Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9834921, 1 page
https://doi.org/10.1155/2023/9834921

Retraction

Retracted: Multidimensional Discrete Big Data Clustering Algorithm Based on Dynamic Grid

Wireless Communications and Mobile Computing

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Li, "Multidimensional Discrete Big Data Clustering Algorithm Based on Dynamic Grid," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4663816, 9 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4663816, 9 pages
https://doi.org/10.1155/2022/4663816

Research Article

Multidimensional Discrete Big Data Clustering Algorithm Based on Dynamic Grid

Xiaolei Li

College of Finance and Information, Ningbo University of Finance and Economics, Ningbo 315175, China

Correspondence should be addressed to Xiaolei Li; lixiaolei@nbufe.edu.cn

Received 13 January 2022; Revised 21 February 2022; Accepted 25 February 2022; Published 12 March 2022

Academic Editor: Mohammad Farukh Hashmi

Traditionally, the data clustering algorithm is lack of comprehensive performance, leading to low clustering purity and long clustering time. In addition, the consistency between the clustering results and the original data distribution is not strong. Therefore, the multidimensional discrete big data clustering algorithm based on dynamic grid was put forward. Firstly, multidimensional discrete big data was processed in advance. The principal component analysis was used to reduce the dimension of data. The concept of entropy was introduced to divide the key attributes and noncritical attributes, so as to extract the key attributes. According to the results of data preprocessing, the dynamic grid was partitioned. According to the results, OptiGrid in data clustering algorithm was used to achieve the data clustering. The experimental results show that the clustering purity of this algorithm is between 95% and 100%, which is significantly higher than the traditional algorithm. Therefore, the multidimensional discrete big data clustering algorithm based on dynamic grid has better comprehensive performance, closer clustering shape to the original data distribution, higher clustering purity, and faster execution efficiency.

1. Introduction

With the rapid development of information technology, Internet and cloud computing, the amount of information is increasing explosively. The way of accessing information becomes simple and efficient. How to efficiently process the information, extract the needful content, and shorten the time from getting information to taking favorable actions becomes a pressing need. The data mining is to solve these problems. In modern society, the big data is full of the society, so the data mining for big data has become a hot research [1]. As an important research method in data mining, it is reasonable to apply the clustering technology to the analysis and research of big data. Therefore, scholars have paid more and more attention to the data clustering problem. Moreover, the research on big data clustering is very important in real life and production.

At present, some research results are given. For example, Reference [2] proposed a data clustering method based on K-means algorithm. This method extracted a lot of data samples from massive data. According to the principle of Euclidean distance similarity based on the best clustering center, the clustering results of evaluation samples were constructed and the suboptimal solution was removed from the clustering results. Then, the evaluation result was weighted, and thus to obtain k clustering centers, which were regarded as the big data clustering center. In the process of clustering, the calculation time was too long. It is necessary to improve this method. In Reference [3], a data clustering method based on rapid regional evolution was proposed. This method was able to reduce the dimension of data. On this basis, the cardinal number of clustering fuzzy membership was used to perform the fuzzy clustering algorithm on cluster data, so as to achieve automatic clustering, but the clustering result of this method was quite different from the actual result. In Reference [4], a clustering method based on representative point was presented. All the data samples were initialized and clustered. According to the deleted invalid clustering data, the average density of clusters was adjusted. The density peak algorithm was used to integrate all data again and update the clustering center point. However, this algorithm has the problem

that the clustering results are not consistent with the original data distribution.

Due to the shortcomings in above methods, a multidimensional discrete big data clustering algorithm based on dynamic grid was put forward. This algorithm divides the grid in neighborhood of each dimension by the data points, and dynamically adjusts the grid structure.

In the subspace clustering stage, the lowest density points of grid unit are found to divide the data space. And then, the clustering subspace is found; the high-density grids are connected in the subspace and thus to complete the clustering. Finally, the programming language is used to implement the above algorithm. The experimental data set contains the real data set and the synthetic data set. The proposed algorithm is tested. The results show that the proposed algorithm is effective in solving their own problems, so it has higher comprehensive performance.

2. Basic Definitions

2.1. Overall Flow of Multidimensional Discrete Big Data Clustering Algorithm Based on Dynamic Grid. As a new data object, the big data is formed with the continuous development of information technology. It is a series of flowing data. A big data is a series of continuous and infinite data, just like a flowing river, and the "water" in river is composed of data. In the present-day world, the application of information technology in various industries has reached a new height, which makes the fields of various societies continue to contribute new data to the outside world. Meanwhile, they are also the recipients of data. The dataflow transformation may bring knowledge with guiding significance to our production activities [5, 6]. With the improvement of people's living standards, mobile communication devices play a more and more important role in people's lives. The communication between people through mobile communication devices is more and more frequent. Thus, a lot of big data is formed. It is necessary to analyze these big data, so as to help enterprises adjust their production activities, improve their competitiveness, allocate resources, and maximize the utilization.

The modern society is a society of information explosion. With the rapid development of information technology, the data collection has become more and more easy, so that the scale of database becomes larger and larger, and the complexity of data also becomes higher and higher. For example, some attributes of trade transaction data may reach hundreds of dimensions, or even higher. Due to the influence of "dimension effect", many clustering methods that perform well in low-dimensional data space are not good in multidimensional space [7]. Traditionally, the clustering method has two difficulties in multidimensional data clustering. On the one hand, the attributes of some data points are too "treasured" in the data space, so they cannot find clusters. On the other hand, the data distribution in multidimensional space is generally sparse. The traditional clustering method usually applies Euclidean distance to the clustering of data objects. In the big data space without distance meaning, this method is obviously no longer applicable. It is more difficult for people to deal with more and more complex and huge data. At the same time, these data contain useful or interesting information for people. People expect to do a good job in knowledge extraction and application in these data, so as to make correct decisions in real life. It is necessary to analyze multidimensional data timely. Thus, the cluster analysis of multidimensional data emerges as the times require [8–10].

Compared with the traditional clustering method, the clustering algorithm based on grids is more suitable for multidimensional discrete big data. The algorithm thought is to divide the data space into grid units, and the clustering process is carried out on the grid. When the grid is small enough, the data points falling into the same grid are similar, which belong to the same cluster. The processing time of algorithm only depends on the number of units in each dimension of quantization space and the number of independent objects. The clustering efficiency is high. This algorithm is able to find the clusters with any shape and size. The specific research process is shown in Figure 1.

2.2. Multidimensional Discrete Big Data Processing. The high-dimensional data clustering technology refers to the clustering technology in high dimensional data space. Basically, the high-dimensional data space is a large-scale high-dimensional data set. The "dimension effect" determines that the difficulty of traditional clustering algorithm in high-dimensional dataflow clustering. "Dimension effect" means that when the dimension of data object increases, the computational complexity also increases exponentially. Therefore, how to counteract the influence of "dimension effect" on high-dimensional data flow clustering has become a problem [11].

2.2.1. Dimension Reduction. The essence of dimensionality reduction technology is to reduce the dimension of high-dimensional data objects. Traditionally, when the clustering algorithm was used to cluster them in low-dimensional data space, the method of attribute transformation for data objects was often adopted. At present, the main methods of dimensionality reduction include the principal component analysis, the self-organizing mapping network, and the wavelet transform.

Principal component analysis (PCA) is a widely used method of dimensionality reduction. When there are nd -dimensional data in a data set, the $n \times m$-order covariance matrix is calculated at first, and then k eigenvectors in this matrix are calculated. These vectors show the main features of data set. After that, the original high-dimensional data is projected along feature vectors. In this way, the dimension of high-dimensional data set is reduced. Finally, the traditional clustering algorithm can be adopted. The specific process is shown in Figure 2:

(i) Step 1: take each sample in data set as a column vector, and arrange them in columns to form a matrix with n rows and m columns

(ii) Step 2: subtract the mean value from each row vector (each variable) of matrix, so that the mean value

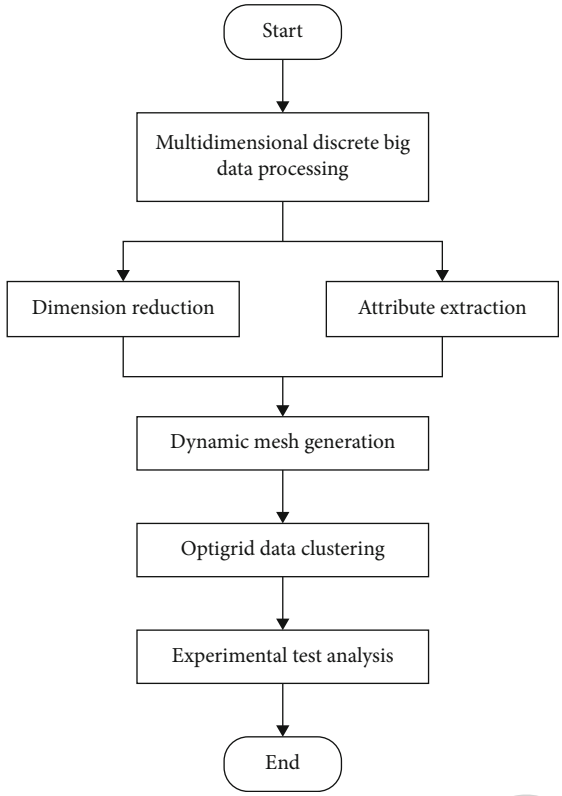


Figure 1: Research flow of clustering algorithm based grids.

of new row vectors is 0, so that a matrix X of new data set is obtained

(iii) Step 3: calculate the covariance matrix of X, the feature value λ and the unit feature vector e of covariance matrix

(iv) Step 4: according to the order of feature values, arrange the unit feature vectors into a matrix, so as to get the transformation matrix P, and then calculate the principal component matrix by PX

(v) Step 5: calculate the variance contribution rate and variance cumulative contribution rate by the feature value, and take the first k principal components whose variance cumulative contribution rate is more than 85%, or directly take the first k principal components if it is necessary to reduce to specific k -dimension [2]

2.2.2. Attribute Extraction. For high-dimensional data, the importance of each attribute is different. When calculating the distance between objects, it is not necessary to consider all attributes, but only two or three representative attributes. During the grid division and the grid spacing calculation, the computing complexity can be greatly reduced. The specific measures is to introduce the concept of entropy to divide key attributes and noncritical attributes, and thus to extract the key attributes [12–14].

The information entropy of information x represents the average uncertainty before information x occurs,

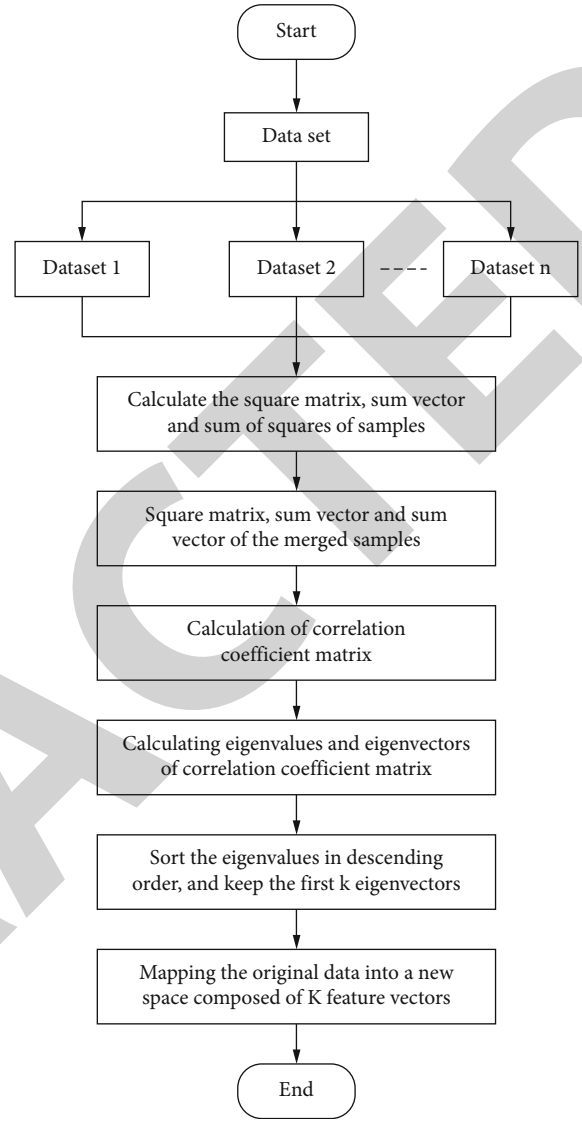


Figure 2: Dimensionality reduction process of principal component analysis.

namely the randomness of occurrence of x. If x is a variable of n different values, $x_1, x_2, \cdots, x_n$, the information entropy is defined as:

$$Y(X) = \sum_{i=1}^{n} f(x_i) \log f(x_i), \tag{1}$$

where $f(x_i) = k_i / \sum_{i=1}^{n} k_i$, and k_i is the number of occurrences of value x_i, n represents the attributes of each data in the data space.

The information entropy of each attribute is calculated and sorted from large to small to get an information entropy vector $\langle Y(1), Y(2), \cdots, Y(n) \rangle$, and the first two or three attributes are taken into consideration.

2.3. Dynamic Grid Generation. In the clustering based on grids, the number of grids in a dimension determines the

computational complexity of algorithm and the final clustering effect. The mesh generation methods contain the fixed mesh generation and the variable mesh generation. For the high-dimensional data space with large density distribution change, if the granularity is fine, the computational amount will be greatly increased. If the granularity is too coarse, the clustering effect is not ideal. Therefore, the grid partition is a very important step in grid clustering algorithm.

In traditional grid clustering algorithms, it is very important to choose the size of granularity of grid unit. If the partition granularity is chosen too small, the number of grid units will be increased, leading to the increase of amount of calculation. Meanwhile, this may make the number of data points falling into the grid units too little. And then, it may not satisfy the requirement of density threshold and it had to be ignored. If the granularity size is too large, the clustering accuracy will be reduced [15]. In Figure 3, the granularity of grid units is too large and the clustering quality is low, but there are a relatively small number of grid units, so the clustering speed is fast. In Figure 4, the granularity of grid units is small and the clustering quality is high, but the number of grid units is relatively large, so the clustering speed is slow. Moreover, the traditional grid clustering algorithm has some problems in incremental data processing. In processing the real-time data flow, how to deal with the problem that the measurement value of new data point in a certain dimension exceeds the grid space is the key.

In this algorithm, the space is divided into the grid structure for clustering analysis. In practice, the more centralized the spatial distribution of data, the less the actual number of grid, the better the clustering effect. At present, most of the researches on mesh generation focus on how to deal with the grid data units, rather than the analysis of partitioning methods. This section introduces the mesh generation methods commonly used in cluster analysis. In existing theoretical researches, the mesh generation methods mainly include the static grid and dynamic grid [16].

On the basis of satisfying the partition parameters specified by users, the static mesh generation is to divide the data space into several grid units with the same size. The grid unit with certain data points is called grid cell density. The grid units in this method only store the statistical information of data falling into the cell, such as the sum of data points and the number of data points. This mesh generation method is adopted by Wave Cluster and CLIQUE. Wave Cluster has a good effect in processing low-dimensional data space, and its result exceeds many excellent clustering algorithms. Based on Wave Cluster, CLIQUE considers the high-dimensional partition, but users need to specify the global density threshold. Meanwhile, its time complexity is relatively high.

The dynamic grid partition method refers to the strategy of separate processing. On this basis, the data space is divided recursively and thus to reduce the scale of problem. All the regions in original space are divided repeatedly until the same types of data points are included in the same region. These regions are the final grid units (Figure 5). OptiGrid is a typical dynamic grid clustering algorithm, which is built on the spatial data distribution.

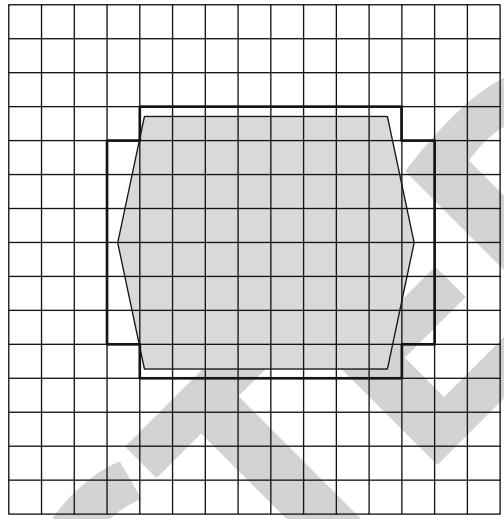


Figure 3: Mesh generation of large granularity.

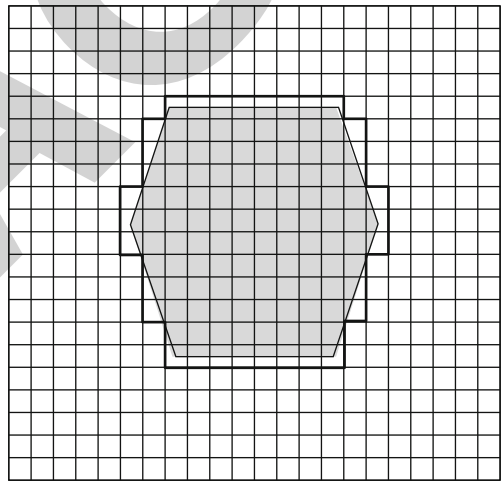
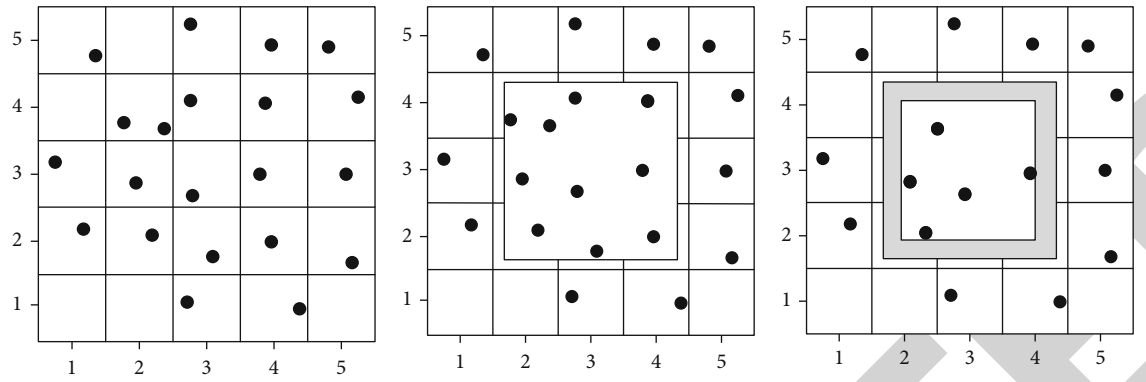


Figure 4: Mesh generation of small granularity.

It selects the optimal grid division by corresponding density information [17].

The static grid and dynamic grid have their own advantages and disadvantages. The static grid is good at processing the low-dimensional data, but it is not suitable for dealing with high-dimensional data. Through one-time data scanning, clusters of arbitrary shape are found. The clustering accuracy depends on the size of grid unit. If the unit is too large, the quality of cluster will be reduced. On the contrary, if the unit is too small, the accuracy may be better, but the time complexity of algorithm will increase. The dynamic grid is less affected by the spatial dimension, so it is suitable for processing massive high-dimensional data. It does not need users to specify the partition parameters. According to the density distribution of data, dynamic grid is able to divide the space. Generally, the volume of grid unit based on dynamic grid division is larger than that of static grid division. The accuracy of clusters formed by dynamic grid division is affected in low-dimensional space. Meanwhile, it



(a) Original grid (b) Extended grid (c) Extended area

Figure 5: Schematic diagram of principle of dynamic mesh generation.

is necessary to scan the data many times during the division, so it is time consuming [18].

The grid partition methods in multidimensional space are all extended on the basis of 2D grid partition methods. The proposed algorithm is researched based on the data in two-dimensional space. After considering the gridding time and accuracy, the dynamic grid method is applied in the multidimensional grid.

2.4. OptiGrid Data Clustering. After the density values of grid nodes are obtained, this algorithm needs to use these density values for the final clustering operation. The main idea of clustering algorithm based on grids is to divide the data space into some units, and these units compose the grid structure. This algorithm can use these grid structures for the clustering operation [19]. The main advantage of clustering algorithm based on grids is to map the data points in data set to the divided grids. Obviously, the efficiency of clustering algorithm is not influenced by the number of data points. The most common clustering algorithms include STING, CLIQUE, GCHL, O-cluster, OptiGrid, and GDILC.

Optigrid algorithm is a clustering algorithm based on density function mesh segmentation proposed by Hinneburg in 1999. This algorithm gives an effective gridding way in data space, and it is not affected by dimension and noise. The basic point of OptiGrid is to use the contraction mapping of data to determine the best cutting plane, and then the position of cutting plane is selected as the minimum value of density function distribution of data on the contraction mapping. This algorithm will give up some dimensions without the best cutting plane. A good cutting plane has two characteristics:

(a) The cutting plane needs to divide the data in the area with small density

(b) The cutting plane should be able to find clusters

The first limit guarantees that the cutting plane will not split a class cluster, and the second limit guarantees that this cutting plane is conducive to the grid clustering. Optigrid adopts the nonuniform grid division based on data, which not only considers the data distribution and more accurate division of space but also ensures that all clusters can be found, avoiding the low accuracy of most of algorithms in high-dimensional data. Meanwhile, the speed of grid clustering will not be affected. Therefore, OptiGrid gridding result can be directly output, without the subsequent processing. This method is an effective spatial gridding method.

The first limit guarantees that the cutting plane will not split a cluster, and the second one guarantees that such cutting plane is conducive to grid clustering. Optigrid uses the non-uniform grid division method based on data, which not only considers the distribution information of data, but also ensures that all clusters can be found, avoiding the low accuracy of most algorithms in high-dimensional data, and the speed of grid clustering will not be affected. Therefore, the results of OptiGrid gridding can be directly output as the results of clustering, without the need for subsequent processing like clique and sting. This method is a very effective spatial gridding method.

The clustering process of OptiGrid algorithm is shown in Table 1.

3. Experimental Test Analysis

In the above chapter, the data flow clustering algorithm based on density and grid was introduced in detail and analyzed theoretically. In order to verify the effectiveness of multidimensional discrete big data clustering algorithm based on dynamic grid, the clustering shape, efficiency, and accuracy of proposed algorithm was compared with the data clustering methods in Reference [2], Reference [3], and Reference [4] through experiments, and then the results analysis was given.

3.1. Experimental Environment and Data Set. The experimental environment of algorithm is as follows: Microsoft Windows XP Professional operating system, Genuine Intel (R) CPU, 1.73 GHz, and 1GB memory. This algorithm was written in *C* ++ language, and the data set used in algorithm includes the artificial data set and the real data set. The experimental data were processed by MATLAB software.

Table 1: Clustering process of OptiGrid algorithm.

Input: data set D; q: min-cut-score) Output: clustering results
(1) Determine a set of compressed mappings, $P = \{p0, p1,..., pk\}$
(2) Calculate all mappings of dataset D, $D \longrightarrow p0(D), p1(D),...,pk(D)$
(3) Initialize the cutting plane list set, best cut⟵Q, cut⟵Q
(4) For $i = 0$ to k do
(A) cut⟵to determine the best local cut (Pi (D)) (B) cut-score⟵Pi (D) best local cut (Pi (D)) score (C) add all cutting planes with scores greater than min-cut-score to best-cut
(5) If best cut⟵Q, then D is a cluster
(6) Select Q cutting planes with the highest score from the best cut, and delete the remaining
(7) A multidimensional grid set G is constructed by selecting the optimal cutting plane, and all the points in D are mapped to g
(8) Determine the optimal grid in grid set G and add it to cluster like set C
(9) Check and delete unqualified clusters in cluster collection C
(10) For each cluster c_i in C do OptiGrid (c_i, q, min-cut-score)

3.1.1. Synthetic Data Set. The clustering effect of algorithm on data sets of arbitrary shape can be tested by synthetic data sets. In Matlab platform, a two-dimensional data set of arbitrary shape is generated, as shown in Figure 6. This data set contains two attributes and five classes. Each letter corresponds to a class. There are 6500 data points, and the noise ratio is 6.4%.

3.1.2. Data Set in Real Environment. In this experiment, KDD CUP-99 data set is adopted. This data set is a real data set generated by DARPA intrusion detection evaluation project of MIT Lincoln Laboratory in 1998. Various user types, different network traffic, and attack mean are simulated. There are five million data records in the whole data set. The exception types are divided into four categories: DOS denial of service, R2L unauthorized remote host access, U2R unauthorized local super user privilege access, and PROBING port monitoring or scanning. There are twenty-two kinds of attacks. If the normal access traffic *NORMAL* is included in this data set, the whole data set can contain five categories, and they are marked as 1, 2, 3, 4, and 5, respectively. Each data record contains forty-one features, including 32 continuous features and 9 discrete features. Because the whole data set is huge, only 10% of the data are selected for the clustering.

It is necessary to preprocess the data before the experiment of real data set. The distance-based method is often used to calculate the similarity in the clustering algorithm. There are two kinds of features in KDD CUP99 data set, namely, the continuous feature and discrete feature. The threshold values of different features will influence the final similarity measurement greatly. For continuous feature attributes, the measurement methods are different. Generally, the smaller the measuring unit is, the larger the codomain of variables will be, which will affect the final clustering results. In other words, the influence on clustering will be greater when calculating the distance between data. In order to reduce the influence of measuring unit selection on final clustering result and eliminate the influence caused by difference between attribute measurements, it is necessary to standardize the attribute values and map the dimension data to [0,1] interval.

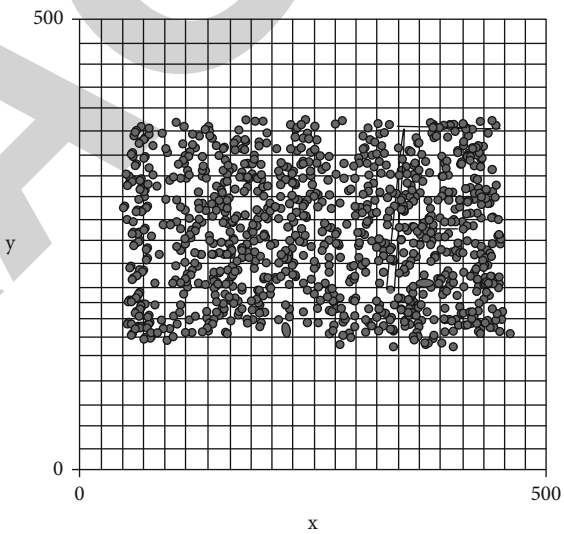


Figure 6: Synthetic 2D data set.

3.2. Cluster Shape. In order to test the ability of the proposed algorithm to find the clusters of arbitrary shapes, two artificial datasets with complex shapes are generated randomly. Finally, the clustering shapes are compared with the methods in Reference [2], Reference [3], and Reference [4]. Figure 7(a) is the clustering results obtained by the algorithm based on dynamic grid. Figure 7(b) is the clustering results obtained by the method in Reference [2]. Figure 7(c) is the clustering results obtained by the method in Reference [3]. Figure 7(d) is the clustering results obtained by the method in Reference [4].

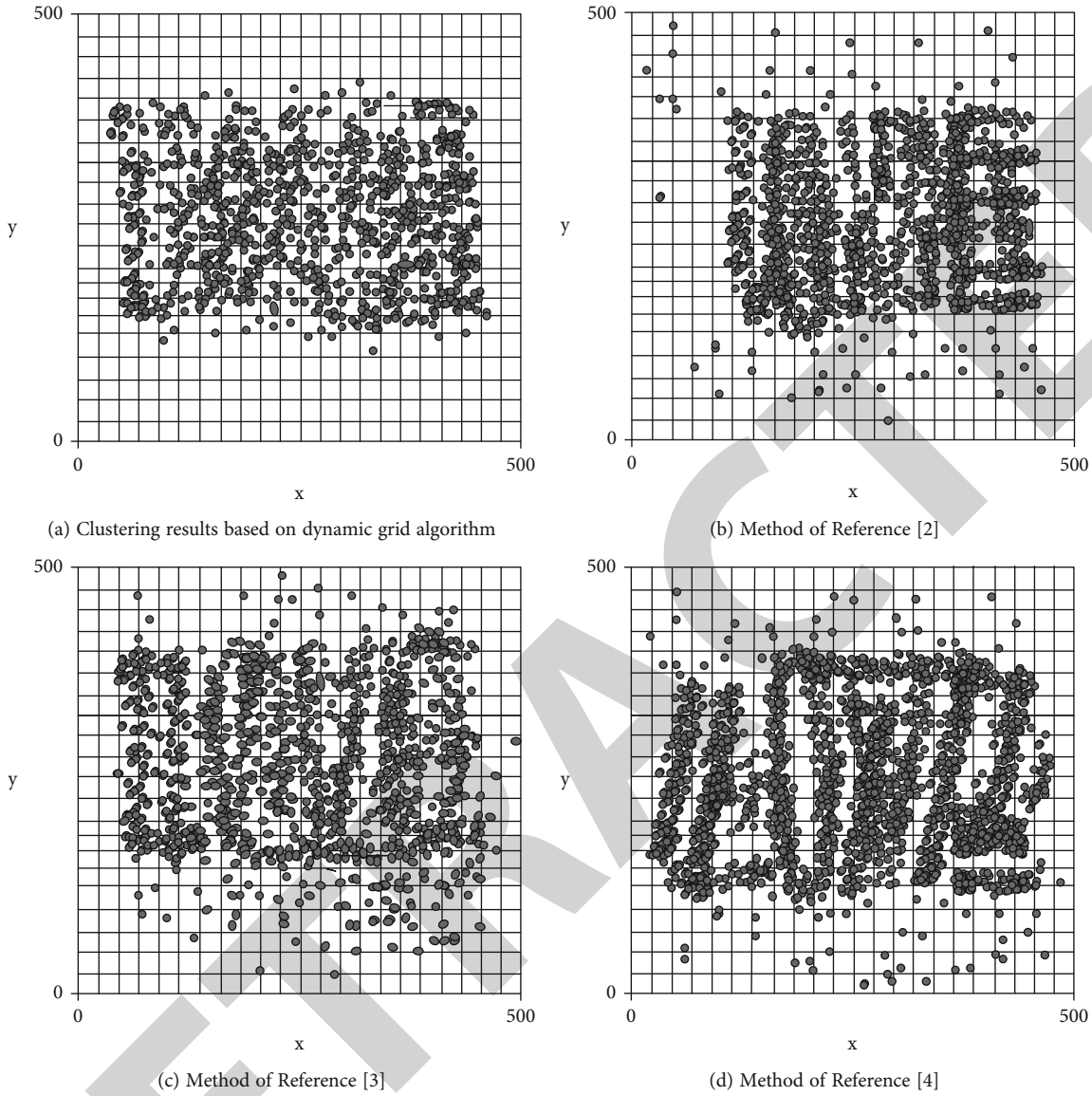


Figure 7: Comparison of cluster shapes in different methods.

Figure 7 shows that the clustering shape of the proposed algorithm is closer to the original data distribution compared with the methods in Reference [2], Reference [3], and Reference [4]. The subdivision of transitional mesh units for cluster boundary makes the cluster boundary of algorithm more accurate.

3.3. Cluster Purity. The cluster purity is the proportion of the largest number of clusters in a cluster result. Figure 8 shows the clustering purities of different methods on the synthetic data set when the dimension of data set is 50, 60, 70, 80, or 90.

In Figure 8, the clustering purity of the algorithm in this paper is between 95%–100%, that of the reference [2] algorithm is between 88%–96%, that of the reference [3] algorithm is between 92%–95%, and that of the reference [5] algorithm is between 83%–96%. With the increase of dimension of data set, the clustering purity of the proposed algorithm is always higher than that of the methods in Reference [2], Reference [3], and Reference [4]. Because the proposed algorithm only needs to change the number of grid units and sibling linked lists when the dimension increases, and the methods in Reference [2], Reference [3], and Reference [4] need to map the data space globally.

3.4. Execution Efficiency. The data flow clustering algorithm must have high execution efficiency to keep up with the arrival speed of data flow. Therefore, the efficiencies of different methods in *KDD-CUP99* data set are tested, and the time spent in processing different data volume is taken as the evaluation index. Experimental results are shown in Figure 9.

In Figure 8, when the data volume is 20 KB, the data clustering time of this algorithm is 4 s, that of Reference [2] algorithm is 6 s, that of Reference [3] algorithm is 7 s, and that of Reference [4] algorithm is 11 s; when the data volume is 40 KB, the data clustering time of this algorithm

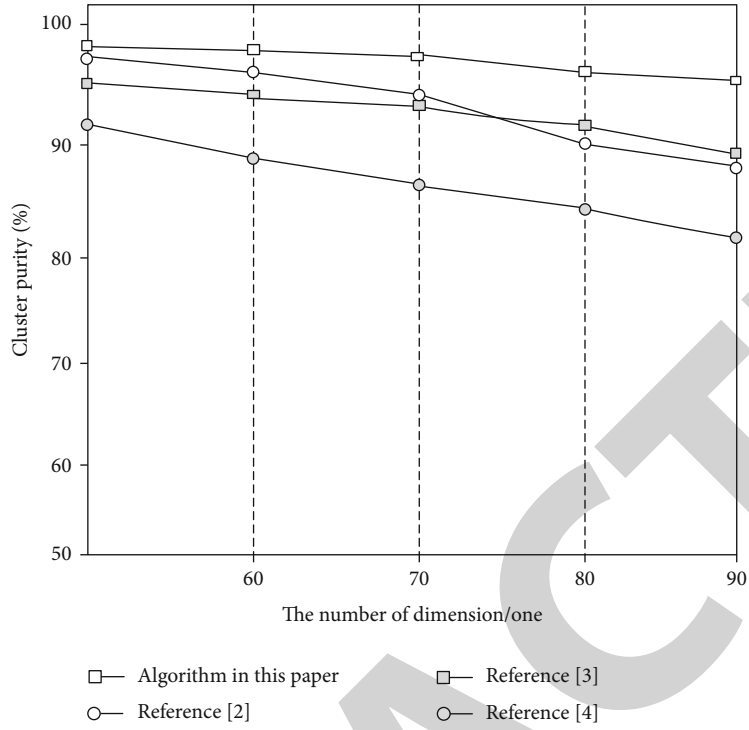


Figure 8: Cluster purity of four methods.

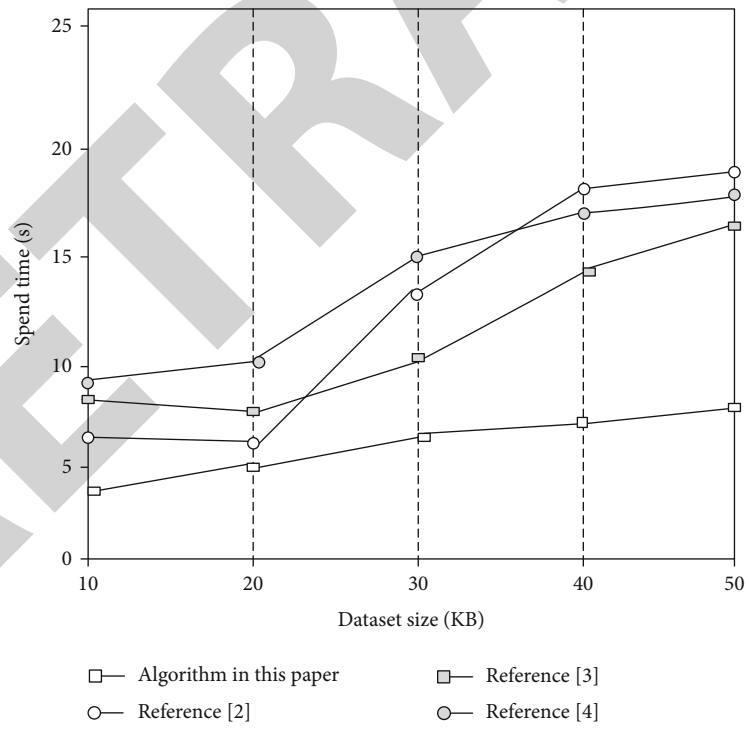


Figure 9: Execution efficiency.

is 6 s, that of Reference [2] algorithm is 18 s, and that of Reference [3]. The data clustering time of the algorithm is 13 s, and that of the reference [4] algorithm is 16 s. Figure 9 shows that with the increase of data volume, the increase rate of execution time of the proposed algorithm is lower than that of traditional method. The proposed algorithm dynamically adjusts the data density value at run time and adjusts the grid unit detection time interval at the same time, so that the proposed algorithm is able to adapt to the density distribution of current data space, which avoids the

frequent adjustment for the grid cluster and improves the computing efficiency.

4. Conclusions

The clustering analysis is an important part of data mining algorithm. It is also an analysis activity in data mining. The clustering algorithm is the core of overall clustering analysis, which determines the quality of all clustering results. At present, how to improve the clustering efficiency and reduce user's cost and burden under the premise of ensuring the stability and effectiveness of algorithm has become an interesting research. Because the traditional clustering method has high requirement for computer hardware resource, the time of massive data clustering operation is long and the clustering effect is not good. Therefore, a new clustering algorithm based on grids. The clustering quality and performance of algorithm is proved by experiments. The experimental results show that the data clustering time of this algorithm is at least 4 s, which is significantly less than the traditional algorithm. Experiment results show that the proposed algorithm has better overall performance. Due to the limitation of time and research ability, there are still many deficiencies. It is necessary to further research data flow clustering problem. The main problems include the following: although the proposed algorithm has better clustering quality and performance, but the spatial complexity is still high. This is the content to be further researched. At present, the experiment is based on numerical data, and the actual data has high-dimensional attributes. This is also a problem to be studied in the future.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

It is declared by the authors that this article is free of conflict of interest.

Acknowledgments

National FL Teaching and Research Project by Shanghai FL Education Press (2015-2017). Research on the Business Writing in a Blended Learning Environment under the Internet Age (2015FJ0008B)

References

[1] J. C. Fan and J. Wang, "A two-phase fuzzy clustering algorithm based on neurodynamic optimization with its application for PolSAR image segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 72–83, 2018.

[2] X. F. Mai, J. Liu, X. Wu et al., "Stokes space modulation format classification based on non-iterative clustering algorithm for coherent optical receivers," *Optics Express*, vol. 25, no. 3, pp. 2038–2050, 2017.

[3] J. L. Huang, Q. S. Zhu, L. J. Yang, D. Cheng, and Q. Wu, "QCC: a novel clustering algorithm based on Quasi-Cluster Centers," *Machine Learning*, vol. 106, no. 3, pp. 337–357, 2017.

[4] J. Zhou, L. Chen, C. P. Chen, Y. Wang, and H. X. Li, "Uncertain data clustering in distributed peer-to-peer networks," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 29, no. 6, pp. 2392–2406, 2018.

[5] R. He, Q. Li, B. Ai et al., "A Kernel-power-density-based algorithm for channel multipath components clustering," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7138–7151, 2017.

[6] P. Li, S. H. Lee, and J. S. Park, "Development of a global batch clustering with gradient descent and initial parameters in colour image classification," *IET Image Processing*, vol. 13, no. 1, pp. 161–174, 2019.

[7] L. Huang, G. S. Wang, Y. Wang, W. Pang, and Q. Ma, "A link density clustering algorithm based on automatically selecting density peaks for overlapping community detection," *International Journal of Modern Physics B*, vol. 30, no. 24, article 1650167, 2016.

[8] S. H. Han and G. Yi, "High performance clustering algorithm for analysis of protein family clusters," *The Journal of Supercomputing*, vol. 72, no. 5, pp. 1878–1896, 2016.

[9] Z. Lv, L. Qiao, M. S. Hossain, and B. J. Choi, "Analysis of using blockchain to protect the privacy of drone big data," *IEEE Network*, vol. 35, no. 1, pp. 44–49, 2021.

[10] Z. Lv, D. Chen, R. Lou, and H. Song, "Industrial security solution for virtual reality," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6273–6281, 2021.

[11] B. Li, R. S. Liu, J. J. Cao, J. Zhang, Y. K. Lai, and X. Liu, "Online low-rank representation learning for joint multi-subspace recovery and clustering," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 335–348, 2018.

[12] Q. M. D. Lohani, R. Solanki, and P. K. Muhuri, "Novel adaptive clustering algorithms based on a probabilistic similarity measure over Atanassov intuitionistic fuzzy set," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3715–3729, 2018.

[13] Z. Lv, D. Chen, and Q. Wang, "Diversified technologies in Internet of vehicles under intelligent edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2048–2059, 2021.

[14] T. Wang, W. Liu, J. Zhao, X. Guo, and V. Terzija, "A rough set-based bio-inspired fault diagnosis method for electrical substations," *International Journal of Electrical Power & Energy Systems*, vol. 119, article 105961, 2020.

[15] Y. Li, J. Wang, and T. Ding, "Clustering-based chance-constrained transmission expansion planning using an improved benders decomposition algorithm," *IET Generation, Transmission & Distribution*, vol. 12, no. 4, pp. 935–946, 2018.

[16] L. Cong, S. Ding, L. Wang, A. Zhang, and W. Jia, "Image segmentation algorithm based on superpixel clustering," *IET Image Processing*, vol. 12, no. 11, pp. 2030–2035, 2018.

[17] Z. K. Bao, J. G. Liu, and H. F. Zhang, "Identifying multiple influential spreaders by a heuristic clustering algorithm," *Physics Letters A*, vol. 381, no. 11, pp. 976–983, 2017.

[18] Y. W. Jiang, "Simulation of multi-dimensional discrete data efficient clustering method under big data analysis," *Computer Simulation*, vol. 36, p. 2, 2019.

[19] G. X. Wang, "Research on automatic recognition of error data in rescue command and dispatching database," *Automation & Instrumentation*, vol. 6, pp. 1–3, 2018.