Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 2816102, 8 pages
https://doi.org/10.1155/2020/2816102

Research Article

Density Peak Clustering Based on Relative Density Optimization

Chunzhong Li and Yunong Zhang

Institute of Statistics and Applied Mathematics, Anhui University of Finance & Economics, Bengbu 233000, China

Correspondence should be addressed to Chunzhong Li; 120120038@aufe.edu.cn

Received 29 September 2019; Revised 6 May 2020; Accepted 25 May 2020; Published 11 June 2020

Guest Editor: Filipe J. Marques

Among numerous clustering algorithms, clustering by fast search and find of density peaks (DPC) is favoured because it is less affected by shapes and density structures of the data set. However, DPC still shows some limitations in clustering of data set with heterogeneity clusters and easily makes mistakes in assignment of remaining points. The new algorithm, density peak clustering based on relative density optimization (RDO-DPC), is proposed to settle these problems and try obtaining better results. With the help of neighborhood information of sample points, the proposed algorithm defines relative density of the sample data and searches and recognizes density peaks of the nonhomogeneous distribution as cluster centers. A new assignment strategy is proposed to solve the abundance classification problem. The experiments on synthetic and real data sets show good performance of the proposed algorithm.

1. Introduction

As an unsupervised machine learning algorithm, clustering groups sample data into reasonable class based on similarity between sample points. Such process tries to make the similarity between samples inside a same cluster as high as possible and the similarity between samples in different clusters as low as possible. Many different types of clustering algorithms are proposed in different applications. In general, clustering can be divided into divisive clustering [1–3], hierarchical clustering [4, 5], grid-based algorithms [6, 7], model-based algorithms [8, 9], and density-based algorithms [10, 11]. In practical applications, data sets are various and complex with high dimensions, which brings a huge challenge to clustering. Some scholars put forward the idea of considering multiple clustering algorithms comprehensively, that is, integrated clustering [12, 13], which effectively improves the accuracy of clustering. With the development of cluster analysis theory and technology, it plays an increasingly important role in image processing, machine learning, artificial intelligence, natural language processing, pattern recognition, information retrieval, and bioinformatics [14].

Clustering by fast search and find of density peaks (DPC) [15] proposes a totally new clustering frame and the type of redefining clustering center. The structures of data are mapped into two-dimensional space (local density and the nearest distance), in which centers are recognized and clusters are grouped. With DPC, density peaks of sample data are easily and quickly found and DPC also shows high efficiency in assignment and elimination of noises. However, there are still limitations in clustering with DPC. (1) There is no unified density measurement, and parameter d_c is difficult to set because it is related with specific problems. (2) Clustering centers need to be selected manually, which is qualitative analysis with subjective factors. As a result, objective and reasonable centers are difficult to find in decision graphs. (3) In terms of sample distribution, sample points are assigned to the nearest clusters with high density, which easily results in continuous transmit of the mistake clustering. (4) According to the definition of distance δ_i, two points would be selected as clustering centers if density of the two points is both the highest and belongs to the same cluster, which means one cluster is divided into two clusters mistakenly. (5) DPC shows limits in clustering of data sets with high dimension, unevenly distributed density, and noises.

To improve DPC, a new algorithm is proposed from two aspects, density measurement and assignment of the remaining points. The classical DPC algorithm uses global density, which cannot effectively identify the density peaks in the low density area. In this paper, the d_c nearest information of samples is employed to calculate the local relative density, in attempt to recognize the centers of data set with nonhomogeneous distribution. To solve the overclassification problem in DPC, a new assignment strategy with sorting of local density and defining of corresponding distances of data samples is proposed. Based on the two improvements, a density peak clustering algorithm based on relative density optimization (RDO-DPC) achieves satisfied clustering results on synthetic and real data sets with various density types and irregular shapes.

The reminder of the paper is organized as follows: Section 2 introduces the definition and process of classical DPC and related works; density peak clustering algorithm based on relative density (RDO-DPC) algorithm is proposed in Section 3; experiments on synthetic and real data sets are shown in Section 4; and Section 5 gives conclusion and prospect.

2. Related Works

2.1. DPC Algorithm. Clustering by fast search and find of density peaks (DPC) [15] could find the clusters of various densities and shapes with a simple strategy. The fundamental principle of DPC is that the ideal density peaks possess two essential features: (1) the local density of the peak is higher than the density of the neighbors; (2) the distances between different peaks are relatively longer. To find density peaks meeting the two above conditions, DPC introduces local density ρ_i of sample i and the corresponding distance δ_i, which is the distance from i to j, the sample whose local density is higher than i, and which is the nearest sample to j.

Local density depends on distance, which means it can be regarded as a function of the distance, for example, kernel function. One of the local densities is defined by cut-off kernel:

$$\rho_i = \sum_{j \neq i} \chi\left(d_{ij} - d_c\right), \tag{1}$$

where d_{ij} represents the distance between point i and j. Positive number d_c is the appointed parameter. The value of $\chi(d_{ij} - d_c)$ is set as 1 if $d_{ij} < d_c$; otherwise, it is set as 0. The other local density can be defined by Gaussian kernel:

$$\rho_i = \sum_{j \neq i} \exp\left\{-\left(\frac{d_{ij}}{d_c}\right)^2\right\}. \tag{2}$$

d_c in equations (1) and (2) can control the influence of neighbors on sample points, which equal the function of neighborhood ε. When data set is of large scale (number of points it contains), clustering result from DPC is slightly influenced by cut-off distance, and the influence from cut-off distance becomes greater and greater, while data scale becomes smaller. To avoid the influence from cut-off distance on local density, or further on clustering results, DPC employs Gaussian kernel in equation (2) to calculate overall density of the sample, while it is used to cluster small-scale data.

Another feature of ideal clustering center is that the distance between different centers should be as far as possible. As a result, δ_i, the distance from sample i to j which is the nearest to i, and whose local density is larger than i, is defined as

$$\delta_i = \begin{cases} \min\left(d_{ij}\right), & \rho_j > \rho_j, \\ \max\left(d_{ij}\right), & \rho_j = \max\left(\rho_j\right), \end{cases} \quad j = 1, 2, \ldots, n \tag{3}$$

The definition in equation (3) shows that if the density of sample i is the largest local density or the largest overall density, distance δ_i of sample i is far more larger than distance δ_j of the neighbors of i. Therefore, cluster centers are often points with extremely large δ_j, and density ρ_i of those center points is also very large. Through constructing decision graph of distance δ in relative to density ρ, DPC selects sample points with relatively large ρ and δ as cluster centers. For remaining points j, DPC assigns the points to clusters, which are the nearest to j and are larger than j in density, thus completing the distribution of remaining points with high efficiency.

2.2. Related Work. The researchers have improved the DPC [15] in many ways to adapt it to different applications, mainly focusing on the definition of cluster centers and assignment strategy.

In terms of definition of cluster centers, some scholars try to expand the differentiation between cluster center and other sample points, so as to select cluster centers in the decision-making graph, such as the normalization of local density and distance [16], gravitational analogy minimum distance [17, 18], and the Laplacian centrality in the form of no parameter [19]. Although this kind of method expands the differentiation between the density peak point and other points to a certain extent, it is still difficult to determine the cluster centers directly and effectively in some complex decision-making, and it needs manual selection. Therefore, other scholars have proposed a method to quantitatively select the class center based on the decision graph, among which the most prominent algorithms are the fuzzy theory σ principle [20], the normal distribution 3σ criterion [21], the inflexion point [22] of data distribution in the decision graph, the linear fitting of the distribution curve of density and distance product [23], and the Chebyshev inequality [24] or the upper bound of generalized extremum [25]. This kind of method can automatically determine the potential class center of the data set without human intervention. However, due to the influence of multiple density extreme value, it is often necessary to merge subclusters to further optimize the sample allocation effect.

The remaining points assignment strategy of the classical DPC is prone to chain mistaken assignments. Many improvements are proposed to modify the assignment strategy of the classical DPC, such as the distribution

of the remaining points based on the k-nearest neighbor [26, 27], the similarity measurement of the samples based on the shared nearest neighbor [28], the combination of initial clusters with boundary samples [29] or density reachable [30], and the assignment of remaining points in combination with other algorithms [11, 31]. The assignment strategy based on nearest k and shared nearest neighbors takes full consideration of the neighbor information of samples, which is beneficial to get the reasonable cluster assignment of samples. However, the mere consideration of distances between samples cannot reflect the impact of the real cluster attribution on the similarities between samples. The assignment strategy of the remaining points based on the combination of initial clusters works well on multiple density peaks, but it shows high time complexity. Moreover, some algorithms use DPC as the initial cluster center selection strategy, which can better solve the impact of initial cluster center selection on clustering results, but these algorithms all show high time complexity and are not suitable for clustering of large-scale high-dimensional data.

For high-dimensional data with noises, noises filtering standard is constructed based on nearest k, and the clustering centers recognition and remaining points assignment are conducted after filtering of noises [26, 27]. DenPEHC [23] takes sample points with a higher ratio of δ and ρ as noises, but there were still errors and manual factors. Furthermore, dimension reduction is combined to reduce the dimensions of high-dimensional data [32], and then sample points are assigned with nearest neighborhood parameter k. Furthermore, geodesic distance [33, 34] is used to calculate the manifold distance between data points, and isometric mapping is introduced to reduce the dimension of high-dimensional data sets. The above analysis shows that many improvements and optimization are proposed to solve the problems in DPC, and results are satisfying. However, many problems still exist in clustering of complex data sets, for example, uneven density of clusters, high dimensions, optimization of parameters, recognition of center, noise treatment, and high time complexity.

3. RDO-DPC Algorithm

The proposed RDO-DPC improves the classical DPC from two aspects: the definition of local density and assignment strategy of cluster members. Taking advantage of neighbor information, RDO-DPC defines a new measurement of relative density. Then, cluster centers are selected combining decision graph, so as to obtain satisfying results from the clustering of data set with uneven density between clusters. The remaining points are allocated according to the structure information of data set, which effectively avoid the disadvantage of one-step distribution strategy in DPC.

Recognizing cluster centers of different density areas is the guarantee of effective clustering results. Peaks of low density area are buried in high density peaks with local density definition in equation (2) because the local density of dense area is much higher than that of sparse area. In order to give prominence to peaks of sparse area, relative local density is defined as

$$\widetilde{\rho}_i = \frac{N_i \cdot \rho_i}{\sum_{d_{ij} \le d_c} \rho_i}, \tag{4}$$

where the radius of influence d_c is the p quantile of pairwise distances from the smallest to the biggest. N_i is the number of samples in d_c spherical neighbor of sample i. Revised local density ρ_i is defined as

$$\rho_i = \sum_{j \ne i, d_{ij} \le d_c} \exp\left\{-\left(\frac{d_{ij}}{d_c}\right)^2\right\}, \tag{5}$$

where the strict condition $d_{ij} \le d_c$ in equation (5) is equivalent to truncated Gaussian kernel function in order to eliminate the interference from samples far away. Compared with classical DPC, relative local density (4) and (5) can recognize the cluster centers of regions with different densities by employing relative index rather than absolutely index.

The ideal cluster centers of DPC possess two features: one is that local density is higher than the density of samples around, and the other one is that cluster centers are far away from each other. It is shown that distance also is important in selection of cluster centers. As a result, cluster centers are often samples with a higher density and larger distance. If there are two largest density peaks in one cluster, the two points will be both selected as cluster centers according to equation (3). The result is that one cluster is mistakenly divided into two clusters, which eventually leads to unsatisfied clustering results. Therefore, relative density is ranked before calculation of the density higher than $\widetilde{\rho}_i$ and the shortest distance to sample i, which can help the distinction of two largest density peaks. The corresponding distance of q_i is defined as

$$\delta_{q_i} = \begin{cases} \min_{j<i}\left(d_{q_i q_j}\right), & i \ge 2, \\ \max_{j \ge 2}\left(\delta_{q_j}\right), & i = 1, \end{cases} \tag{6}$$

where $\{q_i\}_{i=1}^n$ represents the subscript sequence of one descending order of $\{\rho_i\}_{i=1}^n$, satisfying $\rho_{q_1} \ge \rho_{q_2} \ge \cdots \ge \rho_{q_n}$. If the biggest local density peaks of q_i and q_j in a data set according to equations (2) or (4) are very close, it is hard to identify the real peak in decision graph. Therefore, q_i and q_j may be recognized as their own cluster centers, respectively. After the ranking of the two peaks, if $\rho_{q_i} \ge \rho_{q_j}$, the distance in corresponding to q_i is set as the largest corresponding distance of other density peaks with equation (6). The distance corresponding to q_j is the distance between q_i and q_j, which weakens the value of δ_{q_j} corresponding to q_j. As a result, q_j is no longer the cluster center.

Combined with equations (5) and (6), the peaks of areas with a greater density difference are easy to be recognized in decision graph, and the discriminability is strengthened with the decision distances that the peaks are

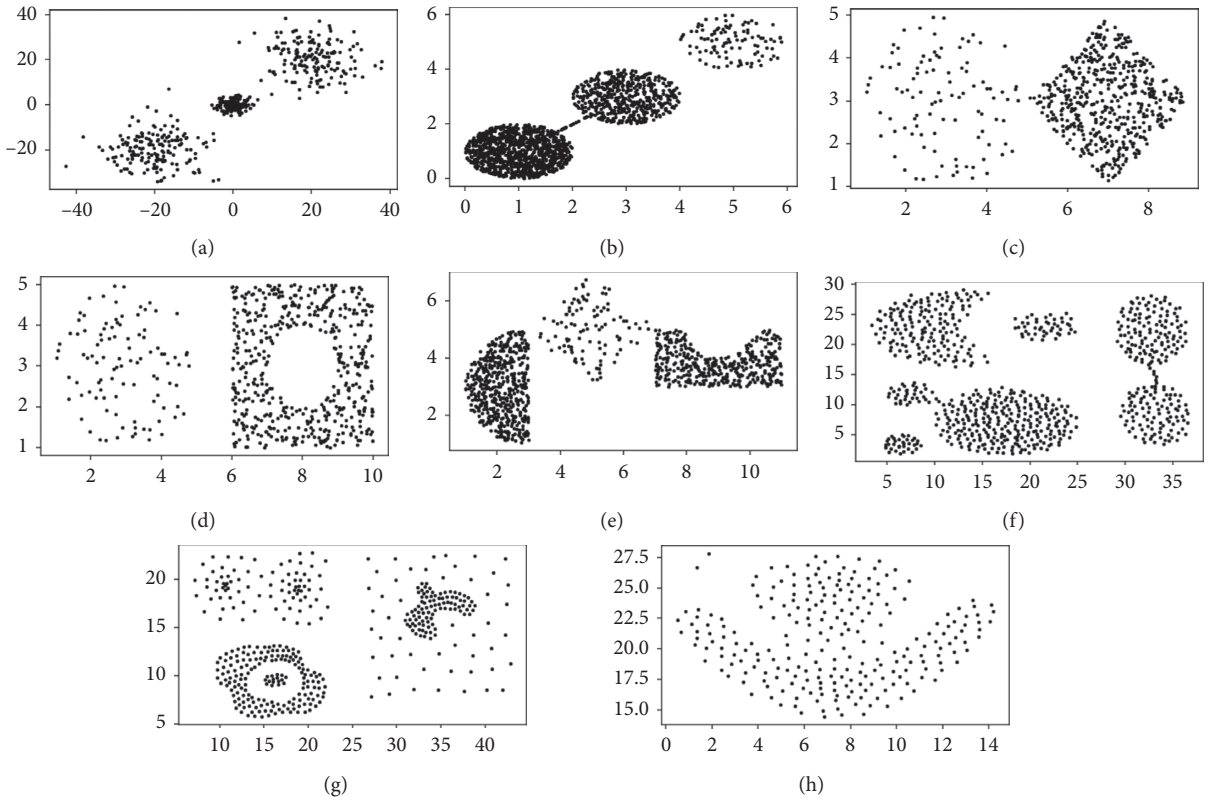

Figure 1: Two-dimensional exhibition of 8 synthetic data sets: (a) DS1; (b) DS2; (c) DS3; (d) DS4; (e) DS5; (f) aggregation; (g) compound; (h) flame.

corresponding to. Therefore, a stronger generalization ability is obtained. RDO-DPC algorithm is formed, as shown in Algorithm 1.

RDO-DPC takes relative density as measurement of density. With relative density, density calculation of each point is restricted in d_c scope, and the values are only related to points inside d_c neighbor scope. The relative closeness of samples with sample inside d_c scope can be revealed more clearly, and local information of each point and its sample point inside d_c scope can also be shown clearly. Therefore, RDO-DPC suits not only data sets with relatively even density between clusters but also data sets with obvious density differences between clusters.

The time complexity of RDO-DPC is $O(n^2)$, which consists of the measurement of relative local density $\widetilde{\rho}$ and the assignment of remaining points based on the nearest distance δ. The computation of $\widetilde{\rho}$ lies in the Euclidean distance between sample points and the determination of d_c neighborhood, whose computing complexity is $O(n^2)$. The assignment strategy of the remaining points based on nearest distance δ employs the classical sorting algorithm, whose computing complexity is $O(n^2)$.

4. Experiments

In this section, 8 synthetic and 7 real data sets are employed to test the new proposed algorithm. The data sets used are greatly different from each other in density distribution, scale, shapes, and so on. Among those data sets, DS1–DS5, aggregation, compound, and flame are synthetic two-dimensional data sets, which are shown in Figure 1. And the 7 real data sets are from UCI machine learning repository.

In the experiment, the clustering results of RDO-DPC are compared with that of the classical DPC. Both algorithms, RDO-DPC and DPC, need the setting of cut-off distance d_c, which is defined as the distance at p% in the ascending sequence of all distances among samples.

The clustering results are measured with AMI (adjusted mutual information) and ARI (adjusted Rand index) [35]. The value range of the two indexes is [0, 1], and the larger the value is, the better the clustering result is. Besides, the clustering results of two-dimensional synthetic data sets are labelled with different colors, and the centers are labelled with red star to give clear view of the results. The results shown in this section are both the best results from RDO-DPC and DPC with best parameters. In this way, the algorithms are better judged concerning their adaptability to data sets of different types and clustering effectiveness.

Eight two-dimensional synthetic data sets are employed to test the clustering efficiency of RDO-DPC and DPC. Both the two algorithms found centers quickly and assigned the reminder samples effectively. Some comparative

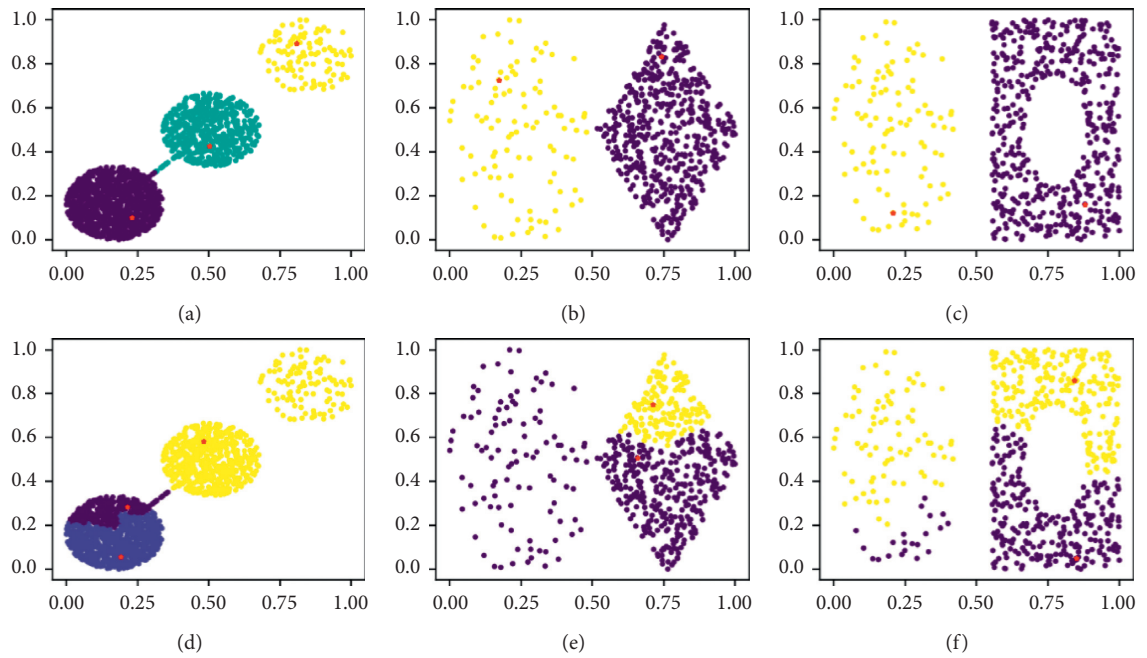

Figure 2: Some comparative clustering results of RDO-DPC and DPC with proper parameters on some synthetic data sets: (a) RDO-DPC on DS2; (b) RDO-DPC on DS3; (c) RDO-DPC on DS4; (d) DPC on DS2; (e) DPC on DS3; (f) DPC on DS4.

Input:
Sample matrix $\mathbf{X} \in \mathbf{R}^{n \times m}$ and cut-off ratio parameter $\mathbf{p}$
Output:
Clustering label $\mathbf{y} \in \mathbf{R}^{n}$
(1) Calculate distance matrix
(2) Calculate relative local density ρ_i according to equation (4)
(3) Calculate distance δ_{q_i} with equation (6)
(4) Draw decision graph and select cluster centers
(5) Assign points to centers according to the nearest distance principle
(6) Clustering result

Algorithm 1: RDO-DPC.

visualization results of synthetic data sets are shown in Figure 2, in which sparse clusters can be recognized, and excessive clustering can be avoided.

The validation of the comparative clustering results of the 8 synthetic data sets is shown in Table 1, which includes ARI, AMI, and their variances. The variances of ARI and AMI are expressed as "ARI.Var" and "AMI.Var" in the table. With different parameter p, AMI and ARI varied more or less. The variance in the accuracy with different parameter p is given in the table to show the validation of the RDO-DPC algorithm. Furthermore, the ARI and AMI listed here in the table are the best with the proper parameter p. Compared with the DPC algorithm, RDO-DPC exhibits superior performance in clustering of data sets with extremely large density differences among clusters and with various shapes.

The comparison of the quantitation indexes between RDO-DPC and DPC shows obvious superiority of RDO-DPC. RDO-DPC is slightly lower than DPC in clustering indexes of DS1 but was apparently higher than DPC in indexes of other data sets. The superior performance of RDO-DPC is because of its employment of relative density in clustering of data sets with uneven density among clusters. Therefore, RDO-DPC can recognize cluster centers more effectively and correctly and assign the remaining points correctly, thus achieving better clustering results than DPC.

Seven real data sets from UCI machine learning repository are employed to test the performance of RDO-DPC and classical DPC. These benchmark data sets include data of high dimensions, complicated structures, and various shapes. With different parameter p, the efficiency of the two algorithms varies slightly. AMI and ARI are employed to measure the different clustering results, and the variance in the accuracy and best parameters are listed in Table 2.

Table 1: Clustering results measured with AMI and ARI on eight synthetic data sets.

Algorithm	RDO-DPC					DPC		
	ARI	ARI.Var	AMI	AMI.Var	p	ARI	AMI	p
DS1	0.988	0.010	0.980	0.004	10.0	0.994	0.989	2.0
DS2	0.979	0.000	0.959	0.000	12.8	0.585	0.606	2.0
DS3	1.000	0.035	1.000	0.013	8.0	—	0.095	2.0
DS4	1.000	0.003	1.000	0.004	8.1	0.015	0.041	1.5
DS5	0.904	0.000	0.829	0.000	10.0	0.691	0.696	1.5
Aggregation	0.895	0.002	0.882	0.000	13.0	0.755	0.860	2.0
Compound	0.789	0.008	0.773	0.002	12.5	0.546	0.697	2.0
Flame	0.476	0.002	0.421	0.003	13.0	0.327	0.403	2.0

ARI.Var and AMI.Var represent the variance in ARI and AMI, respectively.

Table 2: Clustering results measured with AMI and ARI on real data sets.

Algorithm	RDO-DPC					DPC		
	ARI	ARI.Var	AMI	AMI.Var	p	ARI	AMI	p
Iris	0.759	0.008	0.793	0.002	13.0	0.720	0.767	2.0
Seeds	0.787	0.005	0.736	0.002	6.8	0.734	0.717	2.0
Wine	0.742	0.002	0.746	0.000	2.0	0.672	0.706	2.0
E. coli	0.796	0.002	0.699	0.001	9.5	0.309	0.443	2.0
Wdbc	0.850	0.001	0.762	0.001	10.5	—	—	1.0
Zoo	0.883	0.040	0.771	0.015	8.0	0.363	0.321	2.0
Glass	0.266	0.000	0.378	0.001	4.5	0.224	0.246	1.0

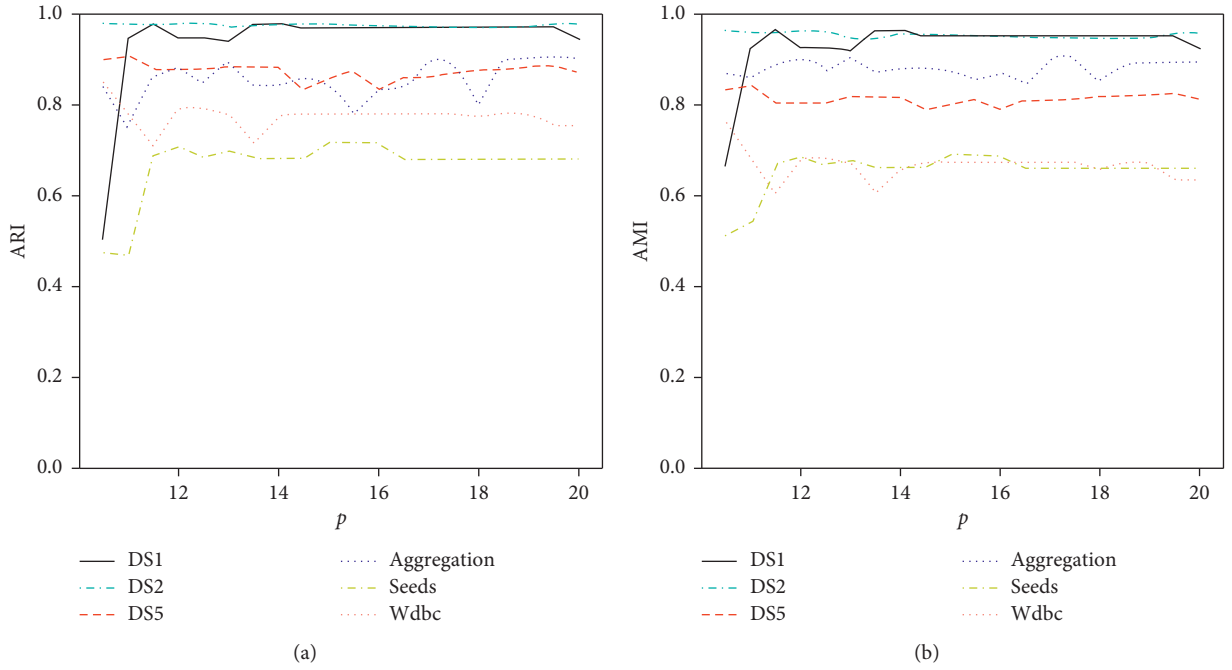

Figure 3: Parameter sensitivity analysis (measured with ARI and NMI) of RDO-DPC on some synthetic and real data sets.

From Table 2, the contrastive results of the two algorithms real data sets show the superior performance of the proposed RDO-DPC, which can find the center and meaningful group of real data sets. Especially for data set Wdbc, DPC could not find cluster centers and recognize meaningful groups of the data set because of its deficiency in clustering of high-dimensional data. ARI and AMI of the proposed RDO-DPC shows that RDO-DPC performs well on high-dimensional data.

The robustness of the new algorithm is also considered. In RDO-DPC, d_c is important because it is used to determine the relative density of each sample, which has impact on

many critical steps in clustering. The value of d_c is closely related with parameter p, which means d_c determines the performance of RDO-DPC.

Figure 3 lists the influence of different values of p on ARI and AMI of some synthetic and real data sets. The robust interval of p is suggested to be set from 10 to 20 for the proposed algorithm in the experiments. As shown in Figure 3, the accuracy of new algorithm remains stable overall with respect to p.

The above comparative results on synthetic and real data sets show that the new proposed algorithm RDO-DPC is effective in the clustering of data sets with extremely large density differences among clusters and with various shapes. And the algorithm is robust overall. In terms of data sets with low number of records and huge number of features, the new algorithm also shows certain efficiency although clustering on such data sets is difficult.

5. Conclusions

Based on neighborhood information of samples, relative density is introduced in this paper. The introduced relative density is used to describe the relative density between each sample and the samples around it and takes full advantage of the information of adjacent samples, thus facilitating the effective find of centers and distinction of clusters of different densities. In addition, the assignment strategy of the original DPC is also improved. The experiments on different types of data sets show that the proposed algorithm can perform effectively on data sets with arbitrary shapes, uneven density, and high dimensions, avoiding the mistaken assignment of samples of the original DPC. Compared with classical DPC, the proposed RDO-DPC not only considers the local density of the samples but also the relative density, which enables RDO-DPC to cluster data sets with uneven density with a higher efficiency. For further research, the reduction of calculation complexity is still an important problem.

Data Availability

The 7 real data sets used in this paper are from UCI machine learning repository. The other data sets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study presented in this article was supported by the National Science Foundation of China (Grant nos. 61305070 and 61703001).

References

[1] L. Bai, X. Cheng, J. Liang, H. Shen, and Y. Guo, "Fast density clustering strategies based on the k-means algorithm," *Pattern Recognition*, vol. 71, pp. 375–386, 2017.

[2] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?" *Pattern Recognition*, vol. 93, pp. 95–112, 2019.

[3] Kanika, R. Kanchan, Sangeeta, and Preeti, "Visual analytics for comparing the impact of outliers in k-means and k-medoids algorithm," in *Proceedings of the 2019 Amity International Conference on Artificial Intelligence*, IEEE, Dubai, UAE, February 2019.

[4] S. A. Mondal, "An improved approximation algorithm for hierarchical clustering," *Pattern Recognition Letters*, vol. 104, no. 1, pp. 23–28, 2018.

[5] B. Perret, J. Cousty, S. J. Ferzoli Guimarães, Y. Kenmochi, and L. Najman, "Removing non-significant regions in hierarchical clustering and segmentation," *Pattern Recognition Letters*, vol. 128, no. 1, pp. 433–439, 2019.

[6] T. Boonchoo, X. Ao, Y. Liu, W. Zhao, F. Zhuang, and Q. He, "Grid-based DBSCAN: indexing and inference," *Pattern Recognition*, vol. 90, pp. 271–284, 2019.

[7] C. Deng, J. Song, R. Sun, S. Cai, and Y. Shi, "GRIDEN: an effective grid-based and density-based spatial clustering algorithm to support parallel computing," *Pattern Recognition Letters*, vol. 109, pp. 81–88, 2018.

[8] G. Celeux, C. Maugis-Rabusseau, and M. Sedki, "Variable selection in model-based clustering and discriminant analysis with a regularization approach," *Advances in Data Analysis and Classification*, vol. 13, no. 1, pp. 259–278, 2019.

[9] M. Liang, Q. Li, Y. Geng, J. Wang, and Z. Wei, "REMOLD: an efficient model-based clustering algorithm for large datasets with spark," in *Proceedings of the IEEE International Conference on Parallel & Distributed Systems*, IEEE Computer Society, Shenzhen, China, December 2017.

[10] I. Gialampoukidis, S. Vrochidis, I. Kompatsiaris, and I. Antoniou, "Probabilistic density-based estimation of the number of clusters using the DBSCAN-martingale process," *Pattern Recognition Letters*, vol. 123, pp. 23–30, 2019.

[11] S. Pourbahrami, L. M. Khanli, and S. Azimpour, "Improving neighborhood construction with apollonius region algorithm based on density for clustering," *Information Sciences*, vol. 522, pp. 227–240, 2020.

[12] M. Yousefnezhad, A. Reihanian, D. Zhang, and B. Minaei-Bidgoli, "A new selection strategy for selective cluster ensemble based on diversity and independency," *Engineering Applications of Artificial Intelligence*, vol. 56, pp. 260–272, 2016.

[13] M. Mojarad, H. Parvin, S. Nejatian, and V. Rezaie, "Consensus function based on clusters clustering and iterative fusion of base clusters," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 27, no. 1, pp. 97–120, 2019.

[14] M. Richard and D. David, "Clustering analyses methods: strategies and algorithms," *Reviews in Theoretical Science*, vol. 4, no. 2, pp. 153–158, 2016.

[15] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[16] Y. Ju, "Research on manifold-based density peaks clustering algorithm," Thesis, IEEE, Yangzhou, China, 2016.

[17] J. Jiang, D. Hao, Y. Chen, M. Parmar, and K. Li, "GDPC: gravitation-based density peaks clustering algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 502, pp. 345–355, 2018.

[18] J. Jiang, Y. Chen, D. Hao, and K. Li, "DPC-LG: density peaks clustering based on logistic distribution and gravitation," *Physica A: Statistical Mechanics and its Applications*, vol. 514, pp. 25–35, 2019.

[19] X.-H. Yang, Q.-P. Zhu, Y.-J. Huang, J. Xiao, L. Wang, and F.-C. Tong, "Parameter-free Laplacian centrality peaks clustering," *Pattern Recognition Letters*, vol. 100, pp. 167–173, 2017.

[20] R. Mehmood, R. Bie, H. Dawood, and H. Ahmad, "Fuzzy clustering by fast search and find of density peaks," in *Proceedings of the International Conference on Identification, Information and Knowledge in the Internet of Things*, pp. 258–261, IEEE, Beijing, China, 2016.

[21] B. Qiu and L. Cheng, "Parameter-free clustering algorithm based on Laplace centrality and density peaks," *Journal of Computer Applications*, vol. 38, no. 9, pp. 2511–2514, 2018.

[22] Z. Yang, H. Wang, and Y. Zhou, "A clustering algorithm with self-adapting truncation distance and clustering center," *Data Analysis and Knowledge Discovery*, vol. 2, no. 3, pp. 39–48, 2018.

[23] J. Xu, G. Wang, and W. Deng, "DenPEHC: density peak based efficient hierarchical clustering," *Information Sciences*, vol. 373, pp. 200–218, 2016.

[24] J. Ding, Z. Chen, X. He, and Y. Zhan, "Clustering by finding density peaks based on Chebyshev's inequality," in *Proceedings of the Chinese Control Conference*, pp. 7169–7172, IEEE, Chengdu, China, July 2016.

[25] J. Ding, X. He, J. Yuan, and B. Jiang, "Automatic clustering based on density peak detection using generalized extreme value distribution," *Soft Computing*, vol. 22, no. 9, pp. 2777–2796, 2017.

[26] J. Jiang, Y. Chen, X. Meng, L. Wang, and K. Li, "A novel density peaks clustering algorithm based on k nearest neighbors for improving assignment process," *Physica A: Statistical Mechanics and its Applications*, vol. 523, pp. 702–713, 2019.

[27] J. Xie, H. Gao, W. Xie, X. Liu, and P. W. Grant, "Robust clustering by detecting density peaks and assigning points based on fuzzy weighted k-nearest neighbors," *Information Sciences*, vol. 354, pp. 19–40, 2016.

[28] R. Liu, H. Wang, and X. Yu, "Shared-nearest-neighbor-based clustering by fast search and find of density peaks," *Information Sciences*, vol. 450, pp. 200–226, 2018.

[29] P. Jia, J. Fan, and Y. Peng, "An improved clustering algorithm by fast search and find of density peaks based on boundary samples," *Journal of Nanjing University(Natural Sciences)*, vol. 53, no. 2, pp. 368–377, 2017.

[30] Y. Liu, Z. Ma, and F. Yu, "Adaptive density peak clustering based on k-nearest neighbors with aggregating strategy," *Knowledge-Based Systems*, vol. 133, pp. 208–220, 2017.

[31] Y. Zhou, K. Ting, and M. J. Carman, "Density-ratio based clustering for discovering clusters with varying densities," *Pattern Recognition*, vol. 60, pp. 983–997, 2016.

[32] M. Du, S. Ding, and H. Jia, "Study on density peaks clustering based on k-nearest neighbors and principal component analysis," *Knowledge-Based Systems*, vol. 99, pp. 135–145, 2016.

[33] X. Xu, Y. Ju, Y. Liang, and P. He, "Manifold density peaks clustering algorithm," in *Proceedings of the 3th International Conference on Advanced Cloud and Big Data*, pp. 311–318, IEEE, Yangzhou, China, 2016.

[34] L. A. R. Calla, L. J. F. Perez, and A. A. Montenegro, "A minimalistic approach for fast computation of geodesic distances on triangular meshes," *Computers and Graphics*, vol. 84, pp. 77–92, 2019.

[35] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: is a correction for chance necessary?" in *Proceedings of the Annual International Conference on Machine Learning ICM L'09*, pp. 1073–1080, Montreal, Canada, June 2009.