

Research Article

Heuristic Clustering Based on Centroid Learning and Cognitive Feature Capturing

Chunzhong Li , Yunong Zhang, and Xu Chen

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

Correspondence should be addressed to Chunzhong Li; l20120038@aufe.edu.cn

Received 15 January 2019; Revised 22 March 2019; Accepted 27 March 2019; Published 8 April 2019

Academic Editor: Oscar Reinoso

Copyright © 2019 Chunzhong Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As one of the typical clustering algorithms, heuristic clustering is characterized by its flexibility in feature integration. This paper proposes a type of heuristic algorithm based on cognitive feature integration. The proposed algorithm employs nonparameter density estimation and maximum likelihood estimation to integrate whole and local cognitive features and finally outputs satisfying clustering results. The new approach possesses great expansibility, which enables priors supplement and misclassification adjusting during clustering process. The advantages of the new approach are as follows: (1) it is effective in recognizing stable clustering results without priors given in advance; (2) it can be applied in complex data sets and is not restricted by density and shape of the clusters; and (3) it is effective in noise and outlier recognition, which does not need elimination of noises and outliers in advance. The experiments on synthetic and real data sets exhibit better performance of the new algorithm.

1. Introduction

Clustering is an automatic process of partitioning data set with proper similarity measurement. In the process, data in same group have maximum similarity, while data in different group have minimum similarity. As unsupervised learning, clustering is greatly influenced by similarity measurement and is closely related with priors in application fields. Clustering is extensively applied in fields such as biology [1, 2], computer vision [3, 4], geological exploration [5–7], and information retrieval [8], because it shows excellent advantages in automatic grouping.

Clustering is developed into many types of algorithms in its successful application in those fields. According to different clustering processes, it is generally classified into two types: agglomerative clustering and partitional clustering. Agglomerative clustering initially sees all data as one group and then conducts the fusion of clusters based on certain principles, thus forming proper data groups. Partitional clustering redistributes the existing data groups and forms proper clusters, such as k-means [9] and fuzzy c-means [5–7].

According to different similarity measurements, clustering can also be subdivided into many types: local density-based clustering, clustering based on density estimation, clustering based on matrix calculation, clustering based on graph calculation, grid-based clustering, and so on.

Clustering based on local density feature is a typical clustering algorithm. Simple calculation and high efficiency are the advantages of this type of algorithm, such as clustering with density peaks (CDP) [10] and its generalized version [11, 12]. This type of algorithms endows every datum with a local density feature and redefines centroid, or defines cluster merging mechanism, to obtain clustering results. Such algorithm is effective, but it is difficult to find centroid with poor decision graph for some complicated data sets.

Density-based clustering algorithm constitutes a significant proportion of the clustering research, such as mean shift [13] and scale space filtering (CSSF) [14]. This type of algorithms employs nonparameter density estimation to estimate distribution of the data set. Area of high density represents centroids, while area of relatively low density represents boundaries. Then, centroids and boundaries are found with gradient descent or level set, and data are

automatically divided into groups. Density-based algorithm is effective in dealing with Gaussian type data sets and is widely applied in protein sequencing and computer vision. Such algorithms depend highly on density estimation and gradient descent, so they are not effective in clustering of data set with arbitrary density feature and arbitrary shape.

One typical algorithm of clustering is spectral clustering [15, 16] and subspace clustering [17]. Spectral clustering constructs similarity Laplace matrix and conducts dimensional reduction through eigenvalue decomposition. Finally, clustering result is obtained through k-means in low-dimensional space. Spectral clustering fully employs characteristic mapping and dimensional reduction, which enables the algorithm to be effective in dealing with manifold structure data set. However, its high computing complexity and sensitivity to parameters make it difficult in its application into large scale data sets.

There are two problems for clustering: one is definition of similarity, including Euclidean and non-Euclidean dissimilarity and the other is similarity treatment. Different algorithms employ different methods for similarity treatment. The first problem is subjective because it is greatly related with priors in application fields. The second problem lies in centroid estimation or boundary description. For most clustering algorithms, some clustering mistakes in the middle of the process will go down to the final clustering result, which means that the ability of error correction needs to be improved.

This paper proposes a heuristic clustering algorithm based on cognitive feature capturing. The proposed algorithm is a centroid learning process and captures data structure features with a new type of similarity measurement to obtain better clustering result. In feature description, three cognitive features are described: neighbourhood, density difference, and connectivity. Similarity measurement (a new kernel function) is established to capture the three mentioned features. In clustering, the paper proposes a heuristic algorithm based on centroid learning. The proposed algorithm possesses great expansibility, which enables priors supplement and misclassification adjusting during clustering process. The proposed algorithm, to some extent, weakens the subjective dependence on similarity measurement and allows misclassification adjusting and prior interference in the process of clustering.

In Section 2, we introduce a new similarity measurement model and explain its way of capturing clustering structures. Section 3 mainly establishes a heuristic clustering model based on local centroid learning. Execution of the algorithm and grouping strategy are given in Section 4. Sections 5 and 6 present experiments and conclusion.

2. Similarity Definition for Cognitive Feature

In clustering, cognitive features are of great significance for identification of clusters, such as neighbourhood, density difference, and connectivity. Samples of closer distance tend to be recognized as in one cluster. Samples of greater density difference tend to be recognized as in different clusters, while samples of little density difference can extend continuously.

The capturing of those features helps in the recognition of clusters, which always depends on similarity measurement.

For data set $X=\{x_i \in R^d \mid i = 1, 2, \dots, n\}$, the similarity measurement of an arbitrary point x_i with its neighborhood can be regarded as probability of x_j belonging to local centroid x_i and is expressed as

$$p(x_j \mid x_i) = \frac{I(x_j \mid x_i, \alpha)}{Z} \exp \left\{ -\frac{\text{dist}^2(x_j - x_i)}{\sigma^2(x_i)} \right\} \quad (1)$$

where Z is normalization coefficient, so as to satisfy $\int_{-\infty}^{+\infty} p(x)dx = 1$. $\text{dist}(x_j, x_i)$ represents the Euclidean distance between two data points. $\sigma(x_i)$ is a function of x_i and is defined as

$$\sigma(x_i) = \max \{ \epsilon, r(\text{ENN}(x_i)) \} \quad (2)$$

where $\epsilon > 0$, $r(\text{ENN}(x_i))$ is radius of the effective nearest neighbourhood and $\text{ENN}(x_i)$ is the effective nearest neighborhood of x_i :

$$\begin{aligned} \text{ENN}(x_i) \\ = \left\{ y \in \text{NN}(x_i) \mid \text{dist}(y, x_i) < \frac{1}{2}r(\text{NN}(x_i)) \right\} \end{aligned} \quad (3)$$

where $\text{NN}(x_i)$ is nearest neighbourhood of x_i . $r(\text{ENN}(x_i))$ in Equation (2) is defined as

$$r(\text{ENN}(x_i)) = \begin{cases} r(\text{NN}(x_i)), & \text{ENN}(x_i) \neq \emptyset \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

The exponent section in Equation (1) is a type of radial basis function. Combined with local scale parameter, such function can measure the similarity of neighbours with different density features. With certain σ , x_j and x_i are of greater similarity when Euclidean distance $\|x_j - x_i\|$ between them is smaller. Equation (2) gives the affecting region of centroid x_i . When all neighbors duplicate together, scale is set as a nonzero small constant ϵ to avoid denominator being zero. Meanwhile, the effective nearest neighbourhood in Equation (2) is able to eliminate the influence from noises and outliers. When $\text{ENN}(x_i) = \emptyset$, $\sigma = \epsilon$, and x_i has weaker relation with other samples, which means x_i can be recognized as noise and outlier.

In Equation (1), $I(x_j, x_i \mid \alpha)$ can be expressed as

$$I(x_j, x_i \mid \alpha) = \begin{cases} 1 & \frac{|\rho(x_j) - \rho(x_i)|}{\rho(x_j) + \rho(x_i)} < 1 - \alpha \\ 0 & \text{Otherwise} \end{cases} \quad (5)$$

where α represents significant difference between two scalars. According to significance testing in statistics, α is set as 5% or 10% in common. Local density can be obtained through neighbours:

$$\rho(x \mid k) = \frac{1}{k} \sum_{y \in \text{NN}(x)} \text{dist}(x, y) \quad (6)$$

Equation (5) can weaken the similarity between points with greater density difference, so as to recognize noises and boundaries, and to continue the connectivity through points with smaller density difference.

Similarity measurement (1) can better capture features like nearest neighbourhood, density difference, and connectivity in clustering and can be applied in heuristic clustering to obtain satisfying results.

3. Procedure of Centroid Estimation

As one of the most important types in clustering algorithms, centroid estimation continuously searches for the centroids of local area or of the whole data set. In data set X , an arbitrary sample x_i can be regarded as the initial centroid, and $p(x_j | x_i)$ is the probability of the nearest neighbor x_j belonging to centroid x_i . Then, the probability distribution of sample x_j is expressed as

$$P(x_j | x_i) = \sum_{i=1}^n \frac{1}{n} p(x_j | x_i) \quad (7)$$

If x_i is independent identical distribution variable, distribution of data set X can be expressed as

$$P(X | x_i) = \prod_{j=1}^n P(x_j | x_i) = \prod_{j=1}^n \frac{1}{n} \sum_{i=1}^n p(x_j | x_i) \quad (8)$$

Log likelihood can be expressed as

$$LL(X) = \ln P(X | x_i) = \sum_{j=1}^n \left\{ \ln \frac{1}{n} + \ln p(x_j | x_i) \right\} \quad (9)$$

Partial log likelihood with respect to x_i is

$$\frac{\partial LL(X)}{\partial x_i} = \sum_{j=1}^n \frac{p(x_j | x_i)}{\sum_{i=1}^n p(x_j | x_i)} \frac{1}{\sigma^2} (x_j - x_i) \quad (10)$$

Although $\sigma(x_i)$ is a function in Equation (10), it is computed once in advance and can be viewed as constants. Let $\partial LL(X)/\partial x_i = 0$; estimated value of the centroid can be expressed as

$$\hat{x}_i = \sum_{j=1}^n \frac{p(x_j | x_i)}{\sum_{i=1}^n p(x_j | x_i)} x_j \quad (11)$$

Equation (11) is weighed sum of the influence on centroid from the nearest neighbors of initial centroid x_i , so as to estimate the centroid. Probability density function in Equation (11) can be regarded as normalization processing and can be expressed as

$$\hat{x}_i = \sum_{j=1}^n \frac{g(x_j | x_i)}{\sum_{i=1}^n g(x_j | x_i)} x_j \quad (12)$$

where

$$g(x_j | x_i) = I(x_j | x_i, \alpha) \exp \left\{ -\frac{\text{dist}(x_j, x_i)^2}{\sigma^2(x_i)} \right\} \quad (13)$$

Equation (12) is a process of estimating local centroids, and Equation (13) is a generalized radial basis function. Equations (12) and (13) can be developed into a kind of heuristic clustering algorithm including non-Euclidean similarity and centroid learning.

4. Heuristic Clustering based on Centroid Learning

Combination of local centroid estimation (12) and similarity measurement (13) can make good use of the information of nearest neighborhood, local density, and connectivity and obtain satisfying clustering result effectively.

The new heuristic clustering algorithm is a process of searching for local density maximum (centroid) and is expressed as

$$x_i(t+1) = \sum_{j=1}^n \frac{g(x_j | x_i)}{\sum_{i=1}^n g(x_j | x_i)} x_j(t) \quad (14)$$

where t is iteration steps, and initial sample $x_i = x_i(0)$. In the process of clustering, each sample x_i is an initial centroid, and each centroid can also be neighbors of other centroids.

With gradient (8), the proposed heuristic clustering algorithm can be expressed as

$$\begin{aligned} x_i(t+1) \\ = x_i(t) + \gamma \sum_{j=1}^n \frac{g(x_j | x_i)}{\sum_{i=1}^n g(x_j | x_i)} (x_j(t) - x_i(t)) \end{aligned} \quad (15)$$

where γ is the step size. Different from the adaptive step size in Equation (14), the step size in Equation (15) has to be set manually. However, the fusion of feature information can be controlled more easily in the heuristic clustering.

Heuristic clustering algorithms (14) and (15) are essentially different expressions of a same method. The two equations are both the process of continuous centroid estimation. The convergence of the algorithms depends on the similarity measurement matrix. Clustering process of (14) can be expressed with matrix as

$$X(t+1) = \hat{S} \cdot X(t) \quad (16)$$

where \hat{S} is row normalization matrix based on Equation (13). Since similarity matrix \hat{S} is positive definite matrix, and row sum is 1, the eigenvalue of the matrix satisfies

$$0 \leq \lambda_n \leq \lambda_{n-1} \leq \dots \leq \lambda_1 = 1 \quad (17)$$

Each sample x can be expressed as $x = a_1 e_1 + a_2 e_2 + \dots + a_n e_n$ in vector space spanned by eigenvector. After t calculations, specific sample x can be expressed as

$$\begin{aligned} x(t) &= \widehat{S} \cdot x(t-1) = \dots = \widehat{S}^t \cdot x(0) \\ &= \widehat{S}^t \cdot (a_1 e_1 + a_2 e_2 + \dots + a_n e_n) \\ &= a_1 \left[1 + \sum_{j=1}^n \frac{a_j}{a_1} \lambda_j^t e_j \right] \end{aligned} \quad (18)$$

After t iterations, distance between every two arbitrary points can be expressed as

$$\begin{aligned} \|x_i(t) - x_j(t)\| &= \left\| [e_1(i) - e_1(j)] a_1 \lambda_1^t + \sum_{p=2}^n [e_p(i) - e_p(j)] a_p \lambda_p^t \right\| \\ &\leq \| [e_2(i) - e_2(j)] a_2 \lambda_2^t \| + \| [e_3(i) - e_3(j)] a_3 \lambda_3^t \| \\ &\quad + \dots + \| [e_n(i) - e_n(j)] a_n \lambda_n^t \| \end{aligned} \quad (19)$$

According to Equation (16), when similarity matrix \widehat{S} is positive semidefinite matrix,

$$\lim_{t \rightarrow \infty} \|x_i(t) - x_j(t)\| = 0 \quad (20)$$

Heuristic clustering algorithm (14) or (15) based on feature capturing generates clusters gradually in continuous clustering process. The l cluster can be expressed as

$$\begin{aligned} C_l(X(t)) &= \{x_i(t), x_j(t) \mid \|x_i(t) - x_j(t)\| \leq \epsilon, i, j \\ &= 1, 2, \dots, n\} \end{aligned} \quad (21)$$

In clustering, number of clusters reduces constantly until only one or needed number of clusters is reserved. In the process, since the variation of clusters is regular, some of the clusters remain stable for a rather long period, which is called survival period of the cluster and is expressed as

$$\begin{aligned} \text{life}(C_l(X)) \\ = \operatorname{argmax}_m \{NMI(C_l(X(t)), C_l(X(t+m))) = 1\} \end{aligned} \quad (22)$$

where NMI (normalized mutual information) [18] is clustering index and can evaluate the clustering result. $NMI = 1$ represents that the two clustering results are the same, or that clustering result remains stable for a long period.

5. Experiments

Heuristic clustering algorithm based on centroid learning is a process of centroid estimation, in which the centroids of local area are searched constantly until the satisfying results are obtained. Centroid learning process makes good use of clustering structure information of the data set and performs better in dealing with data sets of manifold structure and various density structures. In this section, the proposed algorithm is tested with synthetic data sets and real data sets

and is also applied to volume rendering to show its good performance.

For heuristic clustering algorithm based on centroid learning, nearest neighbourhood parameter k is a key parameter. Although clustering result is not sensitive to k , it is influenced by k to some extent. If k is extremely small, much more clusters will be obtained because of emphasis on detailed structures of data set, while extremely large k will form only one big cluster because of ignoring of details. Extensive experiments show that $k \in [5, 30]$ for medium data sets, and k is set as 2% of the data set scale for large data sets. Other two parameters are significance testing parameter α and small constant ϵ . ϵ is similar to terminal condition and is an arbitrary constant, and α is a parameter in significance test, which is often set as 5%.

The clustering process of a data set is shown in Figure 1, in which the data set contains 5 satisfying clusters. The cluster number decreases in the clustering procedure, while the consideration of detailed structure information is weakened. Usually, better clustering result is often obtained when whole and local structure features are considered with proper compromise. In clustering process of the new approach, 5 is given as the most stable number of clusters, which is in line with the cognitive recognition of the data set. The clustering result with 6 clusters can also be viewed as a stable result, in which an outlier is regarded as a cluster.

Four synthetic data sets of various density and shape structures are employed to test the efficiency of the new algorithm. The four data sets in Figures 2(a), 2(b), 2(c), and 2(d) are named as ‘‘Airplane,’’ ‘‘Anchor,’’ ‘‘Ring,’’ and ‘‘Swissroll,’’ respectively. The nearest neighbourhood k is set as 8, 8, 8, and 20. α is set as 0 because the algorithm does not need noise and outlier recognition. Figure 2 shows the clustering results of the new algorithm. Table 1 shows the clustering results of other algorithms, including scale space filtering clustering (CSSF) [14], chameleon [19], Spectral-Ng [15], and clustering with density peaks (CDP) [10]. Scale space filtering clustering is based on multiscale feature recognition. Chameleon and spectral algorithms perform better in manifold data set clustering. Although CDP is efficient in clustering of many types of data sets, it is difficult to select the parameters of cut-off, as the parameter is sensitive to decision graph. The comparison of clustering results in Table 1 shows the new algorithm can obtain more satisfying results.

Four real data sets are employed to test the new clustering algorithm, including pen-digit, iris, noise iris, and noise USPS-01. Pen-digit and iris with whole training data set are from UCI machine learning repository, in which pen-digit data set contains 7494 images with 16 attributes and iris data set contains 150 samples with 4 attributes. Noise iris is composed of iris data set and 5% uniform random samples. ‘‘USPS-01’’ data set contains 2200 images which are ‘0’ and ‘1’ classes drawn from USPS handwritten digits, and ‘‘noise USPS-01’’ is ‘‘USPS-01’’ contaminating with 2% residual USPS samples.

The clustering results are compared with those of NRSC [20], Spectral-Ng, STSC [21], and chameleon, to show their robustness to noises. The comparison results measured with NMI are shown in Figure 3. The results in Figure 3 show that

TABLE 1: Comparison of clustering results with NMI.

Data sets	New	CSSF	Chameleon	Spectral-Ng	CDP
Airplane	1.00	0.34	1.00	1.00	1.00
Anchor	0.98	0.79	0.85	0.79	0.98
Ring	1.00	-	0.28	1.00	0.82
Swissroll	1.00	-	1.00	0.02	0.34

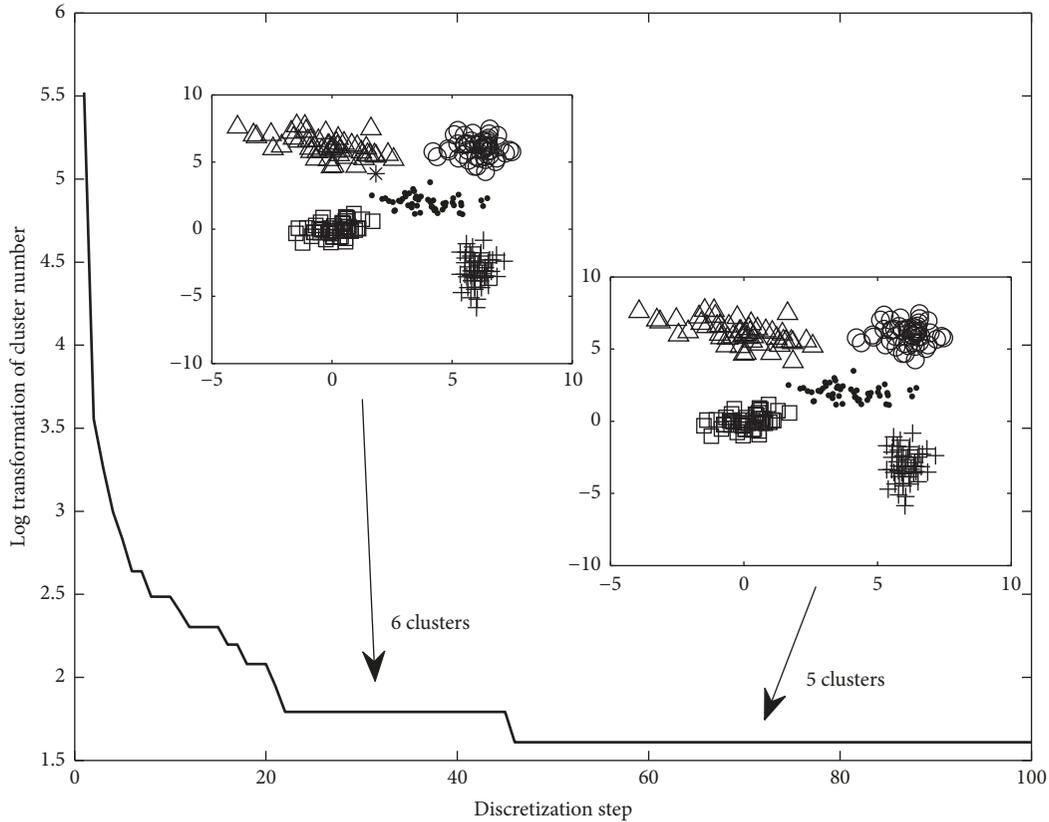


FIGURE 1: Procedure of the clustering. Clustering result (5 clusters) with the longest lifetime is recognized as the final result. The result with 6 clusters is also stable, in which an outlier is recognized as one cluster.

the five algorithms are robust to outliers to some extent. The new algorithm has more advantages compared with other 4 algorithms in noise robustness. The chameleon performs well in pen-digit data set. However, the overall performance of the new clustering algorithm is superior to other four algorithms as shown in Figure 3.

Clustering is applied to volume rendering to improve separability of structures with similar attributes. Volume rendering is a technique used to display a 2D projection of a 3D discretely sampled data set, typically a 3D scalar field. A typical 3D data set is a group of 2D slice images acquired by a CT, MRI, or Micro-CT scanner. Clustering can be used in scalar field to separate similar structures, and the clustering result in scalar field is transformed into visualized image.

The new clustering is employed to separate adjacent structures in two volume rendering examples. One is an engine in industry and the other is keen joints in medical treatment. The results of volume rendering with the new clustering are shown in Figure 4. Figure 4(a) shows part of

adjacent structure (the part in red rectangle in Figure 4(b)) in scalar field. Scanning techniques leads to outliers around adjacent parts in scalar field, which is difficult for clustering. Figure 4(b) shows the visualization of separating similar structures, where different parts are labelled with different colours according to clustering results in scalar field. Figure 4(c) shows the visualization of keen joints with CT scanner, and the visualization of separating structures is shown in Figure 4(d). The experiment on volume rendering exhibits good performance of the new approach.

6. Conclusions

The paper proposes a heuristic clustering algorithm based on cognitive feature capturing. The new approach is effective in integrating cognitive features of data set and conducts clustering based on the features. The proposed approach is not restricted by density and shape features of data set, especially manifold structure

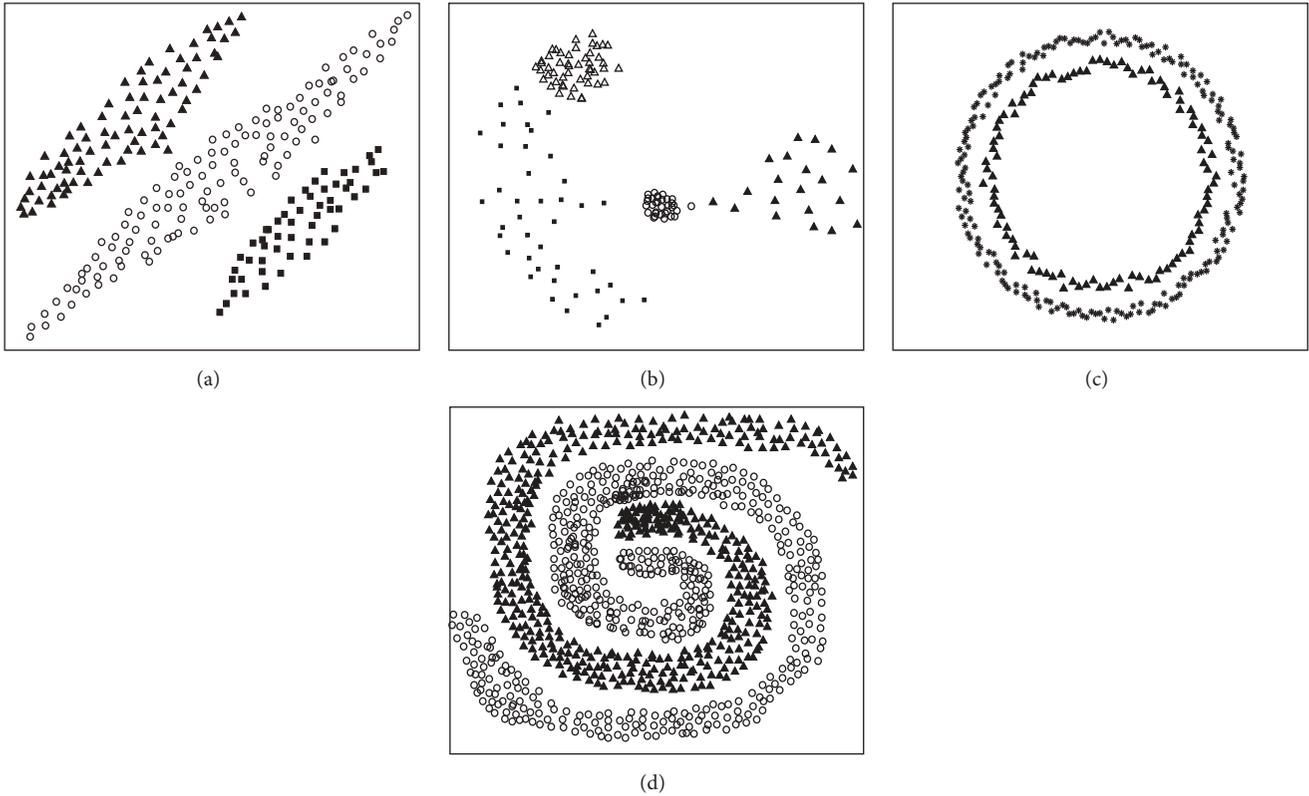


FIGURE 2: Clustering results of the four synthetic data sets by the new clustering. (a) Result with “Airplane” data set. (b) Result with “Anchor” data set. (c) Result with “Ring” data set. (d) Result with “Swissroll” data set. The four data sets have typical cognitive features, such as density difference, connectedness, and manifolds.

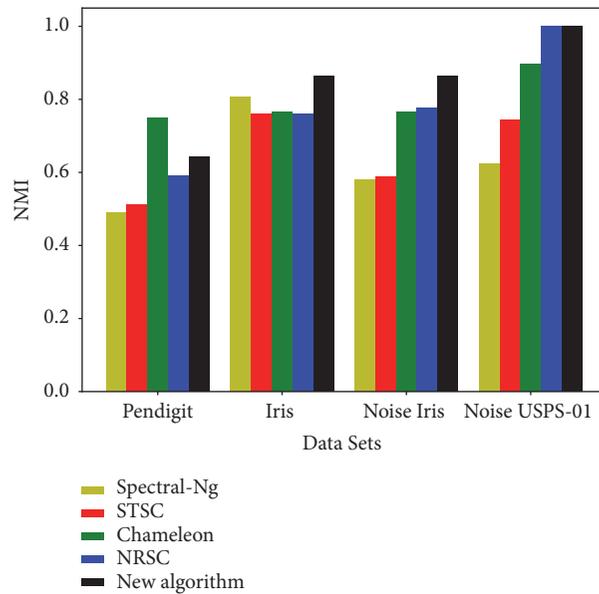


FIGURE 3: Comparison of clustering results measured by NMI.

data after dimensionality reduction. Heuristic clustering shows advantage in integration of data structure and is worth further research concerning feature information feedback and effective usage of data features in clustering process.

Data Availability

The pen-digits and iris data sets are from UCI machine learning repository, and USPS hand written digits data set is from <https://cs.nyu.edu/~roweis/data.html>. The other data

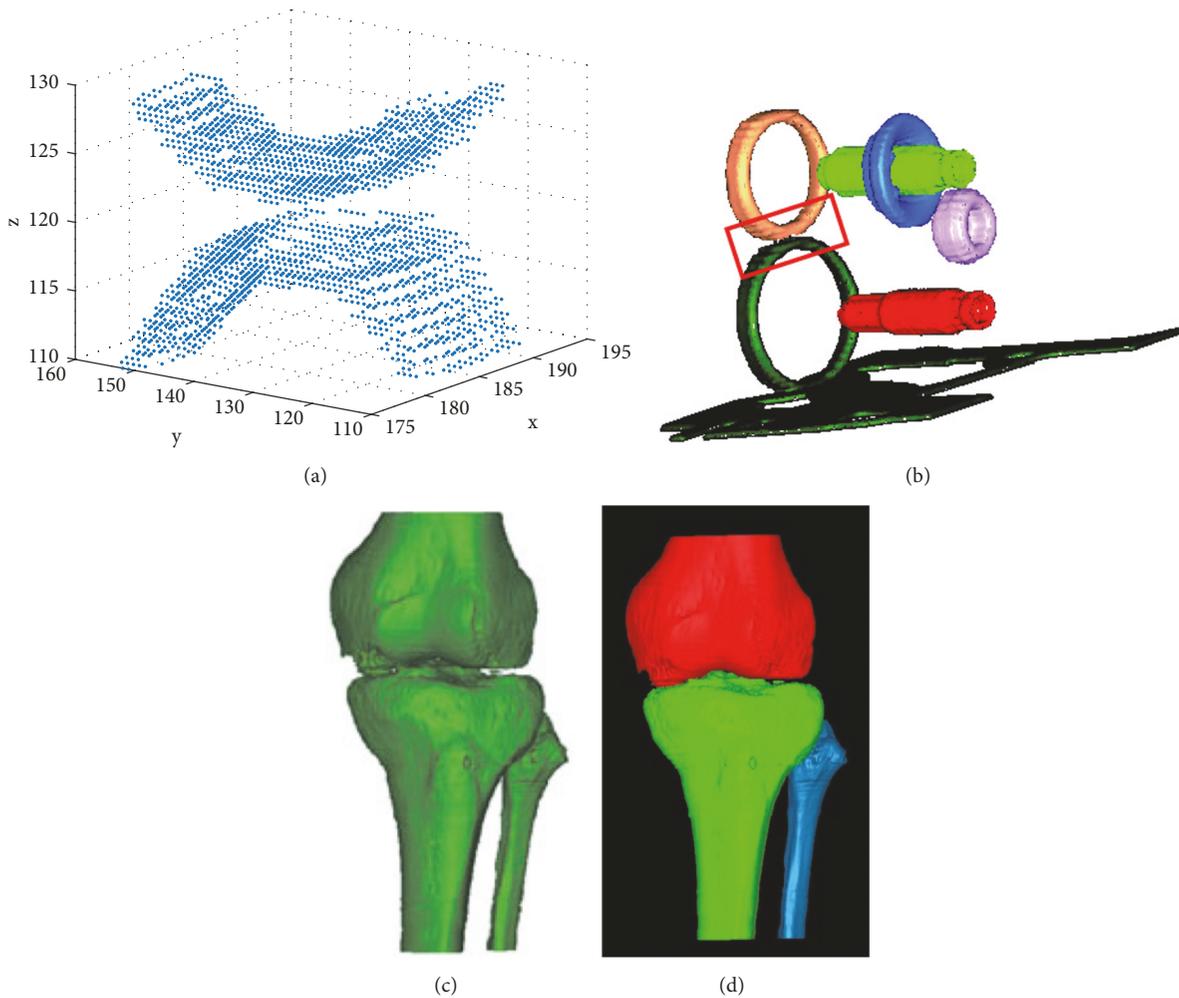


FIGURE 4: Two clustering examples on volume rendering to separate structures with similar attributes. (a) A part of 3D scalar field, which is labelled with red rectangle in (b). (b) The visualization of clustering results in scalar field. The visualization inside the red rectangle is obtained with clustering data points in (a). (c) The visualization of knee joints with CT scanning. (d) The visualization of knee joints by clustering data points in 3D scalar field.

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 is supported by the National Science Foundation of China, Research Grants no. 61305070 and no. 61703001.

References

- [1] A. Sudipta, S. Sriparna, P. Prasanna et al., "Novel symmetry-based gene-gene dissimilarity measures utilizing gene ontology: application in gene clustering," *Gene*, vol. 769, pp. 341–351, 2018.
- [2] Z. Zareizadeh, M. S. Helfroush, A. Rahideh, and K. Kazemi, "A robust gene clustering algorithm based on clonal selection in multiobjective optimization framework," *Expert Systems with Applications*, vol. 113, pp. 301–314, 2018.
- [3] L. Zhang and B. Verma, "Roadside vegetation segmentation with adaptive texton clustering model," *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 159–176, 2019.
- [4] M. Mendonça and L. Oliveira, "ISEC: iterative oversegmentation via edge clustering," *Image and Vision Computing*, vol. 80, pp. 45–57, 2018.
- [5] J. Sun and Y. Li, "Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering," *Geophysics*, vol. 80, no. 4, pp. 1–18, 2014.
- [6] J. Sun and Y. Li, "Joint inversion of multiple geophysical data using guided fuzzy c-means clustering," *Geophysics*, vol. 81, no. 3, pp. 37–57, 2016.
- [7] E. Maag and Y. Li, "Discrete-valued gravity inversion using the guided fuzzy c-means clustering technique," *Geophysics*, vol. 83, no. 4, pp. G59–G77, 2018.

- [8] S. Bhatia and J. Deogun, "Conceptual clustering in information retrieval," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 28, no. 3, pp. 427–436, 1998.
- [9] S. Yu, S. Chu, C. Wang, Y. Chan, and T. Chang, "Two improved k-means algorithms," *Applied Soft Computing*, vol. 68, pp. 747–755, 2018.
- [10] A. Laio and A. Rodriguez, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [11] 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.
- [12] M. Du, S. Ding, and Y. Xue, "A robust density peaks clustering algorithm using fuzzy neighborhood," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 7, pp. 1131–1140, 2018.
- [13] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [14] Y. Leung, J.-S. Zhang, and Z.-B. Xu, "Clustering by scale-space filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1396–1410, 2000.
- [15] A. Y. Ng, M. Jordan, and Y. Weiss, "On spectral clustering analysis and algorithm," in *Advances in Neural Information Processing Systems*, pp. 849–856, MIT Press, British Columbia, Canada, 2002.
- [16] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [17] Y. Zhu, K. M. Ting, and M. J. Carman, "Grouping points by shared subspaces for effective subspace clustering," *Pattern Recognition*, vol. 83, pp. 230–244, 2018.
- [18] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, vol. 3, no. 3, pp. 583–617, 2003.
- [19] G. Karypis, E.-H. Han, and V. Kumar, "CHAMELEON: a hierarchical clustering algorithm using dynamic modeling," *IEEE Computer*, vol. 32, no. 8, pp. 68–75, 1999.
- [20] Z. Li, J. Liu, S. Chen, and X. Tang, "Noise robust spectral clustering," in *Proceedings of the 11th International Conference on Computer Vision*, pp. 68–75, Rio de Janeiro, Brazil, October 2007.
- [21] M. L. Zelnik and P. Peron, "Self-tuning spectral clustering in," *Advances in Neural Information Processing Systems*, pp. 1601–1608, 2005.

