IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 28, 2019, accepted July 14, 2019, date of publication July 19, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929948

An Improvement of Spectral Clustering via Message Passing and Density Sensitive Similarity

LIJUAN WANG[1,3], SHIFEI DING[1,2], AND HONGJIE JIA[1,4]

[1]School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China
[2]Mine Digitization Engineering Research Center, Ministry of Education of the People's Republic of China, Xuzhou 221116, China
[3]School of Information and Electrical Engineering, Xuzhou College of Industrial Technology, Xuzhou 221400, China
[4]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China

Corresponding author: Shifei Ding (dingsf@cumt.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2017XKZD03.

ABSTRACT Spectral clustering transforms the data clustering problem into a graph-partitioning problem and classifies data points by finding the optimal sub-graphs. Traditional spectral clustering algorithms use Gaussian kernel function to construct the similarity matrix, so they are sensitive to the selection of scale parameter. In addition, they need to randomly determine the initial cluster centers at the clustering stage and the clustering performance is not stable. Therefore, this paper presents an algorithm on the basis of message passing, which makes use of a density adaptive similarity measure, describing the relations between data points and obtaining high-quality cluster centers through message passing mechanism in AP clustering. The performance of clustering is optimized by this method. The experiments show that the proposed algorithm can effectively deal with the clustering problem of multi-scale datasets. Moreover, its clustering performance is very stable, and the clustering quality is better than traditional spectral clustering algorithm and k-means algorithm.

INDEX TERMS Spectral clustering, similarity matrix, message passing, clustering stability.

I. INTRODUCTION

The purpose of clustering is divided data points into different clusters according to their similarity, which makes the similarity of data in the same cluster larger and the similarity of data between different clusters smaller. Traditional clustering methods, such as k-means algorithm and FCM algorithm, lack the ability to deal with intractable complex data structures. When the sample space is not convex, these algorithms ease to fall into local optimum [1].

In recent years, spectral clustering has attracted the attention in academia because of its good performance and simple implementation [2]. Also, Spectral clustering can cluster in arbitrary shape sample space and converge to global optimum, especially for non-convex data sets [3]. Spectral clustering algorithm views each point in the data set as the vertex of the graph, and the similarity between any two points as the weight of edges connecting the two vertices. Thus, an undirected weighted graph is constructed. Then, according to a graph partitioning method, the graph is divided into several disconnected subgraphs, and the point set contained in the subgraph is the cluster generated after clustering [4].

There are many traditional graph partitioning methods, such as MinCut, RatioCut, the normalized cut Ncut, and Min-Max cut [5]. The clustering results are obtained by maximizing or minimizing the objective function of graph cut method. Nevertheless, for various graph-partitioning methods, the optimal solution of objective function is often an NP-hard. According to Rayleigh-Ritz theory, the discrete optimization problem of the original one can be solved in polynomial time by relaxing to the real number domain [6]. For graph partitioning, a point may be belong to subset A and the other part belong to subset B, rather than either or the other. Generally, good clustering results can be obtained by making full use of the classification information of the eigenvalues and the vertices contained in the Laplacian matrix of a graph in the clustering process [7]. The method is based

The associate editor coordinating the review of this manuscript and approving it for publication was Kashif Munir.

on the matrix spectral analysis theory to cluster, so it was called spectral clustering. At present, spectral clustering has been widely used in computer vision, data analysis, image processing, video surveillance, automatic control and other fields [8]–[10].

Although the spectral clustering algorithm has achieved good performance in practice, as a new clustering method, spectral clustering is still in the stage of development. There are still a great deal of problems to be further studied. For example, traditional spectral clustering is sensitive to initial values and cannot effectively deal with multi-scale clustering problems [11]. In order to dispose with multi-scale data sets, Zelnik-Manor *et al.* [12] proposed an adaptive spectral clustering technique. Instead of selecting a global parameter σ, it calculates an adaptive parameter σ_{i} for each point x_{i} based on its neighborhoods information, where σ_{i} is the Euclidean distance from point x_{i} to its p'th nearest neighbors. The similarity measure is called an adaptive Gauss kernel function. Considering the data distribution of each point's neighborhoods, self-adjusting spectral clustering can effectively separate the compact clusters contained in the sparse background cluster. Liu *et al.* [11] obtains the implicit cluster structure features in data by local density, and then combines with self-adjusting Gauss kernel function, proposes a spectral clustering algorithm based on shared neighborhoods adaptive similarity. Tao *et al.* [13] puts forward a similarity calculation method between points of manifold structure data to improve the clustering performance of the algorithm. Because of the similarity between data points based on Euclidean distance, the distribution of complex manifold data sets cannot be reflected correctly [14]. In order to solve the problem of traditional spectral clustering algorithm, which is sensitive to scale parameters, and clustering center initialization, we propose a Spectral Clustering Algorithm Based on Message Passing (MPSC) and use density sensitive similarity to measure the similarity of data points. The rest of the paper is organized as follows: Section 2 the ''message passing'' mechanism of AP clustering is introduced to solve the sensitive initialization problem of traditional spectral clustering algorithm. Section 3 introduces the basic theory of AP Clustering algorithm. Section 4 presents the MPSC algorithm and gives its detail steps. Section 5 compares the proposed algorithm with other algorithms in the experiments. and the last part is a conclusion.

II. INITIALIZATION SENSITIVITY ANALYSIS OF SPECTRAL CLUSTERING ALGORITHM

Spectral clustering algorithm transforms the data clustering problem into a graph optimal segmentation problem. The objective function of minimizing or maximizing graph cut method is often an NP discrete optimization problem. Fortunately, spectral method can provide a relaxed solution in polynomial time to this optimization problem [15]. The ''relaxation'' here refers to relaxing the discrete optimization problem to real number domain, and then using some heuristics. The essence of Graph Segmentation can be reduced to the problem of minimizing or maximizing the trace of a matrix, and the task of minimizing or maximizing the trace depends on spectral clustering algorithm.

Generally, spectral clustering algorithm consists of three parts: Reprocessing, Spectral representation and Clustering [16]. Firstly, construct the similarity matrix and Laplacian matrix, then, select eigenvectors and determine the number of clusters.

There are two basic methods for partitioning data sets or graphs: recursive 2-way partitioning and k-way partitioning [17]. The recursive 2-way partitioning method calls the 2-way partitioning algorithm recursively in a hierarchical manner. When the graph is divided into two parts, the same process is applied to the subgraph until the number of clusters meets the requirements or no longer meets the recursive conditions; K-way partition picks out the eigenvectors of Laplacian matrix with implicit clustering information, and uses these eigenvectors to partition the data set into k-way partitions. A typical spectral clustering algorithm for k-way partition is the NJW algorithm [18] proposed by Ng, Jordan and Weiss. The following is Wang *et al.* [19] analysis of the sensitivity of initial values in spectral clustering algorithm.

Proposition 1. Given a dataset $\mathrm{X} = \{x_1, x_2, \ldots..x_n\}$, the similarity matrices obtained by input in different order are $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, $\boldsymbol{D}_1$, $\boldsymbol{D}_2$ are the corresponded diagonal matrices, Laplacian matrices $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, and generating matrices $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$. Then $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are similar, $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are similar, l_1 and l_2 are similar, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are similar.

Proof 1: Assume the dataset is input in sequence of x_1, $x_2 \ldots x_{\mathrm{n}}$ to get the matrix $\boldsymbol{W}_1$. Assume the data set is input in order of $x_n, x_{\mathrm{n}-1} \ldots\ldots x_1$ to get the matrix $\boldsymbol{W}_{2.}$, $\boldsymbol{W}_2$ can be obtained by several elementary transformations of matrix $\boldsymbol{W}_1$, which proves that matrix $\boldsymbol{W}_1$ is similar to matrix $\boldsymbol{W}_2$. Similarly, the matrix $\boldsymbol{W}_2$ obtained by any order of input is similar to $\boldsymbol{W}_1$. Since $\boldsymbol{D}_1$ is a diagonal matrix and its principal diagonal elements are the sum of the corresponding elements of the similar matrix$\boldsymbol{W}_1$, the matrices $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are similar. Because $\boldsymbol{L}_1 = \boldsymbol{D}_1^{-1/2}\boldsymbol{W}_1\boldsymbol{D}_1^{1/2}$, and $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are diagonal matrices, the matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are similar, $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are similar. The corresponding eigenvector $\boldsymbol{X}_2$ is obtained by several elementary transformations of eigenvector $\boldsymbol{X}_1$. It can be proved that the matrix $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are similar. Therefore, the conclusion was proved.

Proposition 1 shows that similarity matrix W and matrix Y are similar, so spectral clustering algorithm is sensitive to the initialization of clustering centers. The essence is that k-means algorithm is used in the last stage of clustering process. To solve this problem, we introduce the ''message passing'' mechanism in AP clustering into spectral clustering, which is used to determine the clustering center to improve the performance of spectral clustering algorithm.

III. BASIC THEORY OF AP CLUSTERING ALGORITHM

Affinity Propagation (AP) was proposed by Frey and Dueek in Science in 2007 [20]. AP algorithm does not need to

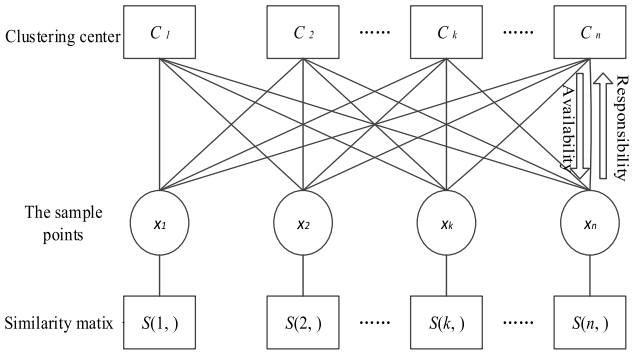


FIGURE 1. Description of the clustering process of AP algorithms.

specify the initial clustering center. Experiments show that AP algorithm has high efficiency, what's more, it takes the similarities between pairs of data points as its input. The similarity matrix, denoted as S (n × n), where the similarity measure s (i, j) = $-//x_i - x_j//^2$ is calculated by the Euclidean distance between pair points, is the fundamental to AP. Responsibility is denoted as : $\boldsymbol{R(i,k)} = \boldsymbol{S(i,k)} - \max\{\boldsymbol{A(i,j)} + \boldsymbol{S(i,j)}\}$ (j ≠ k); Availability is expressed as:

$$\boldsymbol{A(i,k)} = \min\{0, \boldsymbol{R(k,k)} + \sum_{j=1}^{n} \max\{0, \boldsymbol{R(j,k)}\}\} \quad (j \neq i且 j \neq k)$$

are two key factors of AP. The AP algorithm searches for clusters through several iterative processes until a high-quality set of exemplars and corresponding clusters are founded [21], [22]. In order to overcome the problems of oscillation and convergence failure, Meng *et al.* [23] introduced a damping factor (damp) into iterations. The iterative formulas are respectively described as:

$$\boldsymbol{R_T} = \boldsymbol{R_T} \times (1 - \boldsymbol{damp}) + \boldsymbol{R_{T-1}} \times \boldsymbol{damp} \quad (1)$$

$$\boldsymbol{A_T} = \boldsymbol{A_T} \times (1 - \boldsymbol{damp}) + \boldsymbol{A_{T-1}} \times \boldsymbol{damp} \quad (2)$$

where R_{T-1} and A_{T-1} indicate the T-1th iteration's responsibility and availability, damp ϵ [0.5, 1). The clustering process of AP algorithm is shown in Figure 1.

IV. THE PROPOSED SPECTRAL CLUSTERING BASED ON MESSAGE PASSING APPROACH

The key to spectral clustering is the selection of distance measure, which will be smooth with regard to the intrinsic structure of data points. Data in the same cluster should have higher similarity and follow the space consensus. Therefore, the construction of similarity matrix is crucial, which greatly affects the performance of the spectral clustering [24]. While dealing with the complicated datasets, however, the scale parameter $\boldsymbol{\sigma}$ of the Gauss kernel function is fixed, and the similarity simply based on Euclidean distance cannot truly reflect the data distribution and in turn resulting in the poor performance of spectral clustering [25]. Yang *et al.* [26] proposed a density-sensitive similarity measurement, which is able to deal with multi-scale clustering problems and is relatively insensitive to parameter selection. Experiments show that the algorithm can effectively describe the actual clustering distribution.

Defined the adjustable line segment length, which can adjust the distance in regions with different densities:

$$\boldsymbol{L(x_i, x_j)} = (e^{\rho \boldsymbol{d}(x_i, x_j)} - 1)^{1/\rho} \quad (3)$$

where $\boldsymbol{d(x_i, x_j)}$ can be the Euclidean distance between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$, and $\boldsymbol{\rho}$ is the density factor with the value greater than 1.

Density sensitive distance can be defined as follows:

$$\boldsymbol{D_{ij}} = \min \sum_{k=1}^{l-1} \boldsymbol{L(p_k, p_{k+1})}, \quad \boldsymbol{p \in P_{ij}} \quad (4)$$

Algorithm 1 Spectral Clustering Using Message Passing and Density Sensitive Similarity (MPSC)

Input: data set $\boldsymbol{X} = \{\boldsymbol{x_i} | \boldsymbol{i} = 1, \ldots\ldots \boldsymbol{n}\}$, cluster number k
Output: k clusters
Step 1. Calculate the density sensitive distance $\boldsymbol{D_{ij}}$ between each data pair $\boldsymbol{(x_i, x_j)}$ according to Eq. (4);
Step 2. Calculate the similarity $\boldsymbol{w_{ij}}$ between pairwise points $\boldsymbol{(x_i, x_j)}$ by Eq. (5), and construct the similarity matrix $\boldsymbol{W}$ $\boldsymbol{W} \in \boldsymbol{R^{n \times n}}$;
Step 3. Construct the degree matrix of graph D, where D is a diagonal matrix whose $\boldsymbol{(i, i)}$-element is the sum of W's ith row, and the off-diagonal element is 0;
Step 4. Compute the normalized graph Laplacian $\boldsymbol{L}$:

$$\boldsymbol{L} = \boldsymbol{D^{-1/2} W D^{-1/2}};$$

Step 5. Compute the first k eigenvectors $\boldsymbol{u}_{1,\ldots\ldots}, \boldsymbol{u_k}$ of $\boldsymbol{L}$, Let $\boldsymbol{U} = \left[\boldsymbol{u}_1 \vdots \cdots \vdots \boldsymbol{u_k}\right] \in \boldsymbol{R^{n \times k}}$ be the matrix containing the $\boldsymbol{u}_{1,\ldots\ldots}, \boldsymbol{u_k}$ as columns;
Step 6. Normalize each row of the matrix U to unit length and form the matrix $\boldsymbol{Y} \in \boldsymbol{R^{n \times k}}$, where $\boldsymbol{y_{ij}} = \boldsymbol{u_{ij}} / \left[\sum_{j=1}^{n} \boldsymbol{u_{ij}^2}\right]^{1/2}$
Step 7. Treat each row of Y as a point and partition them into k clusters using AP clustering;
Step 8. Assign the original point $\boldsymbol{x_i}$ to cluster $\boldsymbol{j}$ if and only if row $\boldsymbol{i}$ of the matrix $\boldsymbol{Y}$ is assigned to cluster $\boldsymbol{j}$.

Let P = $\{\boldsymbol{p}_1, \boldsymbol{p}_2 \ldots\ldots \boldsymbol{p_l}\}$ denote as the path from $\boldsymbol{p}_1$ to $\boldsymbol{p_l}$ with length l = | P |, while $\boldsymbol{P_{ij}}$ denote that the set of paths connecting with $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$.

Density sensitive similarity measure can be defined as:

$$\boldsymbol{w_{ij}} = \frac{1}{\boldsymbol{D_{ij}} + 1} \quad (5)$$

Compared with Gauss kernel, this similarity measure does not require to introduce kernels and directly calculate the similarity in distance measure. Density sensitive distance indicates that the similarity of data pairs in the same manifold structure

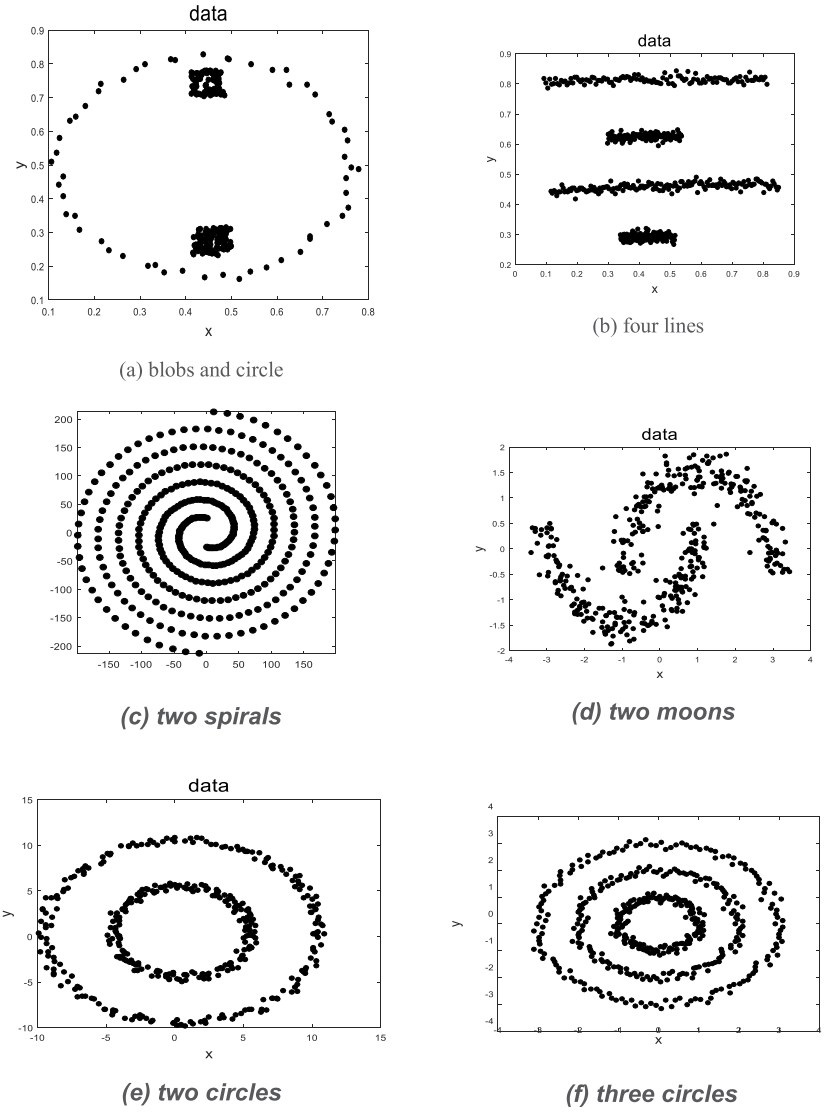


FIGURE 2. Original synthetic datasets.

is higher, while the similarity of data pairs in different manifold structures is lower. The manifold similarity can enlarge the distance between two points on different manifolds and reduce the distance between two points on the same manifold. The distance measure is dependent on the data and adjustable with the local density, i.e., density sensitive.

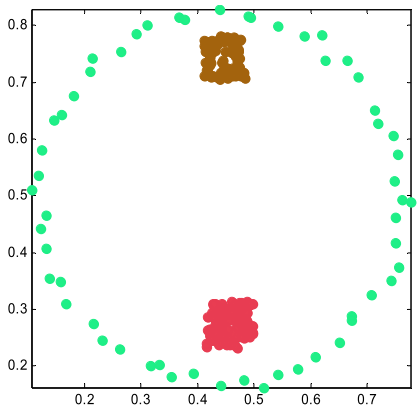
We use the density sensitive distance to improve the accuracy of the spectral clustering algorithm, and propose a Spectral Clustering Algorithm Based on Message Passing (MPSC) algorithm. This algorithm constructs the similarity matrix with the density sensitive distance. Then the objective function is optimized iteratively through message passing mechanism. The detailed steps of MPSC algorithm is shown in Algorithm 1.

For the purpose of overcoming the influence of randomly selected parameters in the traditional algorithm, a message passing mechanism is introduced to optimize the initialization sensitivity in the traditional spectral clustering algorithm, to obtain more stable clustering results.

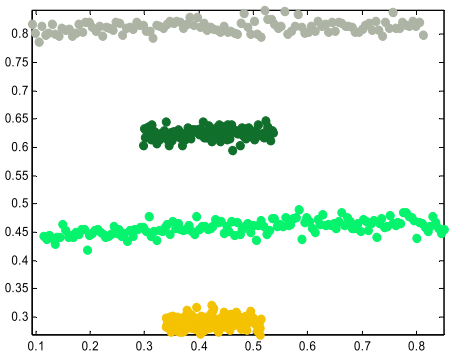
V. EXPERIMENTAL ANALYSIS

A. CLUSTERING ON SYNTHETIC DATASETS

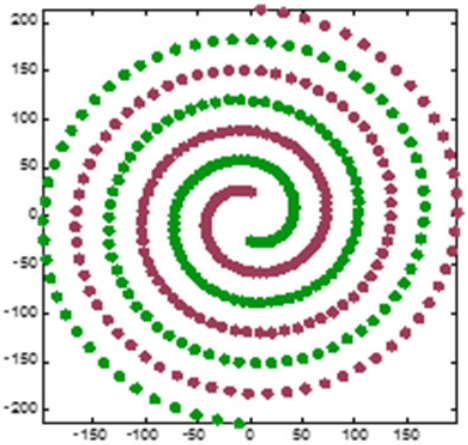
The clustering performance of the proposed MPSC algorithm is compared with K-means algorithm, A K-AP Clustering Algorithm Based on Manifold Similarity Measure (MKAP) [14], NJW algorithm [18] and AP algorithm [20]. All the experiments are conducted on the computer with Intel core i5-8250U processor, 8 GB RAM. The programming environment is MATLAB 2016b. We demonstrate the effectiveness of the improved MPSC algorithm with two prescreening strategies respectively on six synthetic manifold datasets blobs and circle, four lines, two moons, two circles, two spirals and three circles [27], which are illustrated in Fig. 2.



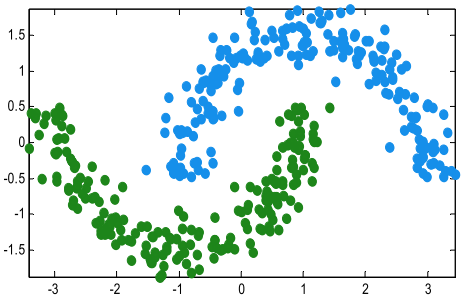
(a) Clustering results of MPSC algorithm on blobs and circle



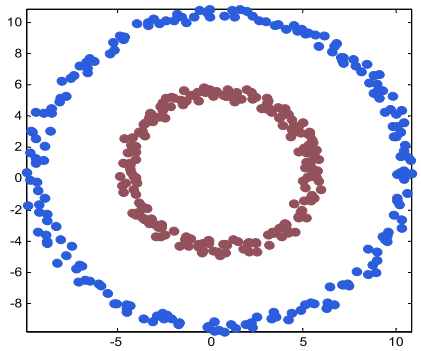
(b) Clustering results of MPSC algorithm on four lines



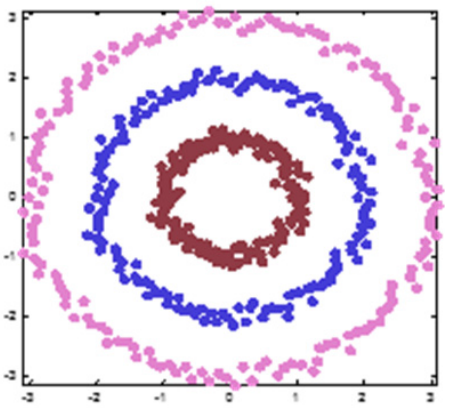
(c) Clustering results of MPSC algorithm on two spirals



(d) Clustering results of MPSC algorithm on two moons



(e) Clustering results of MPSC algorithm on two circles

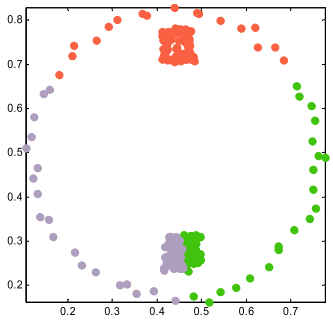


(f) Clustering results of MPSC algorithm on three circles

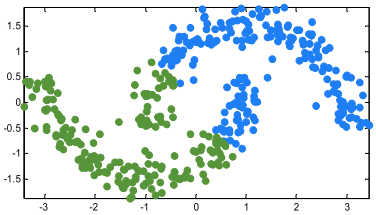
FIGURE 3. **Clustering results of MPSC algorithm on artificial data sets.**

In the experiments, the preference parameter p of AP algorithm is the median of affinity matrix, the maximum iteration is 1000, and the convergence coefficient of iteration is100, scaling factor $\rho = 2$. The clustering results of MPSC algorithm on these six synthetic data sets are presented in Fig. 3.

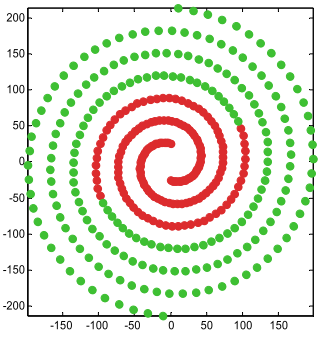
As what we can see from Figure 3, the MPSC algorithm has good performance when dealing with various manifold data.



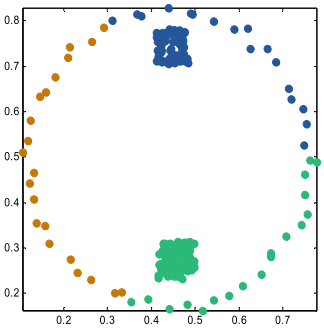
(a) Clustering results of NJW algorithm on blobs and circle (σ=0.1)



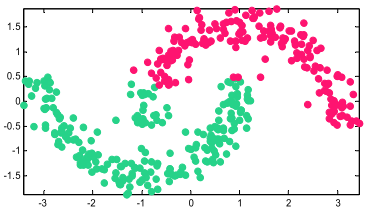
(b) Clustering results of NJW algorithm on two moons (σ=0.1)



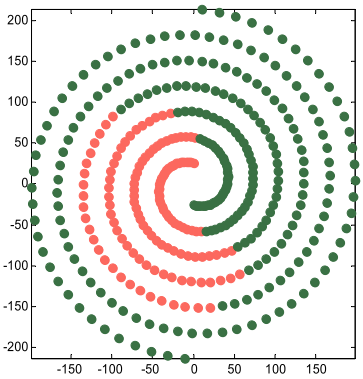
(c) Clustering results of NJW algorithm on two spirals (σ=0.1)



(d) Clustering results of NJW algorithm on blobs and circle (σ=0.2)



(e) Clustering results of NJW algorithm on two moons (σ=0.2)



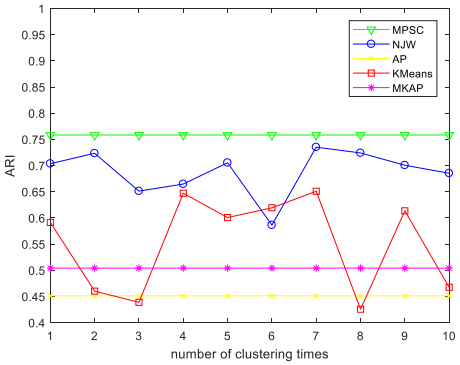
(f) Clustering results of NJW algorithm on two spirals (σ=0.2)

FIGURE 4. **Clustering results of NJW algorithm on artificial data set.**

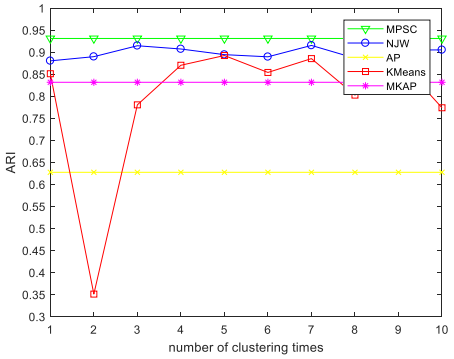
With the help of density sensitive similarity measurement, MPSC algorithm is suitable for the clustering problem on complex datasets. The algorithm uses density-sensitive similarity measure and Laplace transforms to map the original data points into spectral space, which makes the data points in the same class more compact and the data points between different classes more separate. With the NJW algorithm that Wang *et al.* [19] proposed in 2002, Figure. 4 shows the clustering results of NJW algorithm ($\sigma = 0.1$ and $\sigma = 0.2$) on blobs and circles, two moons and two spirals artificial datasets. It shows that NJW algorithm is sensitive to the value of scale parameter σ, and it cannot obtain the accurate clustering results on complex datasets. On the contrary, MPSC algorithm can maintain the global consistency of data in the clustering process, thus overcoming the shortcoming that Gauss kernel function cannot process multi-scale datasets.

B. CLUSTERING ON UCI DATASETS

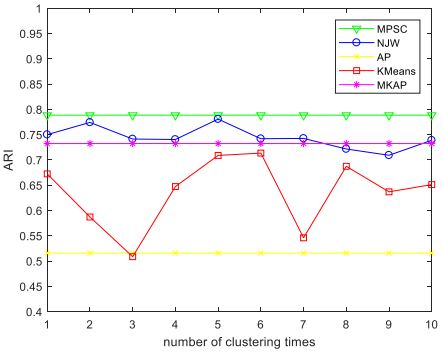
At the present time, to further test the effectiveness of the proposed MPSC algorithm, we conduct a number of experiments with other popular clustering algorithms on more complex datasets from UCI machine learning repository. Table 1 indicates the characteristics of these datasets.



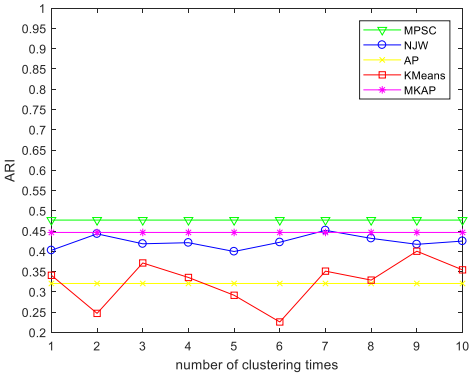
(a) Experimental results of four algorithms on Iris



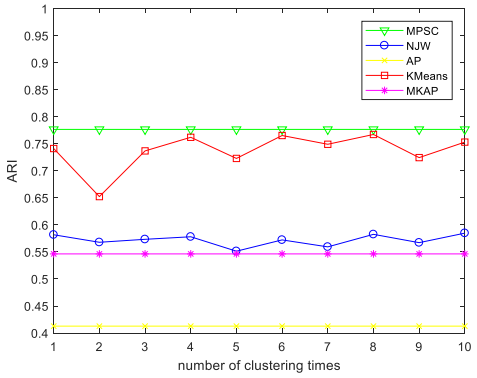
(b) Experimental results of four algorithms on Wine



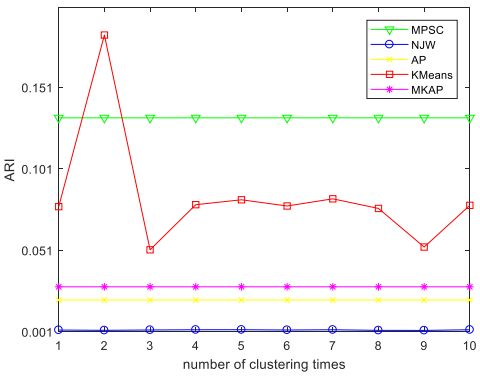
(c) Experimental results of four algorithms on Zoo



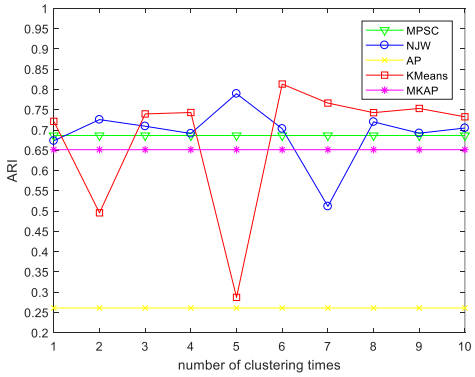
(d) Experimental results of four algorithms on Ionosphere



(e) Experimental results of four algorithms on Glass



(f) Experimental results of four algorithms on Sonar



(g) Experimental results of four algorithms on Seeds

FIGURE 5. **Experimental results of four algorithms on real data sets.**

TABLE 1. Information of real world datasets.

Data set	Number of objects	Number of attributes	Number of classes
Ionosphere	351	34	2
Iris	150	4	3
Wine	178	13	3
Zoo	101	16	7
Sonar	208	60	2
Glass	214	10	6
Seeds	210	6	3

In the experiments, we use three indicators to assess the clustering performance: the running time of algorithm, adjusted rand index (ARI) and clustering accuracy. Comparing the clustering results with the actual partition, there are four possibilities for each pair of data points in the data set.

1) SS: pairwise data belonging to the same class are also divided into the same class when clustering;

2) SD: two data points belonging to the same cluster are divided into different classes;

3) DS: two data points, which do not belong to the same class, are divided into the same class ;

4) DD: two data points that do not belong to the same class are also clustered into different classes.

ARI is based on the relationship of pairwise data points. It can describe the quality of clustering and objectively reflect the advantages and disadvantages of clustering algorithm. It is a common clustering evaluation criterion [28].

Set SS = a, SD = b, DS = c, DD = d, and the calculation equation of ARI is:

$$\boldsymbol{ARI} = \frac{2(\boldsymbol{ad} - \boldsymbol{bc})}{(\boldsymbol{a}+\boldsymbol{b})(\boldsymbol{b}+\boldsymbol{d}) + (\boldsymbol{a}+\boldsymbol{c})(\boldsymbol{c}+\boldsymbol{d})} \tag{6}$$

where ARI ϵ [0, 1], and the smaller the value of ARI is, the worse the clustering quality will be.

The clustering performance of the proposed MPSC algorithm is made a comparison with K-means algorithm, AP algorithm, MKAP algorithm and NJW algorithm. The clustering results of different algorithms are given in Table 2.

ARI index is used as a measure to compare the clustering performance of MPSC algorithm, NJW algorithm, AP algorithm, MKAP algorithm and K-means algorithm. The experimental results are shown in Fig. 5, where the horizontal axis represents the number of clustering while the vertical one depicts the ARI index of the algorithm.

From the experimental data in Figure 5, the curves of K-means algorithm and NJW algorithm fluctuate obviously. The fluctuation of the curves indicates that K-means and NJW algorithm are sensitive to the initialization of clustering centers. The K_means algorithm and NJW algorithm need to randomly select the initial cluster centers, therefore, the clustering results will be poor if the selection of the cluster centers are not appropriate. In contrast, the curves of AP algorithm and MPSC algorithm are greatly smooth, which shows that the performance of AP algorithm, MKAP algorithm and MPSC algorithm is very stable, but the performance of AP and MKAP algorithm is not satisfactory nonetheless. The MPSC algorithm proposed in this paper introduces the "message passing" mechanism of AP algorithm to determine the clustering center. It effectively solves the problem that traditional spectral clustering algorithm is sensitive to the initialization of clustering center. In addition, as what can be seen from Figure 5, the ARI of MPSC algorithm is also significantly better than the other four algorithms. Table 2 gives the average accuracy, average time and ARI index of 10 times clustering of MPSC, NJW, AP, MKAP and K-means algorithms.

TABLE 2. Clustering results of different algorithms on real world datasets.

Dataset	Evaluation index	K-means	AP	NJW	MKAP	MPSC
iris	Accuracy	0.7273	0.6817	0.8534	0.7016	**0.9067**
	Time(s)	0.352	0.582	0.516	0.614	0.588
	ARI index	0.5516	0.4513	0.6850	0.5042	**0.7583**
zoo	Accuracy	0.6534	0.7327	0.6337	0.7528	**0.8119**
	Time(s)	0.353	0.552	0.503	0.6133	0.284
	ARI index	0.6359	0.5158	0.7441	0.7324	**0.7758**
wine	Accuracy	0.4730	0.3652	0.4267	0.4018	**0.5505**
	Time(s)	0.359	0.43	0.484	0.8437	0.402
	ARI index	0.7943	0.6274	0.8986	0.8316	**0.9310**
ionosphere	Accuracy	0.3589	0.3817	0.6182	0.5843	**0.7094**
	Time(s)	1.713	1.049	1.011	1.6541	0.892
	ARI index	0.324	0.3208	0.4237	0.4467	**0.4772**
glass	Accuracy	0.7913	0.4859	0.6542	0.6492	**0.8131**
	Time(s)	0.223	0.438	0.433	0.8215	0.554
	ARI index	0.7375	0.4128	0.5718	0.5464	**0.7767**
sonar	Accuracy	0.3942	0.1201	0.3701	0.3015	**0.4346**
	Time(s)	0.706	0.566	0.595	1.126	0.418
	ARI index	0.0827	0.0206	0.0022	0.0287	**0.1324**
seeds	Accuracy	0.7008	0.3190	**0.7905**	0.6904	0.7194
	Time(s)	0.536	0.429	0.489	0.5120	0.464
	ARI index	0.7006	0.2609	**0.7022**	0.6515	0.6865

VI. CONCLUSIONS

In this paper, a spectral cluster algorithm based on message passing (MPSC) is proposed. To improve the clustering performance of MPSC algorithm, we design a density sensitive similarity measurement. In addition, the density sensitive similarity measure can correctly describe the complex relationships among data points; it can better reveal the distribution of data and maintain the global consistency of dataset.

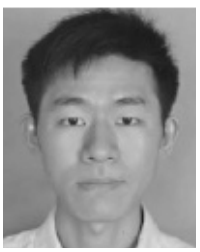
As Experiments on benchmark data sets shown, the clustering quality of MPSC algorithm is superior to that of the K-means, AP, MKAP and NJW clustering algorithm. However, our improved algorithms consume a lot of time when dealing with large-scale data, just like traditional spectral clustering. Therefore, the future study we will consider to improve the clustering efficiency of MPSC algorithm and how to apply MPSC algorithm into web data mining, image segmentation, Speech separation, text recognition and other similar scenes.

REFERENCES

[1] J. G. Sun, J. Liu, and L. Y. Zhao, "Clustering algorithms research," *J. Softw.*, vol. 19, no. 1, pp. 48–61, 2008.
[2] M. C. V. Nascimento and A. C. P. L. F. de Carvalho, "Spectral methods for graph clustering—A survey," *Eur. J. Oper. Res.*, vol. 211, no. 2, pp. 221–231, Jun. 2011.
[3] S. Ding, H. Jia, L. Zhang, and F. Jin, "Research of semi-supervised spectral clustering algorithm based on pairwise constraints," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 211–219, Jan. 2014.
[4] W. Chen and G. Feng, "Spectral clustering: A semi-supervised approach," *Neurocomputing*, vol. 77, no. 1, pp. 229–242, Feb. 2012.
[5] N. Fan and P. M. Pardalos, "Multi-way clustering and biclustering by the ratio cut and normalized cut in graphs," *J. Combinat. Optim.*, vol. 23, no. 2, pp. 224–251, Feb. 2012.
[6] U. Von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.
[7] J. Li, J. Zhou, J. Guan, and S. Zhou, "A survey of clustering algorithms based on spectra of graphs," *CAAI Trans. Intell. Syst.*, vol. 6, no. 5, pp. 405–414, Sep. 2011.
[8] F. Tung, A. Wong, and D. A. Clausi, "Enabling scalable spectral clustering for image segmentation," *Pattern Recognit.*, vol. 43, no. 12, pp. 4069–4076, 2010.
[9] D. Zhang, F. Chen, and H. Peng, "Detecting group-level crowd using spectral clustering analysis on particle trajectories," *Inf. Technol. J.*, vol. 12, no. 1, pp. 174–179, 2013.
[10] L. Ding, F. M. Gonzalez-Longatt, P. Wall, and V. Terzija, "Two-step spectral clustering controlled islanding algorithm," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 75–84, Feb. 2013.
[11] X.-Y. Liu, J.-W. Li, H. Yu, Q.-Z. You, and H.-F. Lin, "Adaptive spectral clustering based on shared nearest neighbors," *J. Chin. Comput. Syst.*, vol. 32, no. 9, pp. 1876–1880, Sep. 2011.
[12] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2005, pp. 1601–1608.
[13] X. M. Tao, S. Y. Song, P. D. Cao, and D. D. Fu, "A spectral clustering algorithm based on manifold distance kernel," *Inf. Control*, vol. 41, no. 3, pp. 307–313, Mar. 2012.
[14] H. Jia, L. Wang, H. Song, Q. Mao, and S. Ding, "A K-AP clustering algorithm based on manifold similarity measure," in *Proc. Int. Conf. Intell. Inf. Process.*, Nanning, China, 2018, pp. 20–29.
[15] S. Ding, B. Qi, H. Jia, H. Zhu, and L. Zhang, "Research of semi-supervised spectral clustering based on constraints expansion," *Neural Comput. Appl.*, vol. 22, no. 1, pp. 405–410, May 2013.
[16] D. Hamad and P. Biela, "Introduction to spectral clustering," in *Proc. IEEE 3rd Int. Conf. Inf. Commun. Technol., Theory Appl.*, Damascus, Syria, Apr. 2008, pp. 1–6.
[17] X.-Y. Cai, G.-Z. Dai, and L.-B. Yang, "Survey on spectral clustering algorithms," *Comput. Sci.*, vol. 35, no. 7, pp. 14–18, Jul. 2008.
[18] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2002, pp. 849–856.
[19] Z. Wang, G. Liu, and E. Chen, "A spectral clustering algorithm based on fuzzy K-harmonic means," *CAAI Trans. Intell. Syst.*, vol. 4, no. 2, pp. 95–99, Apr. 2009.
[20] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[21] L. Sun, R. Liu, J. Xu, S. Zhang, and Y. Tian, "An affinity propagation clustering method using hybrid Kernel function with LLE," *IEEE Access*, vol. 6, pp. 68892–68909, 2018.
[22] X. Zhang, W. Wang, K. Nørvåg, and M. Sebag, "K-AP: Generating specified K clusters by efficient affinity propagation," in *Proc. IEEE Int. Conf. Data Mining*, Sydney, NSW, Australia, Dec. 2010, pp. 1187–1192.
[23] J. Meng, H. Hao, and Y. Luan, "Classifier ensemble selection based on affinity propagation clustering," *J. Biomed. Inform.*, vol. 60, pp. 234–242, Apr. 2016.
[24] X. Zhang, J. Li, and H. Yu, "Local density adaptive similarity measurement for spectral clustering," *Pattern Recognit. Lett.*, vol. 32, no. 2, pp. 352–358, 2011.
[25] Y. Wang, Y. Jiang, Y. Wu, and Z.-H. Zhou, "Spectral clustering on multiple manifolds," *IEEE Trans. Neural Netw.*, vol. 22, no. 7, pp. 1149–1161, Jul. 2011.
[26] P. Yang, Q. Zhu, and B. Huang, "Spectral clustering with density sensitive similarity function," *Knowl.-Based Syst.*, vol. 24, no. 5, pp. 621–628, 2011.
[27] L. Wang, L.-F. Bo, and L.-C. Jiao, "Density-sensitive spectral clustering," *Act Electronica Sonica*, vol. 35, no. 8, pp. 1577–1581, Aug. 2007.
[28] J. Xie, S. Jiang, W. Xie, and X. Gao, "An efficient global K-means clustering algorithm," *JCP*, vol. 6, no. 2, pp. 271–279, Feb. 2011.

LIJUAN WANG received the M.S. degree from Northeastern University, Shenyang, China, in 2013. She is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China. Her research interests include multi-view learning, spectral clustering, and machine learning.

SHIFEI DING was born in Qingdao, China, in 1963. He received the Ph.D. degree. He has authored seven books, more than 210 articles, and more than 30 inventions. His research interests include intelligent information processing, pattern recognition, machine learning, and granular computing.



HONGJIE JIA received the M.S. and Ph.D. degrees from the China University of Mining and Technology, Xuzhou, China, in 2014 and 2017, respectively.

Since 2017, he has been with Jiangsu University, Zhenjiang, China. He has published more than 20 papers in international conferences and journals. His research interests include machine learning, data mining, and spectral clustering.