Hindawi

Research Article

Constructing an Enterprise Management System Using an Improved Clustering Algorithm

Hua Zhu [ID]

*Organization and Personnel Department, Qingdao Vocational and Technical College of Hotel Management,
Shandong 266100, China*

Correspondence should be addressed to Hua Zhu; zhuhua@qchm.edu.cn

Cluster analysis is one of the most commonly used data analysis methods, and K-means is among the most popular and widely used clustering algorithms in recent years. Because of its performance in clustering huge amounts of enterprise data, the K-means algorithm is one of the most commonly used clustering methods in enterprises. Furthermore, in the context of the new era, enterprises are under growing pressure from both domestic and international sources, particularly market competition pressure. Moreover, the traditional K-means dispersed clustering algorithm has several issues when it comes to clustering big data in enterprises, such as instable clustering findings, poor clustering results, and low execution effectiveness. Based on the premise of the clustering algorithm in enterprises, this paper constructs an improved K-means clustering algorithm of the enterprise management system. It enables businesses to quickly address deficiencies, improve information in all areas, supplement adequate resources, and promote growth. The proposed algorithm uses the sample density, the distance between the clusters, and the cluster compact density, and defines the product of the three as different weight densities and the maximum difference as the initial cluster center to solve the problem. It finds the point of sampling with weight density, randomness, and low quality of early cluster center selection. The experimental results show that my planned algorithm has higher execution efficiency, accurateness, a lower error rate, and good stability.

1. Introduction

Enterprises are now widely spread and cheap all over the globe. According to [1], small and medium enterprises account for 99% of the total number of businesses in the country, and they generate 62% of industrial output value. Furthermore, Chinese enterprises employ approximately 76% of the Chinese population. With the increasing use of mobile Internet these days, an increasing number of employees are ready to use various movable applications for word processors and business communication, which increases enterprise management and operational efficiency. According to [2], the current situation of software system application in China indicates that the amount of management systems in large businesses is above 96%, but the number of small and medium firms in such enterprises is less than 6%. According to analysis and investigation, one of the major reasons for this is the limited funds and information

available to small and medium-sized businesses. Besides, in the context of economic globalization, businesses are under increasing market competition. Businesses must better manage their relationships with their customers to achieve sustainable development goals and stabilize economic resources. Furthermore, enterprises have their management system. As time goes on, the management system of enterprises improves at the same rate. Each management system has advantages and disadvantages, so choosing the best management system is the best approach. The existing management system on the market lacks personalized development and has an imperfect functional component for businesses [3].

In addition to the above, an enterprise management system can represent maximum of the operations of corporate management by providing real-time, related, precise, and full data to managers as a basis for decision-making program. Enterprise management systems can assist enterprise managers

in increasing work efficiency while also reducing their workload. There is no sophisticated process design, complex form design, or other such issues. According to [4], enterprise management software emphasizes the completeness of system functions, its process stability, technological advancement, and easy use of the system. The clustering algorithm [5] is a type of unsupervised machine learning model that divides data objects into subsets. There are numerous cluster analysis algorithms, which can be classified as hierarchical, partition, density, grid, and model-based methods. One of the partitioning methods is the K-means clustering algorithm. Because clusters are made up of objects, which are nearer to each other, the algorithm considers compact and impartial clusters to be the ultimate goal. Because the K-means clustering algorithm was suggested, which has gradually been applied in various disciplines, it is of great importance to the algorithm's development by incorporating it with other algorithms and enhancing its algorithm [6]. Despite of the above, the K-means clustering algorithm is quiet one of the most commonly utilized separation clustering algorithms [7].

Because of the rapid development of information technology, various business activities require efficient and accurate information systems for each process; therefore, in this regard the enterprise management system has a positive impact on business operations. In the context of the new era, the design and system that meet the needs of enterprise management have become the center of existing enterprise innovation. According to an analysis of current network construction and innovation, the layout of enterprise management systems faces numerous challenges. Now, only the judicious application of the K-means clustering algorithm with responsive attribute enhancement can provide a solid foundation for the practical development of small businesses [8]. It is important to investigate the system's user rights based on enterprise development and design an enterprise management system based on a clustering algorithm, which ensures that the enterprise management system can control the rights of all users. This paper builds an enterprise management system based on an improved clustering algorithm for the creation against the backdrop of a new era. It investigates the outside world's closer links, such as closely to the trend of the times and the pace, with state and government support. The main contributions of this paper are given as below:

(1) This paper explains and compares the different categories of clustering techniques.

(2) It compares the traditional K-means algorithm with an improved one for the development of an enterprise management system.

(3) Finally, this study takes customer data from a multinational corporation and selects 2000 pieces of worldwide customer data at random from the corporation's customer database. For class classification, the K-means clustering algorithm and the Chen-means algorithm are utilized.

The following parts of this paper are structured as follows: Section 1 deals with clustering and clustering algorithms,
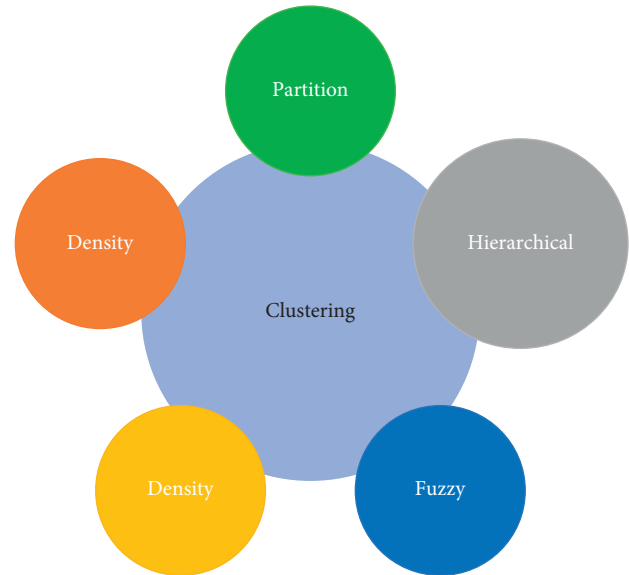

FIGURE 1: Approaches of clustering.

Section 3 clarifies the construction of an evaluation system based on an improved clustering algorithm, Section 4 shows experiment and application analysis, and Section 5 concludes this paper.

2. Clustering Algorithm

Clustering is a machine learning method that groups data points around each other based on their similarities, while the clustering algorithm is a nonsupervised technique, in which the input is un-labeled and solving of problem is based on the algorithm's knowledge gained from attempting to solve like difficulties as exercise schedule. Figure 1 explains five commonly used clustering approaches.

2.1. Clustering Based on Partition. It is a nonsupervised technique for grouping data points about an essential point recognized as a centroid. Clustering on partition splits groups into a predetermined number of clusters based on the distance between data points. It is important to note that each data item can only be assigned to one cluster at a period. The main principle behind the representative clustering algorithm based on the partition is that, first, K points are selected randomly as the initial cluster center, and then the Euclidean distance calculation equation is applied. Second, the data object closest to the cluster center is separated into clusters, and the cluster center is updated based on the average value of all types of data points until it no longer changes. As a result, the algorithm's principle is straightforward to grasp. This clustering algorithm has been demonstrated in practice to be highly efficient for large-scale data sets, but it is sensitive to initial point selection and necessitates the artificial setting of the K value. The results are prone to local optimization, and the algorithm is vulnerable to being troubled by abnormal points and major deviations [9].

2.2. Clustering Based on Hierarchical. Hierarchical clustering, which is based on the hierarchical clustering algorithm, creates a nested clustering tree with different levels based on the similarity of all data points. A nested clustering tree can be built in two ways: agglomerating hierarchical clustering and splitting hierarchical clustering (top-down). The condensed hierarchical clustering algorithm computes the interval between data points in each category and all data points to determine their similarity. Meanwhile, it merges the two groups of data points that are most similar to each other and generates a clustering tree [10]. In split hierarchical clustering, all data points are initially considered a cluster before being divided into two clusters based on the similarity calculation method. This process is repeated until each data point forms its cluster or reaches the termination condition. Finally, a top-down hierarchical clustering tree is established. The hierarchical clustering algorithm is simple to use for finding clusters of various shapes and is suitable for representing the similarity and interval of any pattern with a high degree of flexibility. However, the termination conditions of this type of algorithm are not clearly defined, resulting in an infinite cycle that cannot be terminated in time. Furthermore, the algorithm cannot handle large amounts of data or real-time dynamic data sets. The BIRCH algorithm, CURE algorithm, CHAMELEON algorithm, and others are examples of common hierarchical clustering algorithms.

2.3. Clustering Based on Fuzzy. This is an unsupervised clustering method, likewise known as soft K-means clustering. It assigns every data point to extra than one cluster. Because of the weighting of 0-1, each data point originally belonged to a cluster. Such an algorithm can be applied to the K-means algorithm in that you must specify the amount of clusters users want to produce. The algorithm also repeats itself till the last grouping is reached. Marketing applications can also benefit from fuzzy clustering. It can categorize customers into clusters based on their needs, desires, and purchasing habits, among other things. This algorithm may be beneficial to you because it offers details that can aid in the interpretation of the clusters generated. It is a more organic illustration of consumer behavior and, as a result, can provide retailers who use this strategy with a competitive advantage by allowing them to understand dynamic consumer behavior. However, one disadvantage of this algorithm is that you must specify the number of clusters to be generated as well as a threshold value for membership in the groups. This algorithm is also responsive to the centroids' initial placement.

2.4. Clustering Based on Density. It is also a nonsupervised method for creating a network based on the concentration of data points inside an axis or area. Therefore, after the execution of the algorithm, each cluster can be given an area with a specific range that must comprise the lowest amount of data points to procedure groups. It is capable of analyzing data and making predictions. It can also deal with noise and outliers effectively. However, it is extremely sensitive to the choice of input parameters and may produce poor cluster descriptive terms. It is also inappropriate for data sets of high dimension like product attribute data. The density-based geographic clustering of implementations with noise algorithm is one of the most widely used density-based clustering algorithms. The above algorithm finds and develops high-density areas into specified clusters and noise. Whenever a data set is mapped on an axis, the data points that are just outside the best-fit line are referred to as noise. With an unorganized data set, noise is to be expected. However, high rate of noise can cause the clustering algorithm to malfunction and make it hard to identify strictly delineated clusters. The DBSCAN algorithm is a typical example of a density-based clustering algorithm. It divides the area into clusters, and the space database has a certain amount of noise. It finds any shape of clusters, where the cluster of contractility is stronger, and finally, it finds the cluster set with the greatest density of data points connected [11]. The DBSCAN algorithm first sets the interval and density criteria, to label all data items using these two thresholds. Second, a seed is chosen at random that does not belong to any class, and all data points with a density that the seed can achieve are located. Repeat the processes above until all data analysis items have been assigned to their appropriate categories. If the data determined by the distance threshold are not around any of the core data objects, the data are treated as a noise point for elimination. It can be seen that the DBSCAN algorithm can automatically define the amount of clusters, which is suitable for clusters of any shape. However, for large data sets, large memory consumption is required. At the same time, the uneven density of text data or large distance difference between categories will affect the clustering results.

2.5. Clustering Based on Grid. This type of clustering algorithm first organizes the data points into the grid-like framework inside which the clustering will occur. The algorithm assigns data points to each cell and calculates the density of every cell. A multi-resolution grid data model is used in this algorithm. This means that the algorithm divides the grid space into a limited number of cells in which clustering can happen. The number of populated grid cells, rather than the number of data points in the set, determines the complexity of the clusters.

2.6. K-Means Clustering Algorithm. K-Means clustering algorithm is one of the most commonly utilized clustering algorithms, with simple calculation and fast operation speed [12]. K-Means clustering algorithm is an unsupervised clustering algorithm, which is relatively simple to implement and has a better effect of clustering, so it is extensively utilized. The principle of the K-means clustering algorithm is comparatively simple, and its application is easy, with quick speed of convergence and superior clustering effect. The algorithm is highly interpretable, and the main parameter to be adjusted is only the number of clusters K. The main process of the K-means clustering algorithm is as follows:
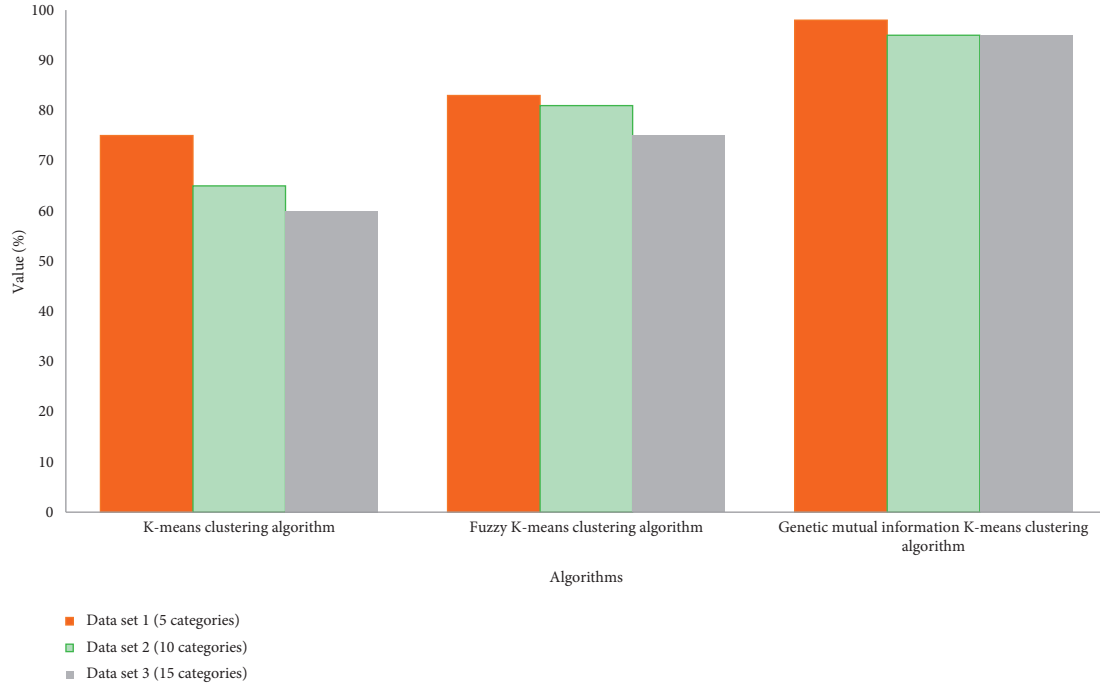
Figure 2: Running results of K-means algorithm and other algorithms.

(i) Firstly, N data samples are given, and the number of cluster categories K of the original data set is determined manually as the initial class center.

(ii) Secondly, divide these samples into K clusters, calculate the distance between each sample and the K centers, and repeatedly calculate new class centers until the class centers no longer change. The similarity between data samples within the same cluster is the greatest, while the similarity between data samples between clusters is the least.

The K-means clustering algorithm can be summarized as follows:

(i) Initialize K objects as the initial clustering center.

(ii) Assign each object to the cluster center which is the latest distance from it.

(iii) Recalculate the cluster center.

(iv) Termination conditions: no object is reassigned, there is no cluster center change, or the sum of error squares is locally minimum.

K-Means clustering algorithm has extensively popularized in numerous fields. Combined with the application practice of the K-means clustering algorithm, its application in gene recognition, feature extraction, text retrieval, and other fields is described [13]. After years of application and practice, the K-means clustering algorithm has been widely popularized and used in many fields and has achieved certain application results. However, the K-means clustering algorithm also has many problems, such as the impact of initial partition accuracy; poor measurement method will also affect the accuracy. Therefore, to improve the running accuracy of the K-means clustering algorithm, people introduce a genetic algorithm, information theory, fuzzy theory, etc., to increase the correctness of the K-means clustering algorithm from multiple dimensions, as shown in Figure 2.

3. Evaluation System Construction Based on Improved Clustering Algorithm

3.1. Traditional K-Means Clustering Algorithm. K values randomly nominated from the sample data set as the initial center point of cluster analysis, and the remaining data items in the sample set were calculated with the initial center point one by one. The data items are classified according to the size of the calculated data values (i.e., near and far), and the data of the whole sample data set are divided into K clusters. The distance between the center point and a single point was calculated for each group of data of these K clusters, and the average value of the distance was obtained, which was the new clustering center point of each cluster [14]. When the above calculation is repeated, the final result of the cluster, namely, k clusters, can be obtained when the center of the cluster calculated in the last two times does not change anymore [15].

K-Means clustering algorithm can carry out unsupervised learning. For sample set,

$$P = [x_1, x_2, \ldots, x_m]. \tag{1}$$

The "K-means" algorithm can cluster the samples into K clusters $C = C_1$, C_2, ..., C_k, and minimize squared error (objective function):

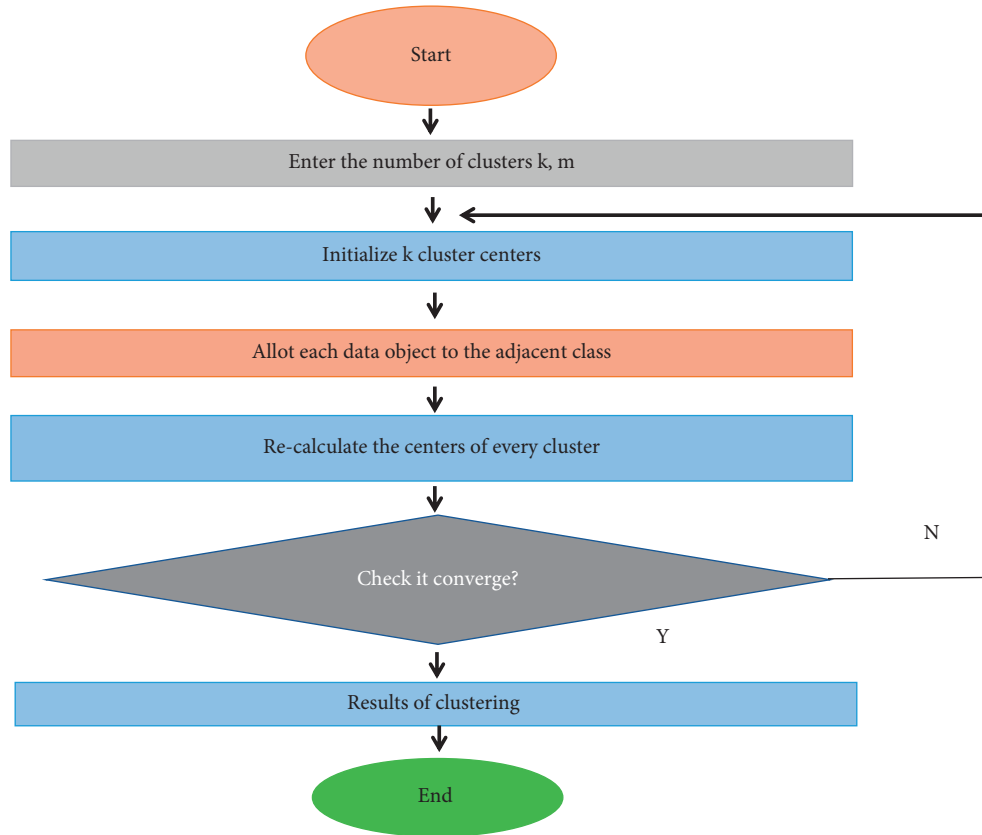$$E = \sum_{\iota=1}^{k} \sum_{x \in C_\iota} \|x - u_\iota\|^2, \tag{2}$$

FIGURE 3: Basic flowchart of K-means clustering algorithm.

where u_i is the center point of cluster C_i:

$$u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x. \tag{3}$$

To sum up, the purpose of the algorithm is to obtain the minimum objective function; that is, the finishing clustering result creates the objective function to obtain a minimum value, to attain enhanced classification effect. The steps of K-means clustering algorithm can be explained below:

Step 1. Given sample set P with size m, randomly select K objects $C = C_1, C_2, \ldots, C_k$ as the initial clustering center.

Step 2. Compute the distance dist(x_i, U_j) among all sample data objects and every cluster center, and divide the corresponding objects into the closed cluster based on the minimum distance.

Step 3. Recalculate the center of each newly formed cluster.

Step 4. The standard measure function is calculated, and the algorithm terminates when certain termination conditions are met. Otherwise, the termination conditions are not met, and the algorithm goes to Step 2.

Figure 3 describes the basic flowchart of the K-means clustering algorithm. According to this diagram, the desired clusters in my case are cluster k and cluster m, which entered into the model, and the k cluster center initialized accordingly. After initialization, the object is assigned to the nearest class, which recalculates the center of every cluster. The model is checked for convergence here; if it is not, Step 2 is executed; otherwise, the output cluster is obtained, which completes the process.

At present, there are basically three stopping criteria for K-means clustering algorithm:

(i) The centroid of afresh-designed clusters does not alter.

(ii) Data points are kept in the identical cluster.

(iii) Reach the extreme number of repetitions.

3.2. Improved K-Means Clustering Algorithm. My improved K-means clustering algorithm does not adopt random selection when choosing the preliminary center point but considers the initial center point by calculating the density of the data set. Related to the conventional K-means algorithm, the randomness and blindness of data center selection are reduced. The steps of my proposed improved K-means clustering algorithm for the construction of an enterprise management system are as follows:

The AvgDist(D), the average distance between all samples, was calculated according to

$$\text{AvgDist}(D) = \frac{2 \times \sum_{j=1}^{n} \sum_{j=i+1}^{n} d(x_i, x_j)}{n \times (n-1)}. \tag{4}$$

TABLE 1: Precision comparison of data sets.

Time	Conventional K-means clustering algorithm	Improved K-means clustering algorithm
1	0.75	0.81
2	0.68	0.82
3	0.85	0.81
4	0.65	0.82
5	0.59	0.80

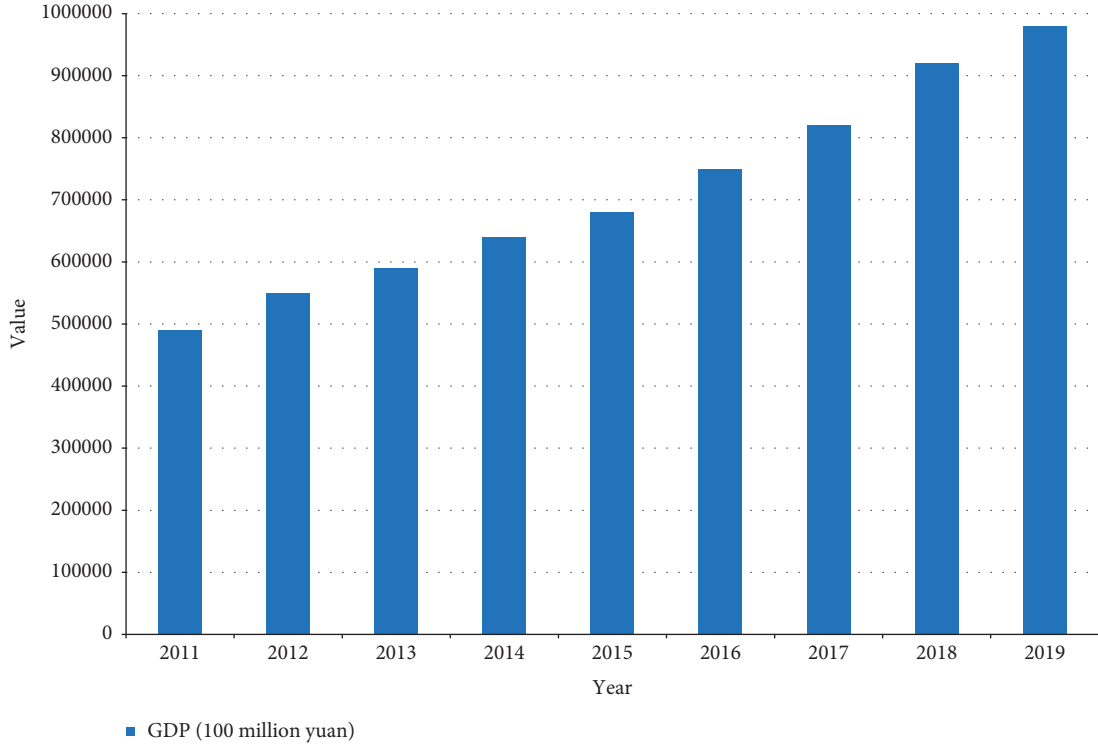FIGURE 4: GDP rate of China from 2011 to 2019.

In data sets D, the density of element (i) in the sample is represented by

$$\rho(\iota) = \sum_{j=1}^{n} f\left[d_{\iota j} - \text{AvgDist}(D)\right]. \quad (5)$$

In, $f(x) = [1, \ x < 0; \ 0, \ x \geq 0]$.

Here, $\rho(i)$ is the amount of samples, which meet the requirement that the distance between other samples to point i is less than AvgDist(D). This sample defines a cluster, within which the average distance is given by

$$\alpha(\iota) = \frac{2 \times \sum_{i=1}^{\rho(\iota)} \sum_{j=i+1}^{\rho(\iota)} d\left(x_{\iota}, x_{j}\right)}{\rho(\iota) \times (\rho(\iota) - 1)}. \quad (6)$$

The distance of clusters, i.e., $s(i)$, denotes the space between the element i of sample and element j of the other sample with a greater native density. So, if the native density of points i of the sample is the max, then $s(i)$ is considered as max $d\{(I, j)\}$, but, if current $\rho(j) > \rho(i)i$, it is given as: min $_{j:\ p(j)>p(i)}\{d(i, j)\}$.

$$S(\iota) = \{\min\{d(\iota, j)\}, \exists j, \rho(j) \rangle \rho(\iota); \max\{d(\iota, j)\}, \exists j, \rho(j) \rangle \rho(\iota)\}. \quad (7)$$

The size of the weight is defined as a number, and the sample density of $\rho(i)$ is multiplied by the distance between clusters $s(i)$ and divided by the average distance of samples within clusters $a(i)$, namely,

$$w = \frac{\rho(\iota) \times s(\iota)}{\alpha(\iota)}. \quad (8)$$

The clustering center of conventional K-means clustering algorithm is a random selection, which often results in local optimization rather than global optimization. In this paper, the improved K-means clustering algorithm is suggested, which can greatly reduce the problem of nonglobal optimal clustering results caused by blind selection of clustering centers and greatly enhance the accuracy and stability of clustering, as shown in Table 1.

The algorithm is as follows: first, sample density D is calculated according to (5). The first cluster center is the element with the largest sample density. Meanwhile, all sample elements satisfying the condition that the distance among the sample in equation (4) and the preliminary cluster center is less than AvgDist(D) are added to the current cluster, and these samples are deleted from set D.

Table 2: Experimental results of the first group.

Algorithm	Group number	Exact grouping quantity	Precision rate (%)	Error grouping number	Error rate (%)
K-Means	8	186	93	14	7
Chen_Means	8	200	100	0	0

Table 3: Experimental results of the second group.

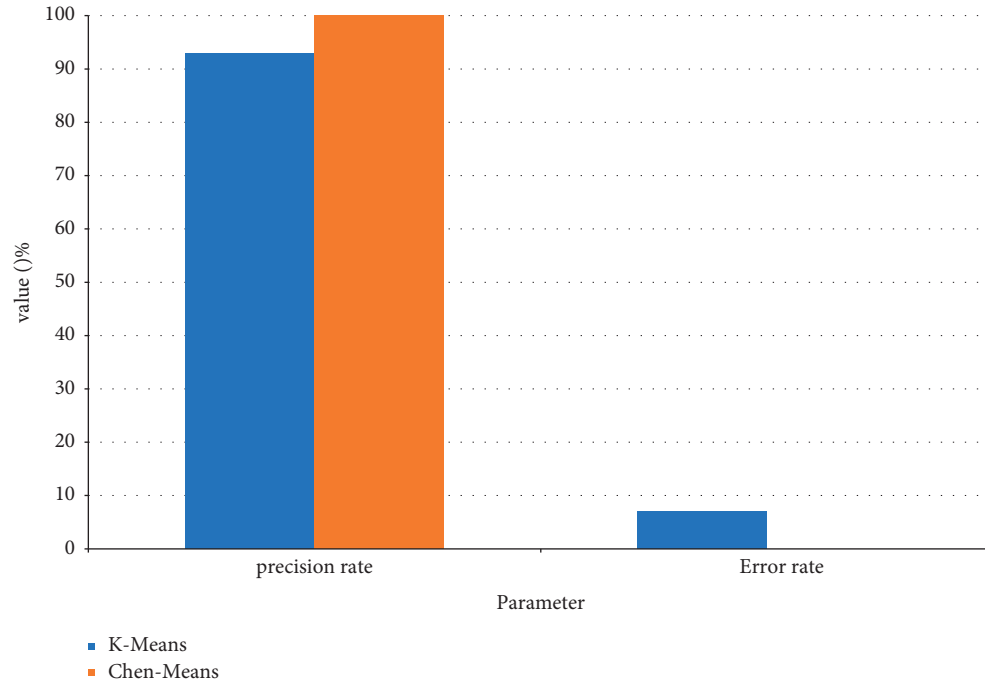| Algorithm	Group number	Exact grouping quantity	Precision rate (%)	Error grouping number	Error rate (%)
K-Means	11	1528	76.4	472	23.6
Chen_Means	12	2000	100	0	0

Figure 5: Comparisons between the K-means algorithm and Chen-means algorithm for the first group.

The second clustering center calculates the weight value w of the rest of elements as per equations (5)–(8), finds the maximum value, and selects the corresponding sample element as the second clustering center. The process is repeated until set D is empty.

4. Experiment and Application Analysis

The GDP growth rate of a country or region can truly reflect the economic development potential and growth space of the region at the current stage, and objectively present the dynamic changes of local economic development. In recent years, China's economy has gradually achieved high-quality growth, as shown in Figure 4.

As a result, building high-quality economic development requires not just bigger economic aggregate, but also a faster rate of economic growth, more development potential, and more growing space. This study takes customer data from a multinational corporation and selects 2000 pieces of worldwide customer data at random from the corporation's customer database. The first set of trials was created to test

the algorithm's effectiveness, and then 200 samples were chosen at random from a pool of 2000 samples for the experiment. For class classification, the K-means clustering algorithm and the Chen-means algorithm are utilized. Table 2 displays the results of the first group.

According to the findings of the experiments, the revised K-means clustering algorithm is more convenient because the K value does not need to be set. Furthermore, because the initial cluster center takes the maximum and minimum values, the drop in classification accuracy caused by random cluster center selection is minimized during the classification process. In the second group of experiments, all 2000 records were used in the experiment. As the data volume was nearly 10 times larger than that of the first group, the number of groups finally obtained might be more, which brought great difficulty to the K-means algorithm to determine the optimal K value. After many experiments, $K = 11$ was finally determined. My experimental results are revealed in Table 3.

As per the results in Table 3, the K-means clustering algorithm divides all samples into 11 groups (the best K value obtained through multiple experiments is 11), and the
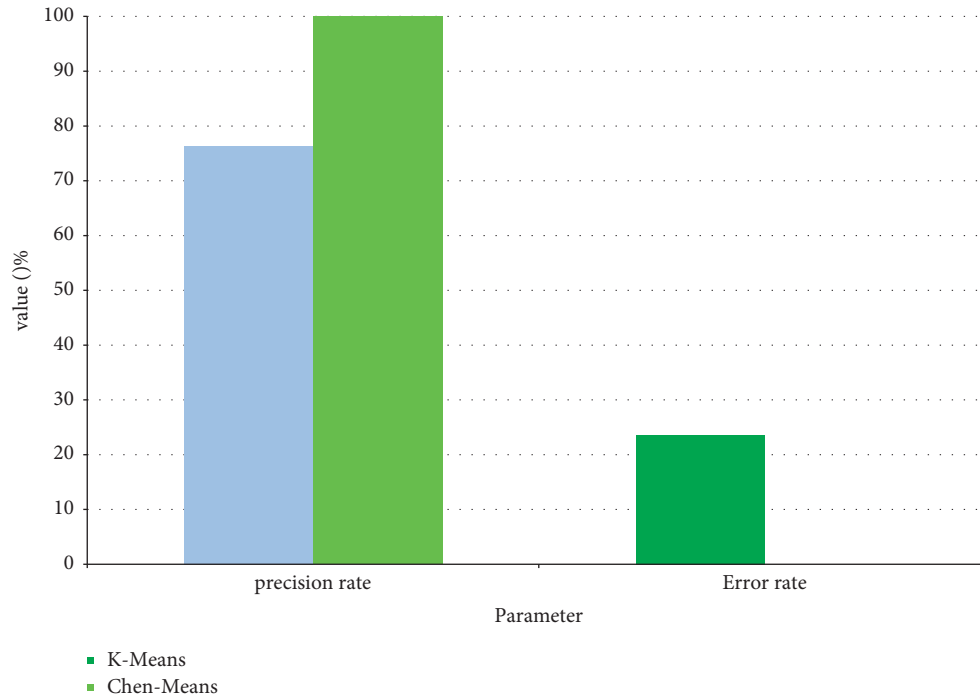
FIGURE 6: Comparisons between the K-means algorithm and Chen-means algorithm for the second group.

best accuracy is only 76.4%. Chen-Means clustering algorithm can still maintain 100% accuracy. The reason is that the determination of the adaptive K value also ensures the maximum group spacing. The iterative process is also an optimization process that keeps the maximum group spacing, so the best classification effect can be achieved in the end. With the increase of data volume, the K-means clustering algorithm will be more difficult, because the amount of classified groups will be more, and the advantages of the Chen-means clustering algorithm will be more obvious. Figure 5 explains the comparisons between the K-means algorithm and my proposed Chen-means algorithm for the first group. From this figure, it is clear that the precision rate (100%) of my proposed algorithm is high as compared to the traditional K-means algorithm (93%) and has zero (0) error rate as compared to the error rate (7%) of traditional K-means algorithm. This shows the efficiency and correctness of my algorithm for the construction of an enterprise management system.

Figure 6 explains the comparisons between the K-means algorithm and my proposed Chen-means algorithm for the second group. From this figure, it is clear that the precision rate (100%) of my proposed algorithm is high as compared to the traditional K-means algorithm (76.4%) and has a zero (0) error rate as compared to the error rate (23.6%) of traditional K-means algorithm. This shows the efficiency and correctness of my algorithm for the construction of an enterprise management system.

5. Conclusions

Recently, enterprise management system based on improved clustering algorithms has a positive impact on enterprise applications. Among other clustering algorithms, K-means has taken great importance these days. When compared with the traditional K-means clustering algorithm, the improved K-means clustering algorithm performs better in enterprise management systems. Keeping the above, this paper constructs an improved K-means clustering algorithm of the enterprise management system that enables businesses to quickly address deficiencies, improve information in all areas, supplement adequate resources, and promote growth. Furthermore, my proposed algorithm uses the sample density, the distance among the clusters, and the cluster compact density, and describes the product of the three as different weight densities and the maximum difference as the initial cluster center to solve the problem. It not only exposes and improves the defects of its algorithm but also efficiently and conveniently handles the large amount of data brought by enterprise business expansion. The experimental results showed that my planned algorithm has higher execution efficiency, accurateness, a lower error rate, and good stability. I believe that shortly, the improved K-means clustering algorithm will play a greater role in enabling enterprises to grow better and the society to develop more happily.

Data Availability

The data sets used and analyzed during the current study are available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflicts of interest.

Authors' Contributions

The conception of the paper and the data processing were completed by Hua Zhu and Hua Zhu participated in the review of the paper.

References

[1] M. Biggeri, "Small and medium enterprises and industrial clusters in BRICS countries," *Handbook of BRICS and Emerging Economies*, Oxford University, Oxford, England, 2020.

[2] X. Yu, F. Jiang, J. Du, and D. Gong, "A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains," *Pattern Recognition*, vol. 94, pp. 96–109, 2019.

[3] Global solver for low-rank subspace clustering, "Global Solver for Low-Rank Subspace Clustering," *Variational Bayesian Learning Theory*, vol. 25, pp. 255–278, 2019.

[4] A. Payne, S. Holt, and P. Frow, "Relationship value management: exploring the integration of employee, customer and shareholder value and enterprise performance models," *Journal of Marketing Management*, vol. 17, no. 7-8, pp. 785–817, 2001.

[5] J. Jiawei Han, "Data Mining: where is it heading?" in *Proceedings of the 13th International Conference on Data Engineering*, San Diego, CA, USA, April 2017.

[6] X. Wang and J. Wang, "Research and improvement on K-means clustering algorithm," in *Proceedings of the 2012 2nd International Conference on Computer and Information Applications (ICCIA 2012)*, Beijing, China, 2012.

[7] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[8] L. Guo, G. F. Liu, and Y. E. Bao, "A study of FCM clustering algorithm based on interval multiple attribute information," *Applied Mechanics and Materials*, vol. 444-445, pp. 676–680, 2013.

[9] Y. Yuan, C. Zhou, B. Qin, and Y. Ou, "A nearest neighbor index clustering algorithm for ecological regionalization based on multi-layer grid model," *Geo-Information Science*, vol. 13, no. 1, pp. 1–11, 2011.

[10] H. Dai, Q. Zhou, and J. Bian, "Markov clustering-based placement algorithm for hierarchical fpgas," *Tsinghua Science and Technology*, vol. 16, no. 1, pp. 62–68, 2011.

[11] Q. Chen, J. I. N. T. A. O. Sun, J. Sun, J. Liu, and B. Zhao, "Nonequilibrium Processes: Recent Accomplishments," 2020, http://Https://elibrary.ru/item.asp?id=44369782&pf=1.

[12] T. Lei and S. Li, "Improved K-means clustering algorithm by combining with multiple factors," in *Proceedings of the 2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC)*, Shanghai, China, April 2021.

[13] S. Zhao, Y. Xiao, Y. Ning, Y. Zhou, and D. Zhang, "An optimized K-means clustering for improving accuracy in traffic classification," *Wireless Personal Communications*, vol. 120, no. 1, pp. 81–93, 2021.

[14] X.-m. Tao, J. Xu, L.-b. Yang, and Y. Liu, "Improved cluster algorithm based on K-means and particle swarm optimization," *Journal of Electronics Information & Technology*, vol. 2010, no. 1, pp. 92–97, 2010.

[15] F.-y. Yuan, X.-c. Zhang, and S.-b. Luo, "Accurate property weighted K-means clustering algorithm based on information entropy," *Journal of Computer Applications*, vol. 31, no. 6, pp. 1675–1677, 2012.