Pers Ubiquit Comput (2016) 20:785–793
DOI 10.1007/s00779-016-0954-4

ORIGINAL ARTICLE

Adaptive fuzzy clustering by fast search and find of density peaks

Rongfang Bie[1] · Rashid Mehmood[1,2] · Shanshan Ruan[1] · Yunchuan Sun[3] · Hussain Dawood[4]

Received: 1 February 2016 / Accepted: 12 June 2016 / Published online: 27 August 2016

Abstract Clustering by fast search and find of density peaks (CFSFDP) is proposed to cluster the data by finding of density peaks. CFSFDP is based on two assumptions that: a cluster center is a high dense data point as compared to its surrounding neighbors, and it lies at a large distance from other cluster centers. Based on these assumptions, CFSFDP supports a heuristic approach, known as decision graph to manually select cluster centers. Manual selection of cluster centers is a big limitation of CFSFDP in intelligent data analysis. In this paper, we proposed a fuzzy-CFSFDP method for adaptively selecting the cluster centers, effectively. It uses the fuzzy rules, based on aforementioned assumption for the selection of cluster centers. We performed a number of experiments on nine synthetic clustering datasets and compared the resulting clusters with the state-of-the-art methods. Clustering results and the comparisons of synthetic data validate the robustness and effectiveness of proposed fuzzy-CFSFDP method.

Keywords Clustering · Decision graph · Fuzzy clustering · Density peaks

✉ Yunchuan Sun
yunch@bnu.edu.cn

Rongfang Bie
rfbie@bnu.edu.cn

Rashid Mehmood
gulkhan007@gmail.com

Shanshan Ruan
shanshan_ruan@mail.bnu.edu.cn

Hussain Dawood
hussaindawood2002@yahoo.com

1 College of Information Science and Technology, Beijing Normal University, Beijing 100875, China

2 Department of Computer Science and Information Technology, University of Management Sciences and Information Technology, Kotli, AJK, Pakistan

3 Business School, Beijing Normal University, Beijing 100875, China

4 Department of Computer Engineering, University of Engineering and Technology, Taxila, Pakistan

1 Introduction

Clustering is a fundamental approach to organize data into distinct groups for finding intrinsic hidden patterns of the data. It can be applied on various fields such as image processing [1–5], cyber security [6, 7], pattern recognition [8, 9], bioinformatics [10–14], protein analysis [15, 16], micro-array analysis [17], and social networks [18]. Clustering algorithms attempt to group more similar data into the same cluster, while dissimilar data are organized into different clusters. Many clustering algorithms have been proposed based on different characteristics and can be further categorized into portioning-based [19–21], density-based [22–28], model-based [29, 30], hierarchical-based [31–34], and grid-based [35] approaches.

K-means [20] is a state-of-the-art clustering algorithm. It partitions data into k number of partitions, and then each partition is iteratively optimized to get the optimized clusters. K-means creates spherical clusters and not sensible to detect outliers or noise in the data. K-means is simple to understand and implement. The effectiveness of K-means is subject to the appropriate knowledge of number of clusters and selection of initial centroids.

Affinity propagation (AP) [36] is effective clustering approach based on message passing between data points.

Unlike K-means, etc., AP does not need the prior knowledge of the number of clusters or selection of initial centroid. However, time and space complexity of AP is much higher than K-means.

Mean shift [37] is a famous kernel density estimation-based clustering algorithm. It is successfully used to image segmentation, image clustering, visual tracking, and air-travel routines. However, the effectiveness of mean shift depends upon the window size, which is used as bandwidth to estimate the densities. Time complexity of mean shift is also higher than K-means.

Recently, density-based clustering approaches have gained popularity among researchers. Density-based clustering algorithms attempt to find arbitrary shapes of clusters in the large spatial domain of datasets even in the presence of noise. These approaches require minimum domain knowledge to organize data into clusters [22].

DBSCAN [23] is a state-of-the-art density-based clustering algorithm that discovers arbitrary shapes of clusters utilizing minimum domain knowledge about data. However, the effectiveness of DBSCAN is subjected to the appropriate selection of input parameters, and it is not fully deterministic for border points and could not perform well in overlapping densities. A various number of variant have been proposed to overcome these limitations, such as OPTICS [24], DBCLASD [25], VDBSCAN [26], ST-DBSCAN [28].

A new density-based clustering algorithm by fast search and find of density peaks (CFSFDP) was proposed by Alex et al. [38]. CFSFDP tends to find density peaks for the selection of the cluster centers, effectively and efficiently with minimum human interaction. CFSFDP provides the heuristic approach of decision graph to the analyzer for the selection of the cluster center. The human-based selection of cluster center is a big limitation toward the spontaneous analysis of data using CFSFDP.

To overcome the aforementioned limitation of CFSFDP, we propose a fuzzy-CFSFDP for adaptive selection of the center clusters. Fuzzy-CFSFDP finds all density peaks and treats each peak as local cluster and then merges local clusters to find the global cluster. The merging process on local clusters merges them into global if more than two density peaks would closer to each other and possess an average density at the shared border region of clusters.

The rest of paper is organized as follows. The related work is presented in Sect. 2. Section 3 describes the problem formulation and proposed method. Experimental results are presented and discussed in Sect. 4, and finally, the concluding remarks and future work are presented in Sect. 5.

2 Related work

CFSFDP provides a unique solution of fast clustering by finding of density peaks in the dataset. CFSFDP is based on two assumptions that the cluster center is a highly dense point as compared with its surrounding neighbors and it is located at a large distance from other cluster centers as compared with its local data points. For each data point i, CFSFDP calculates its local density (ρ_i) and distance (δ_i) from nearest high dense point. ρ_i of a point i is calculated as follow:

$$\rho_i = \sum_j X(d_{ij} - d_c), \tag{1}$$

where

$$X(d) = \begin{cases} 1 & d < 0 \\ 0 & \text{otherwise} \end{cases}$$

where d_{ij} is the distance between point i to j and d_c is the cutoff distance. d_c is an important parameter to calculate the densities and can be selected based on the heuristic approach that in average there exist 1 to 2 % of neighbors in a dataset [38]. The effectiveness of CFSFDP potentially depends upon the appropriate choice of d_c. For small datasets, estimation of ρ_i using Eq. 1 might be affected by a large statistical error [38], in such case the methods of [37, 39] for estimating the densities are suggested. Equation 1 simply counts the number of points that are closer than d_c to i. However, the distance of each data point i can be calculated as follows:

$$\delta_i = \begin{cases} \min_{j:\rho_j > \rho_i}(d_{ij}) & \text{if } \exists\ j\ \ s.t. \rho_j > \rho_i \\ \max_{j:\rho_j > \rho_i}(d_{ij}) & \text{otherwise.} \end{cases} \tag{2}$$

Cluster centers possess large ρ and δ as compared with other cluster points while the data points having higher δ and low ρ are treated as hallo clusters (suitable to declare as noise or outliers).

Data points with high local or global density have the maximum value of δ. Therefore, δ_i is much larger for locally or globally high dense data points. CFSFDP provides a heuristic approach to analyzer for the selection of expected cluster centers manually by plotting calculated statistics of ρ and δ on a decision graph, as shown in Fig. 1b.

Figure 1a contains 28 data points that are shown with decreasing density order, while the calculated crossposting values of ρ and δ are plotted in Fig. 1b. In Fig. 1b, points 1 and 10 have the high density with high value of δ, which is the characteristics of cluster center. However, points 26, 27 and 28 have high values of δ and low values of ρ, hence can be considered as outliers or noise. Thus, decision graph is a key feature of CFSFDP to select cluster centers with minimum human interaction.

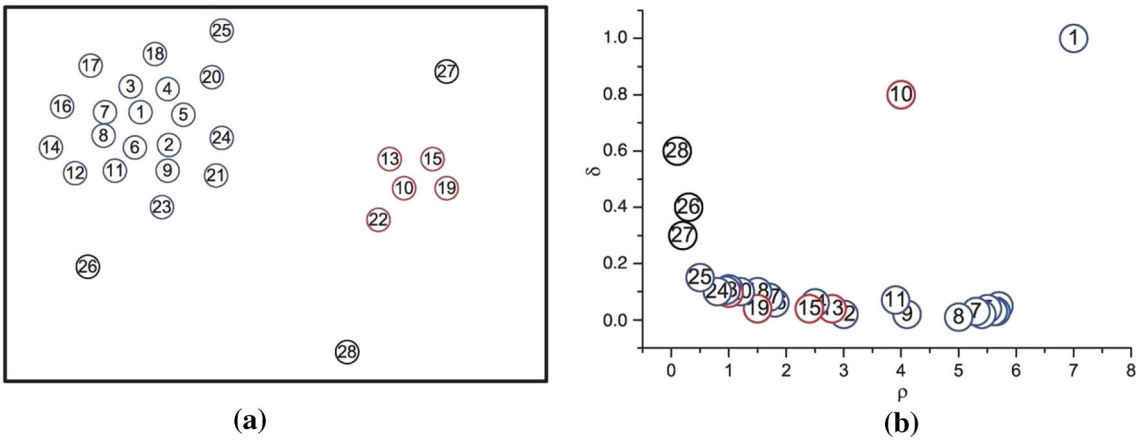

Fig. 1 CFSFDP in two dimensions. **a** Data points distribution. **b** Decision graph for data in **a** [38]

After successful declaration of cluster centers, remaining data points are assigned to the cluster centers in a single round based on minimum nearest distance to cluster center.

Furthermore, to refine and separate noise from the clusters, a border region is identified for each cluster. Border region is defined as a set of points that are at d_c distance from other cluster's points. CFSFDP finds maximum dense points at border region, known as ρ_b. Data points that have higher density than ρ_b are considered as cluster core, and rest of them are declared as noise or outlier know as cluster halos.

Algorithm 1: Fuzzy clustering by fast search and find of density peaks

Require: D distance matrix, **Output:** Organized clusters

1. calculate ρ_i from Eq. 1
2. calculate δ_i from Eq. 2
3. plot ρ and δ on decision graph
4. select cluster centers through decision graph
5. assign remaining points to local cluster centers
6. check the border point conditions for created clusters.

3 Proposed method

In this section, we explain the problem formulation and then propose fuzzy clustering by fast search and finding of density peaks, in detail.

3.1 Problem formulation

In CFSFDP, decision graph is a heuristic approach for analyzers to manually select the expected cluster centers on the basis of high density and high distance values. The human-based selection of cluster centers is a potential barrier toward automatic analysis of data. According to CFSFDP, a cluster center has higher ρ and large value of δ as compared with non-center data points. Thus, on the decision graph, expected clusters always possess large value of δ as compared with the non-cluster data points. However, in some cases single cluster contains more than one density peak and CFSFDP considers each different density peak as a potential cluster center that makes difficult for human to select exact number of clusters in a dataset. These phenomena can be easily observed from decision graph of aggregation dataset, as presented in Fig. 2c. To make effective selection of clusters on decision graph, human should be expert to the domain of underlying dataset.

The key limitations of CFSFDP are: (1) human-based selection of cluster centers is a big barrier in intelligent analysis of data, (2) when one cluster contains more than one density peaks, it is hard to identify cluster centers through decision graph.

3.2 Fuzzy clustering by fast search and finding of density peaks

Fuzzy-CFSFDP is an adaptive way to select exact number of cluster without human intervention and merge clusters if density peaks would closer to each other and possess an average density at the shared border region of clusters. Fuzzy-CFSFDP is based on the fact that:

$$EC_i = (\delta_i) \geq 2\sigma(\delta_i), \tag{3}$$

where, EC_i presents the expected cluster centers, $\sigma(\delta_i)$ is the standard deviation of all distances calculated by Eq. 2. Equation 3 is derived from the definition of cluster center given in CFSFDP. According to CFSFDP, cluster center has large distance from other cluster centers; therefore, the rest of data points should be less than $2\sigma(\delta)$. Therefore, EC_i contains noisy data points and all those points which potentially might be cluster centers. According to CFSFDP, noisy data points possess high value of δ but low value of

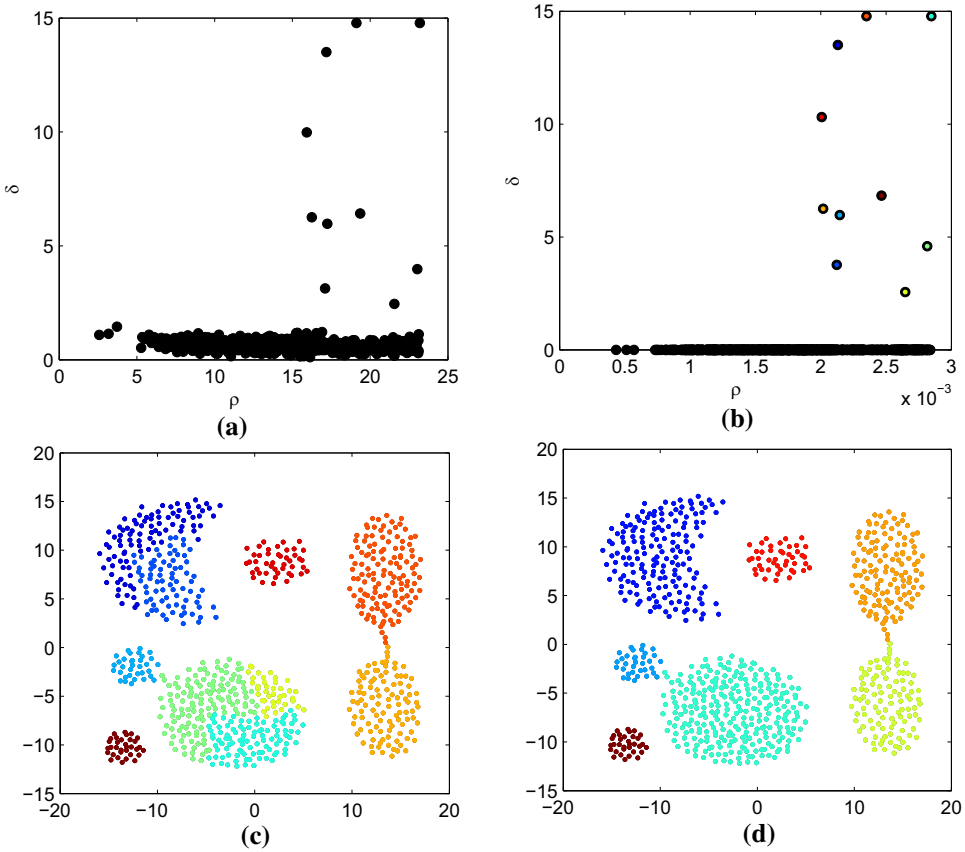

Fig. 2 Decision graph of CFSFDP and clustering results of fuzzy-CFSFDP in aggregation dataset. **a** Decision graph of aggregation dataset representation created by CFSFDP. **b** Presents the decision graph of aggregation dataset created by fuzzy-CFSFDP. *Colored points* represent the number of expected cluster centers, and all non-center points are adjusted as $\delta = 0$ for better understanding of readers. **c** Points out the local clustering results of fuzzy-CFSFDP method without merging of closer densities. **d** Final cluster created by fuzzy-CFSFDP after merging the local clusters into global clusters (color figure online)

ρ. Therefore, the noise from expected cluster centers can be separated by using the following equation:

$$LC_i = EC_i \geq \mu(\rho_i), \tag{4}$$

where LC_i are local cluster centers without noisy data points, and $\mu(\rho_i)$ is the mean of all estimated values of ρ_i. The local cluster centers should be the points that have ρ greater or equal to the average values of ρ, because noisy data points are data points which have low ρ in a dataset. In this way, local cluster centers are the data points that have large distance and higher densities as compared with the neighbor's data points. After the selection of local cluster centers, fuzzy-CFSFDP assigns remaining data points to each cluster center based on their minimum distance from each local cluster center. The next step is to merge the local cluster into the global clusters. To merge clusters into global clusters, fuzzy-CFSFDP finds minimum distance between local clusters and merges into single cluster if that cluster resides at a d_c distance from other cluster with average density.

Algorithm 2: Fuzzy clustering by fast search and find of density peaks

Require: D distance matrix, **Output:** Organized clusters

1. calculate ρ_i from Eq. 1
2. calculate δ_i from Eq. 2
3. find EC_i from Eq. 3
4. find LC_i from Eq. 4
5. assign remaining points to local cluster centers
6. merge local clusters into global clusters
7. check the border point conditions for created clusters.

3.3 Complexity analysis

The fuzzy-CFSFDP uses $\mathcal{O}(n)$ operation to find the expected cluster centers and further takes $\mathcal{O}(n)$ operation to refine the expected clusters to discover local clusters. To merge local cluster into global clusters , proposed method finds two nearest clusters and merges only and only if border points are at d_c distance with average density from other cluster border region. To merge two local clusters into global clusters, it needs $\mathcal{O}(LC_i * LC_j)$ operations, where i and j are data points that belong to cluster LC_i and LC_j.

4 Experiments

To evaluate the robustness of our proposed method, clusters created by fuzzy-CFSFDP are compared with state-of-the-art clustering methods on synthetic clustering datasets. The details of used datasets are shown in Table 1.

4.1 Result and discussion

To evaluate the performance of fuzzy-CFSFDP method, we used the aggregation dataset. In aggregation dataset, some clusters are composition of different densities. The CFSFDP method tends to find the maximum dense point in each density and highlight it as a cluster center in decision graph, as shown in Fig. 2a. To select the exact number of cluster centers from decision graph in Fig. 2a, human should have the domain knowledge of underlying dataset. However, fuzzy-CFSFDP first finds the local cluster centers by utilizing Eq. 3 and then removes the noise points from expected cluster centers by using Eq. 4. After Eq. 4, the local clusters are shown in Fig. 2c. Ten local expected cluster centers are shown in Fig. 2b, where the δ of non-cluster points are adjusted as zero to separate non-cluster points from cluster centers. After the identification of local expected cluster centers, the rest of points are assigned to the cluster centers in a single round. The next step is to merge the expected local clusters into global clusters if two clusters shared the border region and average density greater than d_c is found then they are merged into single cluster. The global clusters of aggregation dataset are shown in Fig. 2d.

In path-based spiral dataset, fuzzy-CFSFDP detects four local clusters as shown in Fig. 3a. In Fig. 3a, colors points presents the cluster centers and have higher value of δ. However, the δ of non-center points are adjusted as zero to separate non-cluster points from cluster centers. The four local clusters of path-based spiral dataset are shown in Fig. 3b. In next step, fuzzy-CFSFDP merges only two local clusters into one cluster because they shared an average threshhold density at shared border region. The global clusters created by fuzzy-CFSFDP are shown in Fig. 3c.

Table 1 The detail description of datasets

Dataset	objects (n)	Dimensions (d)	Classes (k)	Sources
Aggregation	788	2	7	[40]
flame	240	2	2	[41]
Path-based spiral	312	2	2	[42]
R15	600	2	15	[43]
D31	3100	2	31	[43]
Dim2	1650	2	9	[44]
Toys problem	300	2	3	[45]
A1	3000	2	20	[46]
S1	5000	2	15	[47]

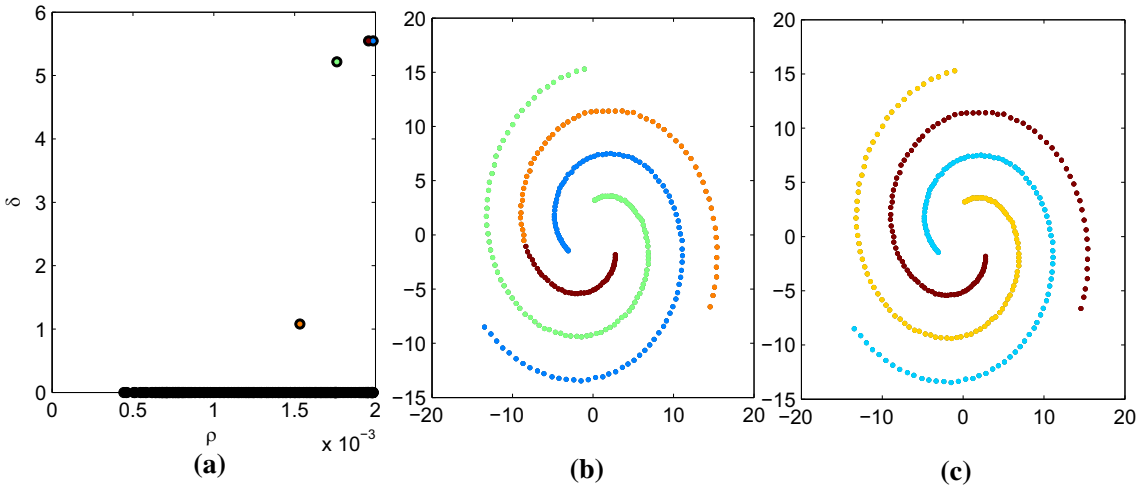

Fig. 3 Fuzzy-CFSFDP detection of cluster centers, expected and global created clusters. **a** Detected expected cluster centers in dataset, the *colored points* are expected local cluster centers and *black point* are cluster non-center points. **b** Presents local cluster of path-based spiral dataset. **c** Points out the global clusters of path-based spiral dataset after merging closed densities into single cluster (color figure online)

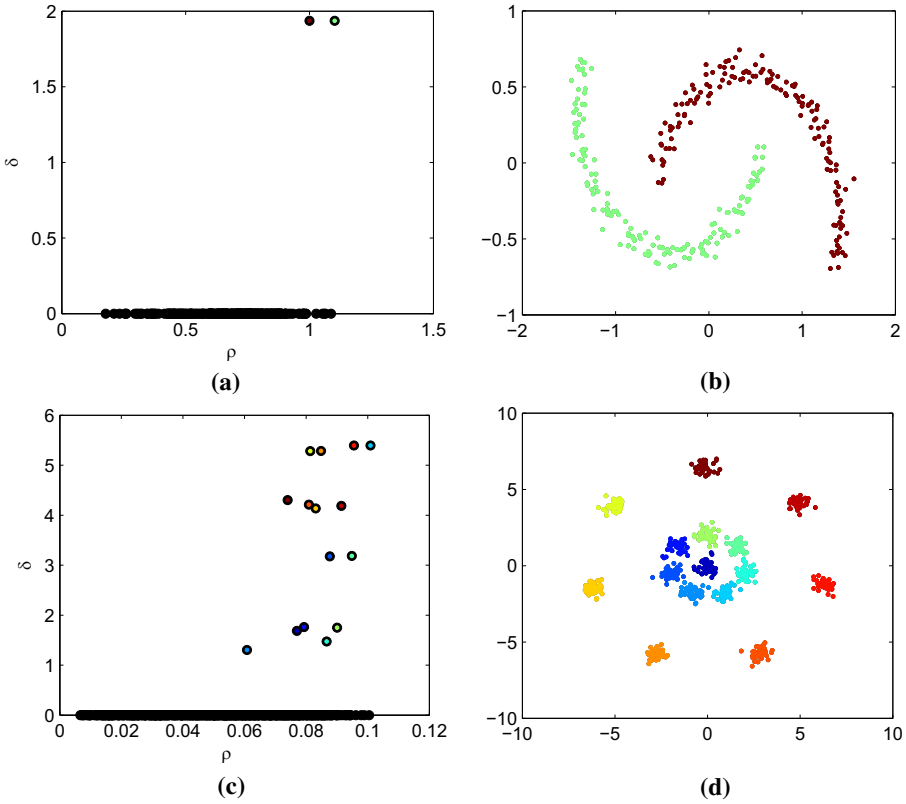

Fig. 4 Identified cluster centers and created clusters in toys problem and R15 datasets by fuzzy-CFSFDP. **a** Two *colored points* presented as expected local cluster centers, which have higher δ value. **b** Clusters of toys problem created by fuzzy-CFSFDP. **c** 15 clusters centers organized by fuzzy-CFSFDP in R15 dataset. **d** Shows the created cluster of R15 dataset by assigning points to cluster centers in a single round (color figure online)

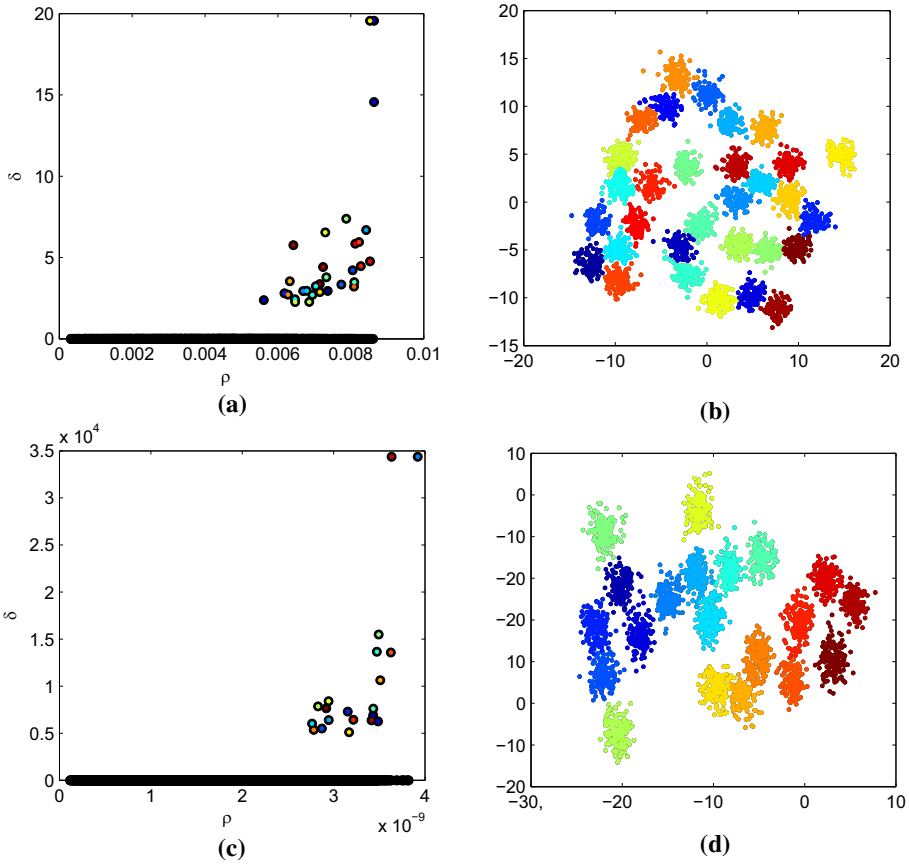

Fig. 5 Detection of cluster centers in large datasets and effectively separation of clusters using proposed method. **a** Identified 31 cluster centers in D31 dataset by fuzzy-CFSFDP. Cluster centers are denoted with *colored points*, and non-center points are shown having 0 value of δ. **b** 31 organized clusters of D31 dataset by fuzzy-CFSFDP. **c** Fuzzy-CFSFDP discovered 20 cluster centers in A1 dataset, having nonzero value of δ. **d** Shows the 20 clusters of A1 dataset, created by fuzzy-CFSFDP (color figure online)

We also use small datasets like toys problem and R15 to evaluate the robustness of proposed fuzzy-CFSFDP method. In both toys problem and R15 datasets, fuzzy-CFSFDP successfully identifies the exact cluster centers as shown in Fig. 4a, c. In both figures, the cluster centers are those points, which have higher value of δ. For non-cluster points δ is adjusted as zero to separate non-clusters points from cluster points. The shared densities at border regions are not sufficient to merge the local clusters into global clusters. So in this case, merging process is skipped and local clusters are considered as global clusters. In these cases, the computational cost of fuzzy-CFSFDP and CFSFDP is almost same. The global clusters of these datasets are shown in Fig. 4b, d, respectively.

To benchmark our proposed fuzzy-CFSFDP method on large datasets, we use Dim-2, A1, D31, and S1 datasets. In all these datasets, fuzzy-CFSFDP successfully identifies exact number of clusters without further merging the local clusters. In these datasets, each cluster contains only a single density peak. If two density peaks do not share common boarder points with average density found in cluster, fuzzy-CFSFDP skips the merging process in such scenario (Fig. 5).

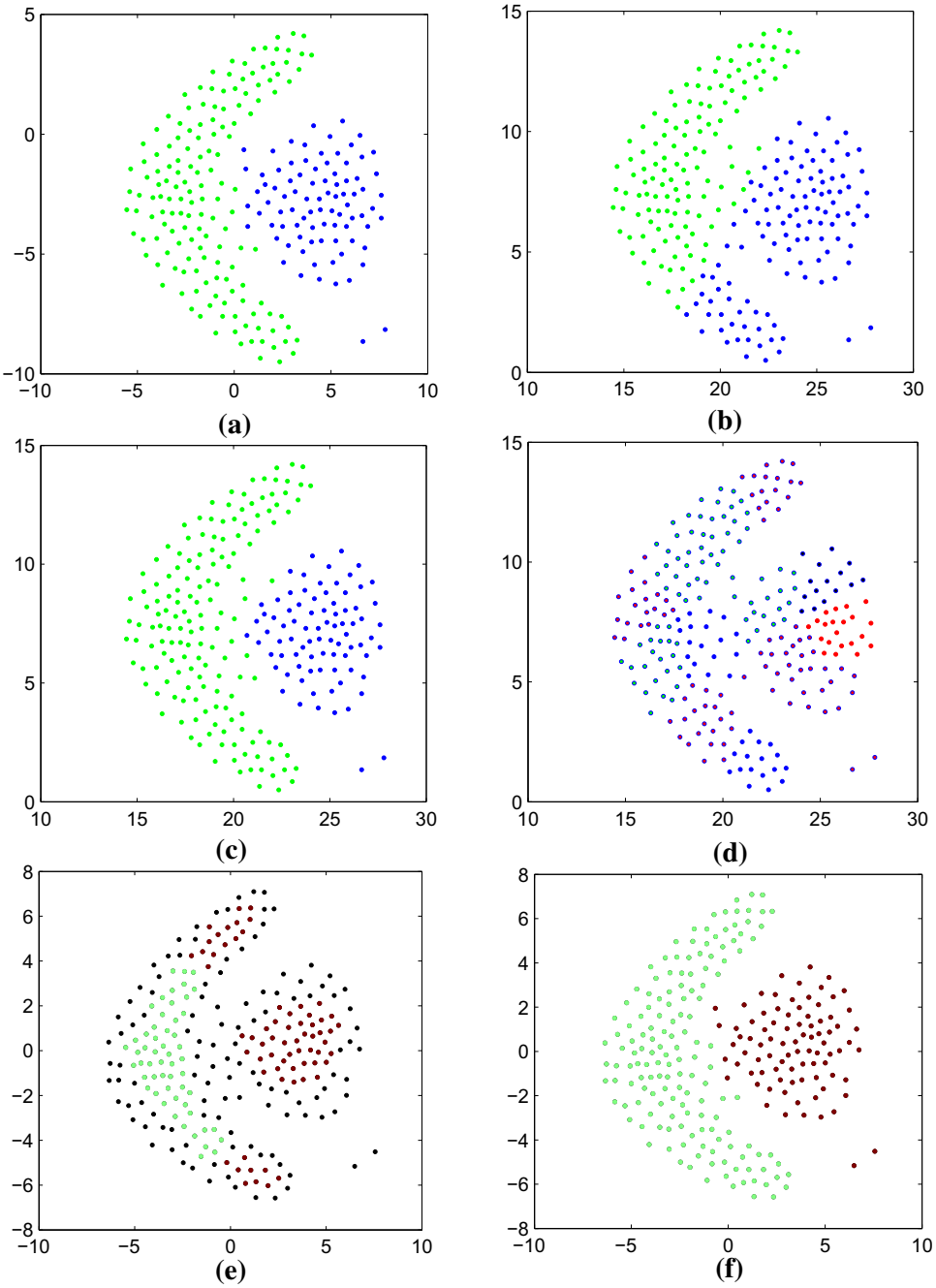

Fig. 6 Comparison of proposed fuzzy-CFSFDP with state-of-the-art clustering algorithms. **a**The ideally separated 2 clusters of flame dataset. **b** Clusters obtained by *K*-means, at $k = 2$. **c** Two clusters created by mean shift clustering method at optimal size of window. **d** 13 clusters created by affinity propagation clustering of flame dataset. **e** Two clusters created by using CFSFDP in flame dataset. **f** Ideal clusters created by fuzzy-CFSFDP of flame dataset

Table 2 The number of local clusters and global clusters in each dataset created by proposed fuzzy-CFSFDP

Dataset	objects (n)	Classes (k)	Local clusters	Global clusters
Aggregation	788	7	7	10
flame	240	2	2	2
Path-based spiral	312	3	2	4
R15	600	15	15	15
D31	3100	31	31	31
Dim2	1650	9	9	9
Toys problem	300	2	3	2
A1	3000	20	20	20
S1	5000	15	15	15

We utilized gene dataset (flame) to evaluate the effectiveness of fuzzy-CFSFDP. We also make comparison with the state-of-the-art methods to validate the robustness of fuzzy-CFSFDP method. On flame dataset, K-means creates spherical clusters because it is not sensitive to detect the connected densities, as shown in Fig. 6b. Affinity propagations create 13 clusters of flame dataset, as shown in Fig. 6d. Both the *K*-means and AP are partition-based clustering algorithms and make partitions of more likely spherical in shapes and hence could not find relations between connected densities. However, the results of mean shift are better than K-means and affinity propagation. Mean shift exactly finds two clusters at optimum value of window size. However, in mean shift it is harder to find the exact number of windows size. The number and shape of clusters depend upon the window size in mean shift. Mean shift miss-classifies only four data points on flame dataset, as shown in Fig. 6c. In CFSFDP, the performance of method highly depends upon the appropriate choice of d_c distance and selection of appropriate cluster centers over decision graph. The flame cluster results of CFSFDP , at dc = 0.710634 (1 % whole dataset) are shown in Fig. 6e. Basically, CFSFDP detects all density peaks in dataset and assigns maximum δ values to each density peak. However, in some datasets different density peaks exist in a single cluster. In this scenario, CFSFDP also assigns maximum distance to each peak in a single cluster that results in a potentially more cluster center. Therefore like flame dataset, CFSFDP could not find better clusters, as shown in Fig. 6e. However, fuzzy-CFSFDP is capable to detect exact number of clusters by utilizing the defined fuzzy rules and merge the local clusters into global clusters to refine the local clusters.

The proposed fuzzy-CFSFDP works in two steps: first it finds the local clusters and then merge the local clusters if different clusters shared a threshhold density at border region of the clusters. Table 2 shows the details of local and global clusters. In all tested datasets, only aggregation and spiral datasets contain connected densities and hence needed the merging process to merge local clusters into global clusters. In rest of other datasets, local clusters and global clusters are same. The fuzzy-CFSFDP does not perform the merging process on these datasets.

5 Conclusions

In this paper, a method for adaptively selection of cluster centers for CFSFDP, named as fuzzy-CFSFDP is proposed. Fuzzy-CFSFDP utilizes the fuzzy rules to select cluster centers for different density peaks and then merges density peaks in case of having similar intrinsic patterns. CFSFDP uses a heuristic approach, known as decision graph to select cluster centers manually. In fuzzy-CFSFDP, we have overcome the limitation of manual selection of cluster centers. Experiments on nine synthetic datasets present the robustness and effectiveness of our proposed fuzzy-CFSFDP method. We also compared the results with the state-of-the-art methods for validation of the effectiveness of our proposed fuzzy-CFSFDP method.

Fuzzy-CFSFDP provides robust performing in static data; however, nowadays more and more data are appearing in a dynamic manner. In future, we will try to make an incremental fuzzy-CFSFDP to deal with stream and big data.

Acknowledgments This research is sponsored by National Natural Science Foundation of China (Nos. 61171014,61371185, 61401029, 61472044, 61472403, 61571049) and the Fundamental Research Funds for the Central Universities (Nos. 2014KJJCB32, 2013NT57) and by SRF for ROCS, SEM.

References

1. Li K et al (2013) Personalized multi-modality image management and search for mobile devices. Pers Ubiquitous Comput 17(8):1817–1834
2. Jiwen L, Erin LV, Xiuzhuang Z, Jie Z (2015) Learning compact binary face descriptor for face recognition. IEEE Trans Pattern Anal Mach Intell (TPAMI) 37(10):2041–2256
3. Lu J, Zhou X, Tan Y-P, Shang Y, Zhou J (2014) Neighborhood repulsed metric learning for kinshipverification. IEEE Trans Pattern Anal Mach Intell (T-PAMI) 36(2):331–345

4. Lu J, Tan Y-P, Wang G (2013) Discriminative multimanifold analysis for face ecognition from a single training sample per person. IEEE Trans Pattern Anal Mach Intell (T-PAMI) 35(1):39–51
5. Lu J, Liong VE, Zhou J (2015) Cost-sensitive local binary feature learning for facial ageestimation. IEEE Trans Image Process (T-IP) 24(12):5356–5368
6. Yan Y, Qian Y, Sharif H, Tipper D (2012) A survey on cyber security for smart grid communications. IEEE Commun Surv Tutor 14(4):998–1010
7. Portnoy L, Eskin E, Stolfo S (2001) Intrusion detection with unlabeled data using clustering. In: Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001) pp 5–8
8. Ahn C-S, Sang-Yeob O (2014) Robust vocabulary recognition clustering model using an average estimator least mean square filter in noisy environments. Pers Ubiquitous Comput 18(6):1295–1301
9. Guo L, Ai C, Wang X, Cai Z, Li Y (2009) Real Time Clustering of Sensory Data in Wireless Sensor Networks. The 28th IEEE International Performance Computing and Communications Conference (IPCCC)
10. Yeganova L, Kim W, Kim S, Wilbur WJ (2014) Retro: concept-based clustering of biomedical topical sets. Bioinformatics 30(22):3240–3248
11. Xu C, Zhengchang S (2015) Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinformatics 37(10):2041–2256
12. Shuji S, Kakuta M, Ishida T, Akiyama Y (2015) Faster sequence homology searches by clustering subsequences. Bioinformatics 31(8):1183–1190
13. Shi Y, Hasan M, Cai Z, Lin G, Schuurmans D (2012) Linear coherent bi-clustering via beam searching and sample set clustering. Discrete Math Algorithms Appl 4(2):1250023
14. Cai Z, Heydari M, Lin G (2005) Clustering binary oligonucleotide fingerprint vectors for DNA clone classification analysis. J Comb Optim 9(2):199–211
15. Nicovich Philip R et al (2015) Analysis of nanoscale protein clustering with quantitative localization microscopy. Biophys J 108(2):475a
16. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658–1659
17. Shaw MKE (2015) K-means clustering with automatic determination of K using a Multiobjective Genetic Algorithm with applications to microarray gene expression data. Dissertation, San Diego State University
18. Chang M-S, Chen L-H, Hung L-J, Rossmanith P, Guan-Han W (2014) Exact algorithms for problems related to the densest k-set problem. Inf Process Lett 114(9):510–513
19. Kannuri L, Murty MR, Satapathy SC (2015) Partition based clustering using genetic algorithm and teaching learning based optimization: performance analysis. Adv Intell Syst Comput 338:191–200
20. MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol 1, no 14, pp 281–297
21. Park H-S, Jun C-H (2009) A simple and fast algorithm for K-medoids clustering. Expert Syst Appl 36(2):3336–3341
22. Lovely Sharma P, Ramya KA (2013) Review on density based clustering algorithms for very large datasets. Int J Emerg Technol Adv Eng 3(12):398–403
23. Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd 96(34):226–231
24. Parimala M, Lopez D, Senthilkumar NC (2011) A survey on density based clustering algorithms for mining large spatial databases. Int J Adv Sci Technol 31(1):216–223
25. Shah Glory H, Bhensdadia CK, Ganatra Amit P (2012) An empirical evaluation of density-based clustering techniques. Int J Soft Comput Eng (IJSCE) 2(1):2231–2307
26. Liu P, Zhou D, Wu N (2007) VDBSCAN: varied density based spatial clustering of applications with noise. In: Proceedings: Service Systems and Service Management 2007, pp 1–4
27. Mehmood R, Zhang G, Bie R, Dawood H, Ahmad H (2016) Clustering by fast search and find of density peaks via heat diffusion. Neurocomputing. doi:10.1016/j.neucom.2016.01.102i
28. Birant D, Kut A (2007) ST-DBSCAN: an algorithm for clustering spatial-temporal data. Data Knowl Eng 60(1):208–221
29. Chen T, Zhang NL, Liu T, Poon KM, Wang Y (2012) Model-based multidimensional clustering of categorical data. Artif Intell 176(1):2246–2269
30. Mann AK, Kaur N (2013) Survey paper on clustering techniques. Int J Sci Eng Technol Res (IJSETR) 2(4):803–806
31. Murtagh F, Contreras P (2012) Algorithms for hierarchical clustering: an overview. Wiley Interdiscip Rev: Data Min Knowl Discov 2(1):86–97
32. Chen N, Ze-shui X, Xia M (2014) Hierarchical hesitant fuzzy K-means clustering algorithm. Appl Math A J Chin Univ 29(1):1–17
33. Jaeger D, Barth J, Niehues A, Fufezan C (2014) pyGCluster, a novel hierarchical clustering approach. Bioinformatics 30(6):896–898
34. Jacques J, Preda C (2014) Functional data clustering: a survey. Adv Data Anal Classif 8(3):231–255
35. Parikh M, Varma T (2014) Survey on different grid based clustering algorithms. Int J Adv Res Comput Sci Manag Stud 2(2):427–430
36. Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315(5814):972–976
37. Cheng Y (1995) Mean shift, mode seeking, and clustering. IEEE Trans Pattern Anal Mach Intell 17(8):790–799
38. Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. Science 344(6191):1492–1496
39. Fukunaga K, Hostetler L (1975) The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inf Theory 21:32–40
40. Gionis A, Mannila H, Tsaparas P (2007) Clustering aggregation. ACM Trans Knowl Discov Data (TKDD) 1(1):1–30
41. Fu L, Medico E (2007) FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. BMC Bioinform vol 8, artical no. 3
42. Chang H, Yeung DY (2008) Robust path-based spectral clustering. Pattern Recognit 41(2):191–203
43. Veenman CJ, Reinders MJT, Backer E (2002) A maximum variance cluster algorithm. IEEE Trans Pattern Anal Mach Intell 24(9):1273–1280
44. Franti P, Virmajoki O, Hautamaki V (2006) Fast agglomerative clustering using a k-nearest neighbor graph. IEEE Trans Pattern Anal Mach Intell 28(11):1875–1881
45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
46. Karkkainen I, Franti P (2002) Dynamic local search for clustering with unknown number of clusters. In: Proceedings of International Conference on Pattern Recognition, vol 16, no 2, pp 240–243
47. Franti P, Virmajoki O (2006) Iterative shrinking method for clustering problems. Pattern Recognit 39(5):761–775