An Optimized Clustering Method with Improved Cluster Center for Social Network Based on Gravitational Search Algorithm

Liping Sun, Tao Tao, Fulong Chen, and Yonglong Luo(✉)

Engineering Technology Research Center of Network and Information Security,
Anhui Normal University, Wuhu 241000, China
ylluo@ustc.edu.cn

Abstract. Data clustering is a kind of data analysis techniques for grouping the set of data objects into clusters. To make use of the advantages of distance measure and nearest neighbor method, we present a hybrid data clustering algorithm based on GSA and DPC (GSA-DPC) algorithm. The optional clustering center set is selected by DPC algorithm. In turn, we optimize the clustering center set to achieve the best clustering distribution under the fame of GSA. Its performance is compared with four related clustering algorithms. The simulation results demonstrate the effectiveness of the presented algorithm.

Keywords: Data clustering · Gravitational Search Algorithm · Density peaks clustering · Social network

1 Introduction

Data clustering aims to divide objects into groups according to some measure of their similarities, so that objects in the same cluster are more similar to each other while other objects in different clusters [1]. With the aid of clustering method, the hidden patterns and trends within data can be revealed [2]. Clustering analysis has been widely applied for scientific and engineering applications [3,4].

At present, the typical clustering methods mainly include partitioned, hierarchical, density and etc. Similarity measure method can have an important impact on clustering algorithms. Distance measure and nearest neighbor method are two widely used measurement methods. K-means algorithm [5] is a classic of distance-based clustering algorithm, which is well known for its efficiency and easily to achieve clustering results with spherical clusters. Density peaks clustering (DPC) algorithm [6] could recognize arbitrary shape clusters on the basis of the nearest neighbor method.

Gravitational Search Algorithm (GSA) [7] is a heuristic optimization method for solving continuous optimization problems. To make use of the advantage of two kinds of distance measurement method, DPC algorithm is applied for the best cluster centers candidates, and the distance-based measure of cluster quality is defined for the fitness function of GSA.

F. Chen and Y. Luo (Eds.): Industrial IoT 2017, LNICST 202, pp. 61–71, 2017.
DOI: 10.1007/978-3-319-60753-5_7

2 Analysis of DPC Algorithm

DPC algorithm provides a new method for data clustering. The algorithm is based on the assumption that cluster centers are characterized by neighbors with relatively lower local density and by a relatively large distance from points with higher densities. Let $O = \{o_1, o_2, ..., o_N\}$ be a dataset of N vectors in $\mathbb{R}^l$. Let d_{ij} denotes the distance between the data point o_i and o_j. For N data points, distance matrix $D = \{d_t | d_t$ is the distance between two data points, $t = 1, 2, ...N^2\}$. Let p(p is set for 2% in [6]) is the percentage of the number of data point set $O.d_c \in D$ is a cutoff distance, where $c = \lfloor N^2 * p + 0.5 \rfloor$.

For each data point i, its local density ρ_i and its distance δ_i from points of higher density are defined by Eqs. (1) and (2), respectively.

$$\rho_i = \sum_j \chi(d_{ij} - d_c). \tag{1}$$

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}). \tag{2}$$

where $\chi = 1$ if $d_{ij} < d_c$ and $\chi = 0$ otherwise.

For the data point with highest density, its δ_i is defined by Eq. (3).

$$\delta_i = \max_j (d_{ij}). \tag{3}$$

Rodriguez and Laio [6] also present another local density computation as Eq. (4).

$$\rho_i = \sum_j exp^{(-d_{ij}^2/d_c^2)}. \tag{4}$$

Table 1. DPC on Zoo dataset with different values of p.

p	d$_c$	Indicators				
		ACC	NMI	REC	F_M	E_V
2%	0	0.3762	0.3482	0.2744	0.3523	0.6477
4%–8%	1.00	0.6337	0.7219	0.6848	0.5998	0.4002
9%	1.03	0.6337	0.7227	0.6855	0.6011	0.3989
10%	1.41	0.6040	0.6968	0.6518	0.5784	0.4216

d_c is the only given parameter in [6], which is used in Eqs. (1) and (4). Experimental results indicate that d_c is zero for the experiment on zoo data set, which will cause the error that division by zero. As shown in Table 1, clustering evaluation indices demonstrate that the quality of the clustering results for zoo is low when $d_c = 0$. It also demonstrates that the value of d_c has significant influence on the performance of the DPC algorithm. This shortage of DPC algorithm means the choice of d_c may lead to wrong choice of cluster centers. And once the clustering centers are identified, the remaining data points would be allocated to a wrong cluster.

3 Analysis of GSA Algorithm

GSA is a kind of heuristic optimization method which is inspired by the law of gravity and mass interactions. We apply GSA to solve clustering problems in this paper for optimizing the selection of clustering centers to achieve the best clustering distribution. Let the i-th agent be a $K * l$-dimensional vector which is represented as $X_i = \{\underbrace{x_{11}^i x_{12}^i ... x_{1l}^i}_{C_1} \cdots \underbrace{x_{j1}^i x_{j2}^i ... x_{jl}^i}_{C_j} \cdots \underbrace{x_{k1}^i x_{k2}^i ... x_{kl}^i}_{C_K}\}$, where C_i corresponds to the center of ith-cluster. The mass value of the X_i is calculated as follows.

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)}. \tag{5}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)}. \tag{6}$$

Where $M_i(t)$ and $fit_i(t)$ represent the mass value and the fitness value of the agent X_i at time t, respectively. $worst(t)$ and $best(t)$ are defined as follows for a minimization problem as follows.

$$best(t) = \min_{j \in \{1, \cdots, N\}} fit_j(t). \tag{7}$$

$$worst(t) = \max_{j \in \{1, \cdots, N\}} fit_j(t). \tag{8}$$

The acceleration of the agent X_i by the law of motion is calculated as follows.

$$a_i^d(t) = G(t) \sum_{j \in Kbest, j \neq i}^{N} rand_j \frac{M_j(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)). \tag{9}$$

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t). \tag{10}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1). \tag{11}$$

Where $Kbest$ is the set of first K agents with the best fitness value and biggest mass value, which is initialized to the value K_0 at the beginning of the GSA. ε is a decimal very close to zero but not zero used to avoid the error that division by zero whenever the similarity between agent x_i and agent x_j is equal to zero. The agent's next velocity $v_i^d(t+1)$ and position $x_i^d(t+1)$ are all based on the current value, respectively. The $rand$ is a uniformly distributed random number in the interval [0, 1].

4 GSA-DPC Clustering Algorithm

Given a data set O, the goal of a clustering algorithm is to obtain a partition $I = \{I_1, I_2, \cdots, I_K\}$ which satisfies $I_i \neq \emptyset$, for $\forall i; \bigcup_{i=1}^{K} I_i = O, I_i \cap I_j = \emptyset$, for $\forall i \neq j$. The widely used in cluster analysis for measuring the quality of finding clusters is the total mean-square quantization error (MSE) [8]. In this paper, the fitness of a clustering on data point set O with cluster C, which is defined as follows.

$$fit(O, C) = \sum_{j=1}^{K} \sum_{o_i \in C_j} \|o_i - C_j\|^2. \tag{12}$$

Where $\|o_i - C_j\|$ denotes the similarity between the point v_i and clustering center C_j.

Recall is a typical criterion to evaluate the performance of clustering algorithm, which is defined as follows.

$$Recall = \frac{\#true\ positive decisions}{\#true\ positive\ decisions\ + \#false\ negative\ decisions}. \tag{13}$$

We elaborate cluster center selection based on DPC algorithm as follows.

Algorithm 1. Cluster center selection based on DPC algorithm

Input: the set of data points $O = \{o_1, o_2, \cdots, o_N\}$, parameter p, number of clusters K.
Output: the cluster center candidate set $C = \{C_1, C_2, \cdots, C_K\}$.
1. Calculate distance matrix D by the similarity measure
2. Sort D by ascending
3. Let d_c be the $\lfloor N^2 \times p + 0.5 \rfloor$ -th element in D
4. While $d_c = 0$
5. $\quad p \leftarrow 2 * p$
6. $\quad$ Let d_c be the $\lfloor N^2 \times p + 0.5 \rfloor$-th element in D
7. Endwhile
8. Calculate $\{\rho_i\}_{i=1}^{N}$ according to Eq. (1) or (4)
9. Calculate $\{\delta_i\}_{i=1}^{N}$ according to Eqs. (2) and (3)
10. Calculate $\{\gamma_i = \rho_i \times \delta_i\}_{i=1}^{N}$
11. Sort $\{\gamma_i\}_{i=1}^{N}$ by descending
12. Choose the first K elements of $\{\gamma_i\}_{i=1}^{N}$ as $\{\gamma_i\}_{i=1}^{K}$
13. Choose K data points of O corresponding to the index of each element in $\{\gamma_i\}_{i=1}^{K}$ as cluster center candidate set C

The proposed GSA-DPC algorithm works as follows.

Algorithm 2. GSA-DPC algorithm

Input: the label vector of cluster center:$C \in \mathbb{R}_{i,K*l}$
Output: the cluster assignment I
1. Initialize an initial population : $X = \{X_1, X_2, \cdots, X_s\}$:
2. $X_1 \leftarrow C$
3. For $i = 2$ to s do
4. $X_i = [x_{11}^1 x_{12}^1 \cdots x_{1l}^1 \cdots x_{j1}^1 x_{j2}^1 \cdots x_{K1}^1 x_{K2}^1 \cdots x_{Kl}^1] \times rd$, where rd is a uniformly distributed random number in the interval $(0, 1]$
5. Endfor
6. do
7. Calculate similarity metric between o_i and C_j, for each $o_i \in O \wedge C_j \in X_t, i = 1, 2, \cdots, N, j = 1, 2, \cdots, K, t = 1, 2, \cdots, s$
8. Allocate each data point to the closest cluster center with $IDX_t(t = 1, 2, \cdots, s)$
9. Evaluate the fitness for each $X_t(t = 1, 2, \cdots, s)$ according to Eq. (12)
10. Choose the best cluster assignment I by evaluate the recall indicator using Eq. (13)
11. Calculate the agent's mass value using the Eqs. (5)-(8)
12. Calculate the agent's acceleration using the Eq. (9)
13. Calculate the agent's velocity using the Eq. (10)
14. Update the searching space X through move the agents using Eq. (11)
15. Until the stop criteria is reached

5 Experimental Results

In this section, the performances of GSA-DPC algorithm are tested through different types of the experiments. We compared our algorithm with density peaks clustering (DPC) algorithm [6], gravitational search algorithm for clustering (GSA-C) [7], Kmeans algorithm [5], and GSA-Kmeans (GSA-KM) algorithm [9] in Accuracy (ACC), Normalized Mutual Information (NMI), Recall (REC), F-measure (F_M), Expect Value (E_V). All the algorithms have been executed upon datasets for 20 times.

5.1 Experimental Results on UCI and Synthetic Datasets

As shown in Table 2, 12 data sets from the UCI machine learning repository (http://archive.ics.uci.edu/ml/) are tested in this paper. We also compared our algorithm with the other four algorithms by four 2-D synthetic datasets. The most famous similarity metric between data points of UCI and synthetic datasets is measured by Euclidean distance (as shown Eq. (14)).

$$d(o_i, o_j) = \sqrt{\sum_{t=1}^{l} (o_i^t - o_j^t)^2}. \tag{14}$$

The comparison of the algorithms in this paper is shown in Table 3. GSA-DPC algorithm outperforms others in both low-dimensional and high- dimensional data sets. To be specific, the clustering results on zoo data set indicate

Table 2. The details of UCI datasets.

DataSets	#Object	#Features	#Clusters
Iris	150	4	3
Wine	178	13	3
Glass	214	9	6
Heart	303	13	2
Liver	345	6	2
Pima	768	8	2
Vote	435	16	2
Breast	277	9	2
Wpbc	198	33	2
Zoo	101	16	7
Sonar	208	60	2
Vehicle	846	18	4

that our method has overcome the low performance caused by the value of d_c. In other words, we optimize the clustering center set to achieve the best clustering distribution.

The aggregation data set is composed of 788 data points and has instances from each of 7 classes. The compound data set contains 399 data instances and divided into 6 classes. The jain data set has 373 samples forming 2 classes. The flame data set is composed of 240 data instances and 2 classes are represented. Figure 1 presents the clustering results of our algorithm with comparison to the other methods. As shown, GSA- DPC gets perfect results on different synthetic datasets. As shown in Table 4, the experimental results illustrate our algorithm is effective in finding clusters of arbitrary shape, density and distribution.

5.2 Experimental Results on Social Networks

The comparison experiments have been carried on four typical social networks, including dolphin social networks (Dolphins), books about US politicians (Polbooks), Zachary's karate club (Karate), and American college football (football) (As shown in Table 5). Due to the characteristics of social network, the similarity metric between data points is measured by correlation coefficients (as shown Eq. (15)).

$$R(o_i, o_j) = \frac{Cov(o_i, o_j)}{\sqrt{Cov(o_i, o_i)Cov(o_j, o_j)}}. \tag{15}$$

Table 3. The results of comparison experiments on UCI datasets.

Algorithm	Iris					Wine				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	0.940	0.826	0.887	0.890	0.110	0.962	0.866	0.933	0.925	0.075
DPC	0.553	0.653	0.555	0.670	0.330	0.787	0.555	0.648	0.673	0.327
GSA-C	0.887	0.742	0.798	0.811	0.189	0.956	0.854	0.923	0.914	0.086
Kmeans	0.826	0.708	0.751	0.779	0.221	0.927	0.810	0.882	0.881	0.119
GSA-KM	0.888	0.743	0.800	0.812	0.188	0.954	0.848	0.919	0.910	0.090
Algorithm	Glass					Heart				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	0.492	0.283	0.407	0.404	0.596	0.791	0.265	0.663	0.676	0.324
DPC	0.509	0.299	0.341	0.432	0.568	0.696	0.137	0.563	0.622	0.378
GSA-C	0.435	0.302	0.386	0.375	0.625	0.691	0.122	0.570	0.596	0.404
Kmeans	0.433	0.324	0.384	0.392	0.608	0.669	0.103	0.558	0.583	0.417
GSA-KM	0.434	0.327	0.383	0.393	0.607	0.708	0.140	0.581	0.609	0.391
Algorithm	Liver					Pima				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	0.572	0.002	0.512	0.638	0.362	0.701	0.073	0.588	0.667	0.333
DPC	0.513	0.000	0.510	0.520	0.480	0.618	0.002	0.550	0.629	0.371
GSA-C	0.556	0.000	0.511	0.605	0.395	0.668	0.052	0.590	0.597	0.403
Kmeans	0.546	0.000	0.510	0.591	0.409	0.667	0.050	0.589	0.601	0.399
GSA-KM	0.557	0.000	0.511	0.605	0.395	0.668	0.052	0.590	0.597	0.403
Algorithm	Vote					Breast				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	0.878	0.490	0.809	0.791	0.209	0.706	0.078	0.647	0.657	0.343
DPC	0.876	0.506	0.807	0.787	0.213	0.516	0.000	0.582	0.540	0.460
GSA-C	0.851	0.444	0.778	0.765	0.235	0.617	0.045	0.622	0.596	0.404
Kmeans	0.852	0.438	0.778	0.770	0.230	0.641	0.051	0.626	0.615	0.385
GSA-KM	0.867	0.469	0.793	0.774	0.226	0.604	0.041	0.619	0.589	0.411
Algorithm	Wpbc					Zoo				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	0.730	0.011	0.644	0731	0.269	0.835	0.813	0.822	0.880	0.120
DPC	0.672	0.033	0.618	0.695	0.305	0.376	0.348	0.274	0.352	0.648
GSA-C	0.627	0.026	0.660	0.595	0.405	0.776	0.794	0.814	0.812	0.188
Kmeans	0.596	0.026	0.652	0.576	0.424	0.694	0.762	0.775	0.722	0.278
GSA-KM	0.601	0.027	0.654	0.577	0.423	0.701	0.790	0.829	0.733	0.267
Algorithm	Sonar					Vehicle				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	0.567	0.015	0.507	0.508	0.492	0.395	0.137	0.313	0.328	0.672
DPC	0.510	0.000	0.498	0.524	0.476	0.363	0.142	0.285	0.358	0.642
GSA-C	0.498	0.012	0.499	0.577	0.423	0.363	0.130	0.295	0.368	0.632
Kmeans	0.545	0.006	0.502	0.523	0.477	0.362	0.111	0.307	0.316	0.684
GSA-KM	0.549	0.008	0.503	0.518	0.482	0.363	0.112	0.308	0.316	0.684

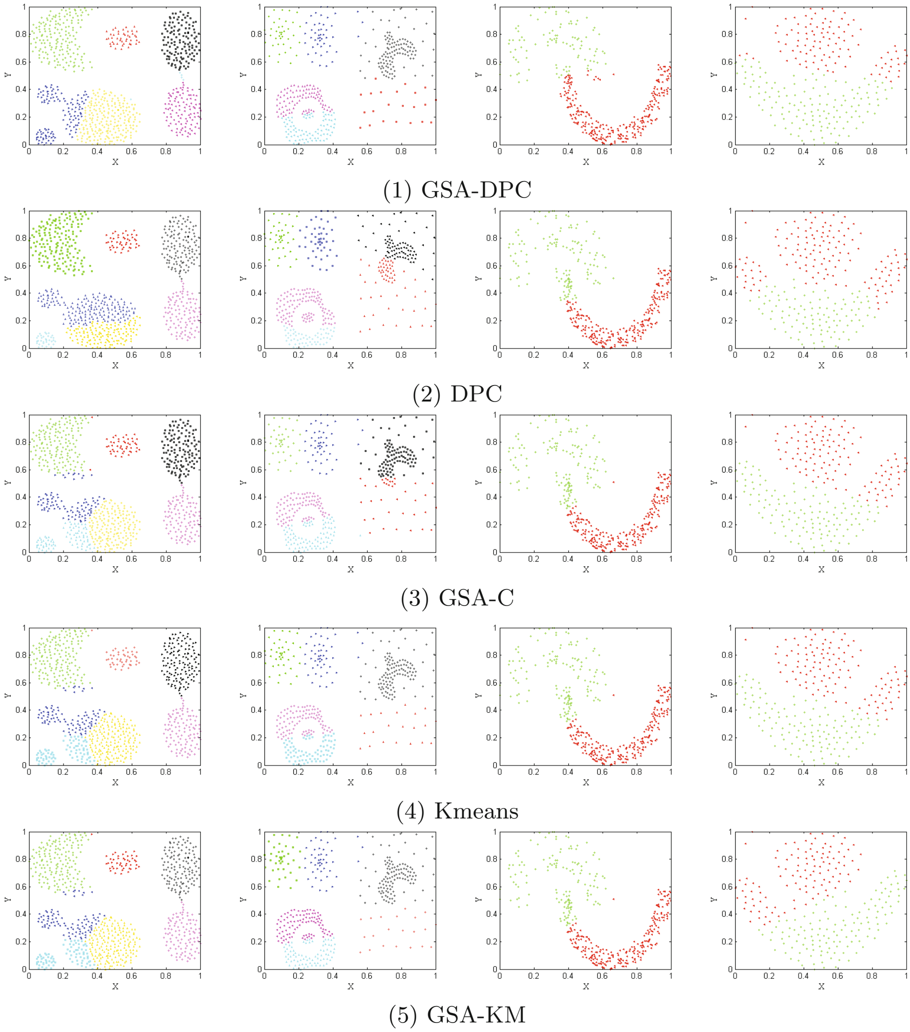

Fig. 1. Cluster assignment of artificial datasets.

We demonstrate the feasibility of our algorithm by experiments and simulations on social network data sets. Table 6 illustrates the performances of all the methods in this paper. DPC method does not need to iterate in the algorithm. Regardless of DPC algorithm, our algorithm is the most stable within the remaining four algorithms.

Table 4. The results of artificial datasets.

Algorithm	Aggregation					Compound				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	0.855	0.847	0.916	0.803	0.197	0.671	0.717	0.757	0.654	0.346
DPC	0.787	0.894	0.906	0.802	0.198	0.667	0.736	0.742	0.643	0.357
GSA-C	0.787	0.842	0.903	0.766	0.234	0.584	0.701	0.730	0.632	0.368
Kmeans	0.784	0.825	0.877	0.759	0.241	0.593	0.693	0.713	0.629	0.371
GSA-KM	0.818	0.839	0.906	0.772	0.228	0.643	0.716	0.746	0.651	0.349
Algorithm	Jain					Flame				
	ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSADPC	0.918	0.588	0903	0.874	0.126	0.867	0.462	0.801	0.778	0.222
DPC	0.895	0.577	0.898	0.837	0.163	0.788	0.413	0.696	0.678	0.322
GSA-C	0.882	0.527	0.882	0.818	0.182	0.845	0.423	0.773	0.746	0.254
Kmeans	0.882	0.527	0.882	0.818	0.182	0.849	0.441	0.779	0.751	0.249
GSA-KM	0.882	0.527	0.882	0.818	0.182	0.840	0.407	0.767	0.740	0.260

Table 5. The details of SNS datasets [10–13].

DataSets	#Nodes	#Links	Source
Football	115	613	American college football (football)
Karate	34	78	Zachary's karate club (Karate)
Polbooks	105	441	books about US politicians (Polbooks)
Dolphins	62	159	dolphin social networks (Dolphins)

Table 6. The result of SNS datasets.

Algorithm		Football					Karate				
		ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	**Max**	0.826	0.872	0.791	0.810	0.190	0.912	0.574	0.826	0.831	0.470
	Min	0.826	0.872	0.791	0.810	0.190	0.618	0.040	0.500	0.530	0.169
	ST.D	0.000	0.000	0.000	0.000	0.000	0.081	0.146	0.088	0.085	0.085
DPC	**Max**	0.817	0.862	0.744	0.783	0.217	0.500	0.046	0.485	0.639	0.361
	Min	0.817	0.862	0.744	0.783	0.217	0.500	0.046	0.485	0.639	0.361
	ST.D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GSA-C	**Max**	0.896	0.906	0.853	0.869	0.365	0.882	0.492	0.780	0.783	0.470
	Min	0.704	0.740	0.577	0.635	0.131	0.618	0.040	0.500	0.530	0.217
	ST.D	0.058	0.044	0.080	0.071	0.071	0.079	0.137	0.087	0.077	0.077
Kmeans	**Max**	0.870	0.887	0.787	0.817	0.392	0.941	0.677	0.883	0.883	0.345
	Min	0.635	0.759	0.510	0.608	0.183	0.529	0.000	0.487	0.655	0.117
	ST.D	0.080	0.042	0.097	0.077	0.077	0.088	0.159	0.091	0.063	0.063
GSA-KM	**Max**	0.800	0.872	0.758	0.766	0.596	0.853	0.401	0.735	0.734	0.266
	Min	0.574	0.598	0.397	0.404	0.234	0.853	0.401	0.735	0.734	0.266
	ST.D	0.095	0.117	0.138	0.158	0.158	0.000	0.000	0.000	0.000	0.000
Algorithm		Polbooks					Dolphins				
		ACC	NMI	REC	F_M	E_V	ACC	NMI	REC	F_M	E_V
GSA-DPC	**Max**	0.676	0.340	0.596	0.590	0.535	0.774	0.247	0.702	0.662	0.480
	Min	0.486	0.098	0.429	0.465	0.410	0.548	0.006	0.553	0.520	0.338
	ST.D	0.059	0.071	0.050	0.034	0.034	0.065	0.084	0.043	0.041	0.041
DPC	**Max**	0.705	0.541	0.725	0.723	0.277	0.677	0.284	0.604	0.593	0.407
	Min	0.705	0.541	0.725	0.723	0.277	0.677	0.284	0.604	0.593	0.407
	ST.D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GSA-C	**Max**	0.686	0.332	0.586	0.597	0.548	0.742	0.185	0.668	0.630	0.482
	Min	0.448	0.077	0.428	0.452	0.403	0.371	0.000	0.546	0.518	0.370
	ST.D	0.064	0.074	0.056	0.041	0.041	0.099	0.058	0.037	0.034	0.034
Kmeans	**Max**	0.781	0.454	0.751	0.711	0.385	0.887	0.510	0.845	0.809	0.444
	Min	0.648	0.339	0.664	0.615	0.289	0.387	0.108	0.561	0.556	0.191
	ST.D	0.041	0.032	0.024	0.024	0.024	0.149	0.150	0.106	0.098	0.098
GSA-KM	**Max**	0.781	0.470	0.760	0.702	0.405	0.871	0.475	0.824	0.785	0.484
	Min	0.638	0.318	0.650	0.595	0.298	0.371	0.000	0.548	0.516	0.215
	ST.D	0.061	0.052	0.039	0.041	0.041	0.176	0.190	0.113	0.108	0.108

6 Experimental Results

In this work, a hybrid clustering method based on gravitational search algorithm (GSA) and density peaks clustering (DPC) algorithm is presented. It tries to exploit the merits of two algorithms simultaneously. The performance of the proposed algorithm is compared with four related approaches. The comparison results show the effectiveness of the proposed method.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (No. 61602009, No.61672039, No. 61572036), Anhui Provincial Natural Science Foundation (No. 1608085MF145), Research Program of Anhui Province Education Department (No. KJ2015A067).

References

1. Jain, A.K.: Data clustering: a review. ACM Comput. Surv. **31**(3), 264–323 (1999)
2. Bonab, M.B., Hashim, S.Z.M., Bazin, N.E.N., Alsaedi, A.K.Z.: An effective hybrid of bees algorithm and differential evolution algorithm in data clustering. Math. Probl. Eng. **2015**(2), 1–17 (2015)
3. Jain, A.K., Dubes, R.C.: Algorithms for clustering data. Technometrics **32**(32), 227–229 (2015)
4. Ashouri, M., Yousefi, H., Basiri, J., Hemmatyar, A.M.A., Movaghar, A.: PDC: prediction-based data-aware clustering in wireless sensor networks. J. Parallel Distrib. Comput. **81–82**, 24–35 (2015)
5. Macqueen, J.: Some methods for classification and analysis of multivariate observations In: Proceeding of Berkeley Symposium on Mathematical Statistics and Probability (1967)
6. Rodriguez, A., Laio, A.: Clustering by fast search and find of density peaks. Science **344**(6191), 1492–1496 (2015)
7. Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S.: GSA: a gravitational search algorithm. Intell. Inf. Manage. **04**(6), 390–395 (2012)
8. Güngör, Z., Ünler, A.: K-harmonic means data clustering with simulated annealing heuristic. Appl. Math. Comput. **184**(2), 199–209 (2007)
9. Hatamlou, A., Abdullah, S., Nezamabadi-Pour, H.: A combined approach for clustering based on k-means and gravitational search algorithms. Swarm Evol. Comput. **6**, 47–52 (2012)
10. Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M.: The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Behav. Ecol. Sociobiol. **54**(4), 396–405 (2003)
11. Zachary, W.W.: An information flow model for conflict and fission in small groups. J. Anthropol. Res. **33**(4), 473 (1977)
12. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. U.S.A. **99**(12), 7821–7826 (2002)
13. Newman, M.E.: Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. **74**(3 Pt 2), 92–100 (2006)