 131–142, 2016. Article (CrossRef Link)

[12] H. Liu, M. Shao, S. Li, and Y. Fu, “Infinite ensemble for image clustering,” in *Proc. of 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 1745–1754, 2016. Article (CrossRef Link)

[13] J. Wu, H. Liu, H. Xiong, J. Cao, and J. Chen, “K-means-based consensus clustering: A unified view,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 1, pp. 155–169, 2015. Article (CrossRef Link)

[14] Liu H, Zhao R, Fang H, et al., “Entropy-based consensus clustering for patient stratification,” *Bioinformatics*, vol. 33, no. 17, pp. 2691-2698, 2017. Article (CrossRef Link)

[15] Zhao X, Liang J, Dang C, “Clustering ensemble selection for categorical data based on internal validity indices,” *Pattern Recognition*, vol. 69, pp. 150-168, 2017. Article (CrossRef Link)
L. Bai, J.Y. Liang, “Cluster validity functions for categorical data: a solution-space perspective,” *Data Mining & Knowledge Discovery*, vol. 29, pp. 1560-1597, 2015. Article (CrossRef Link)

[16] D. Huang, C. D. Wang, and J. H. Lai, “Locally weighted ensemble clustering,” *IEEE Transactions on Cybernetics*, vol. 48, no. 5, pp. 1460–1473, 2018. Article (CrossRef Link)

[17] Iam-On N, Boongoen T, Garrett S M, et al, “A Link-Based Approach to the Cluster Ensemble Problem,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2396-2409, 2011. Article (CrossRef Link)

[18] Yang Y, Jiang J, “Bi-weighted ensemble via HMM-based approaches for temporal data clustering,” *Pattern Recognition*, vol. 76, pp. 391-403, 2018. Article (CrossRef Link)

[19] Liu H, Wu J, Liu T, et al, “Spectral Ensemble Clustering via Weighted K-Means: Theoretical and Practical Evidence,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1129-1143, 2017. Article (CrossRef Link)

[20] Huang, Dong, J. H. Lai, and C. D. Wang. "Robust Ensemble Clustering Using Probability Trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 5, pp. 1312–1326, 2016. Article (CrossRef Link)

[21] Huang D, Wang C D, Peng H, et al, “Enhanced Ensemble Clustering via Fast Propagation of Cluster-Wise Similarities,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1-13, 2018. Article (CrossRef Link)

[22] J. Cao, P. Chen, B. W. Ling, Z. Yang and Q. Dai, “Spectral Clustering with Sparse Graph Construction Based on Markov Random Walk,” *KSII Transactions on Internet and Information Systems*, vol. 9, no. 7, pp. 2568-2584, 2015. Article (CrossRef Link)

[23] Myers, Jerome L. Well, Arnold D., *Research Design and Statistical Analysis 2nd*, Lawrence Erlbaum: 508, 2003, ISBN 0-8058-4037-0.

[24] Z. Li and J. Tang, "Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5343-5355, 2015. Article (CrossRef Link)

[25] Li Z, Tang J, He X, “Robust Structured Nonnegative Matrix Factorization for Image Representation,” *IEEE Transactions on Neural Networks & Learning Systems*, vol. 29, no. 5, pp. 1947-1660, 2018. Article (CrossRef Link)

[26] Zechao L, Jinhui T, Tao M, “Deep Collaborative Embedding for Social Image Understanding,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 41, no. 9, pp. 2070-2083, 2018. Article (CrossRef Link)

[27] Caiming Zhong, Lianyu Hu, Xiaodong Yue, Ting Luo, Qiang Fu, Haiyong Xu, “Ensemble clustering based on evidence extracted from the co-association matrix,” *Pattern Recognition*, Vol. 92, pp. 93-106, 2019. Article (CrossRef Link)

[28] Sarvari H, Domeniconi C, Stilo G, “Graph-based selective outlier ensembles,” in *Proc. of the 34th ACM/SIGAPP Symposium. ACM*, pp. 518-525, 2019. Article (CrossRef Link)

[29] Shannon C E, “A mathematical theory of communication,” *Bell Labs Technical Journal*, vol. 27, no. 4, pp. 623-656, 1948. Article (CrossRef Link)

[30] Seifoddini H K, “Single linkage versus average linkage clustering in machine cells formation applications,” *Computers & Industrial Engineering*, vol. 16, no. 3, pp. 419-426, 1989. Article (CrossRef Link)

[31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. Article (CrossRef Link)

[32] K. Bache and M. Lichman, “UCI machine learning repository,” 2017. [Online]. Available: Article (CrossRef Link)

[33] A. Strehl and J. Ghosh, “Cluster ensembles: A knowledge reuse framework for combining multiple partitions,” *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2003. Article (CrossRef Link)

[34] N. X. Vinh, J. Epps, and J. Bailey, “Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance necessary?,” in *Proc. of ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1073-1080, 2009. Article (CrossRef Link)

Tuoqia Cao is currently pursuing a master degree at Beijing Jiaotong University of Computer and Information Technology. Her current research interests include clustering algorithm.

Dongxia Chang received the B.S. and M.S. degree in applied mathematic from Xidian University, Xi'an, China, in July 2000 and April 2003, and the PhD degree from Tsinghua University, Beijing, China, in July 2009, respectively. From May 2010 to May 2012, she was a post-doctoral research at Beijing Jiaotong University, Beijing, China. Since May 2012, she has been an associate professor at the Institute of Information Science, Beijing Jiaotong University. Her current research interests include pattern recognition, clustering analysis, image analysis.

Yao Zhao received the BS degree from Fuzhou University, China, in 1989, and the ME degree from Southeast University, Nanjing, China, in 1992, both from the Radio Engineering Department, and the PhD degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), China, in 1996. He became an associate professor at BJTU in 1998 and became a professor in 2001. From 2001 to 2002, he was a senior research fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. He is currently the director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, and video analysis and understanding. He serves on the editorial boards of several international journals, including as associate editors of IEEE Transactions on Cybernetics, IEEE Signal Processing Letters, and an area editor of Signal Processing: Image Communication (Elsevier), etc. He was named a distinguished young scholar by the National Science Foundation of China in 2010, and was elected as a Chang Jiang Scholar of Ministry of Education of China in 2013. He is a senior member of the IEEE.