

Research Article

Construction of Student Information Management System Based on Data Mining and Clustering Algorithm

XueHong Yin 

School of Accounting, Zhejiang Gongshang University, Hangzhou 310018, Zhejiang, China

Correspondence should be addressed to XueHong Yin; yinxuehong@mail.zjgsu.edu.cn

Received 8 April 2021; Revised 27 April 2021; Accepted 7 May 2021; Published 18 May 2021

Academic Editor: Zhihan Lv

Copyright © 2021 XueHong Yin. 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.

Data mining is a new technology developed in recent years. Through data mining, people can discover the valuable and potential knowledge hidden behind the data and provide strong support for scientifically making various business decisions. This paper applies data mining technology to the college student information management system, mines student evaluation information data, uses data mining technology to design student evaluation information modules, and digs out the factors that affect student development and the various relationships between these factors. Predictive assessment of knowledge and personalized teaching decision-making provide the basis. First, the general situation of genetic algorithm and fuzzy genetic algorithm is introduced, and then, an improved genetic fuzzy clustering algorithm is proposed. Compared with traditional clustering algorithm and improved genetic fuzzy clustering algorithm, the effectiveness of the algorithm proposed in this paper is proved. Based on the analysis system information management related tools and methods, in response to the needs of the student information management system, a simple student information management system is designed and implemented, which provides a platform and data source for the next application of clustering algorithm for performance analysis. Finally, clustering the students' scores with a clustering algorithm based on fuzzy genetic algorithm, the experimental results show that this method can better analyze the students' scores and help relevant teachers and departments make decisions.

1. Introduction

Data mining started with the research of Knowledge Discovery in Database (KDD). It is a key step in the process of knowledge discovery [1]. Its production has its application background. As the world moves towards an information society, human beings ability to collect, organize, and produce information using information technology has also greatly improved, resulting in the creation of tens of thousands of various types of databases [2]. Data mining research does not only come from the accumulation of mountains. The demand for information processing is developed due to the urgent needs of all aspects of social development, and it plays a huge role in scientific research, technological development, production management, market expansion, commercial operations, and government offices. The academic and business circles at home and abroad attach great importance to the research and

development of data mining technology and software tools. Data mining is an extremely young and active research field, which combines the latest research results of database technology, artificial intelligence, machine learning, statistics, knowledge engineering, object-oriented methods, information retrieval, high-performance computing, and data visualization. After more than ten years of research, many new concepts and new methods have been produced. Especially in recent years, some basic concepts and methods have become clear, and its research is developing in a more in-depth direction [3].

With the continuous expansion of the scale of education, the number of students has increased sharply, which has put a lot of pressure on student management. The informatization of the student information management system is far from satisfying the demand. Therefore, the goal of building a digital campus is proposed, which is to use the Internet. Based on the use of advanced information technology

methods and tools, from the environment (including equipment and classrooms), resources (such as diagrams, handouts, courseware, and information), to activities (including teaching, learning, management, service, and office), all digitized data flows on the Internet, and one line of students, disciplines, colleges, student information management, finance, etc. all realize computer management. This digital campus will accumulate a large amount of data. How to mine the laws implicitly in the large amount of data so as to use these laws to guide the work of the school, improve the management of the entire school, and improve management efficiency are an extremely meaningful work [4]. In response to the above problems, we propose applying the data mining method to the student information management system and extracting useful student information through data mining. Data mining is to extract valuable and interesting knowledge from the data of large databases. This knowledge is implicit, unknown in advance, but potentially useful information. Data mining means collecting some facts or observation data. The decision support process of the model is determined. It is an interdisciplinary subject that combines theories and technologies in many fields such as artificial intelligence, database technology, pattern recognition, machine learning, statistics, and data visualization. As a kind of technology, data mining is in the chasm stage of its life cycle. It needs time and energy to research, develop, and mature gradually, and finally, it is accepted by people.

This paper systematically summarizes cluster analysis, one of the key technologies in data mining, conducts in-depth research on cluster analysis, introduces the current research hot issue—genetic algorithm optimization into cluster analysis—and proposes a fuzzy genetic algorithm clustering; the main content includes the following aspects: on the basis of a brief introduction to the research background of the subject and the significance of the topic, the current status of education informatization and data mining, related concepts of cluster analysis, and clustering are introduced. This paper introduces the current research status of education informatization and data mining, representative clustering algorithms, introduces traditional clustering based on genetic algorithm and clustering based on fuzzy genetic, and designs and implements a student information management system. Finally, the article applies the hybrid clustering algorithm based on fuzzy genetic algorithm to the analysis of student performance and compares and analyzes the clustering based on genetic algorithm and the clustering based on fuzzy genetic algorithm.

2. Related Work

The term Knowledge Discovery from Database (KDD) first appeared in the 11th International Joint Artificial Intelligence Conference [5]. After the first International Academic Conference on Knowledge Discovery in Database (KDD) and Data Mining (DM) was held in Canada in 1995, data mining became popular. It is the concept of “knowledge discovery,” and the deepening of knowledge discovery and data mining is the product of the combination of artificial intelligence, machine learning, and database [6]. The scale of

the KDD International Symposium hosted by the American Association for Artificial Intelligence has grown from the original symposium to the international academic conference. The research focus has gradually shifted from discovery methods to system applications, focusing on the integration of multiple discovery strategies and technologies, and the integration of multiple disciplines.

The domestic research on data mining was a little late and did not form an overall strength. At present, many domestic research institutes and universities are competing to carry out basic theories and applied research on knowledge discovery [7]. These units include Tsinghua University, the Institute of Computing Technology of the Chinese Academy of Sciences, the Third Research Institute of the Air Force, and the Naval Equipment Demonstration Center. Compared with foreign countries, the research in the field of data mining in our country is still in its infancy [8]. The vast majority of work focuses on the design of local algorithms, and there are few integrated system integration designs. Due to the lack of core technology, data mining is only a preliminary application in some domestic fields, such as banking, finance, and GIS. At present, domestic colleges and universities have not carried out extensive research on data mining in the education system. Zhejiang University uses association rule technology for personnel affairs. The information database is digging, trying to find the factors that affect the development of the discipline, and the relationship between the various elements that affect the development of the discipline. Today, as the network application environment and trends are changing, the campus network construction is transitioning from a 10-Gigabit campus network to a “digital campus network.” Digitization of the environment (including equipment and classrooms), resources (such as diagrams, handouts, courseware, and information), and activities (including teaching, learning, management, service, and office), thereby, enhances the efficiency of traditional campuses and expands the traditional function of the campus, ultimately realizing the comprehensive informatization of the education process and achieving the purpose of improving the quality of teaching, scientific research, and management [9]. A new era of education informatization is approaching, and it will gradually enter the “data core era” from the past “basic network core era.” Data mining technology will undoubtedly play an increasingly important role in how to find the hidden laws in a large amount of data and then apply these laws to guide the work of the school. The half-life period of an algorithm being state-of-the-art shrinks also with increasing investments in Data Science and with more and more people being interested in the field of Data Science and Machine Learning. Consecutively, this article might be already out of date in a year. But for now, these are leading techniques that help in the progress of creating better and better algorithms.

The current research and development directions of cluster analysis are as follows:

(1) Research on the scalability of the algorithm: that is, the algorithm should be effective for both small data sets and large data sets. (2) Research on nonnumerical data clustering, which can handle both numerical data and

processing. Nonnumerical data can handle both discrete data and data in continuous domains. (3) Clustering studies of arbitrary shapes can be found. Traditional algorithms using Euclidean distance tend to find spheres with similar density and size clusters, but for other clusters, they may be arbitrary shapes, and it is extremely important to propose an algorithm that can find clusters of arbitrary shapes. (4) Research on algorithms for processing high-dimensional data: many clustering algorithms are good at processing low-dimensional data. In a high-dimensional space, especially considering that such data may be highly skewed and extremely sparse, clustering is extremely difficult. (5) Research on the ability to deal with noisy data: in real applications, most of the data contains, in addition to outliers, unknown data, vacancies, or erroneous data, and some clustering algorithms are sensitive to such data and will lead to low-quality clustering results, so the processing of noise is extremely important. (6) Fuzzy clustering research, such as the clustering of information such as text, image, and sound [10].

3. Cluster Analysis Based on Fuzzy Genetic Algorithm

Lin et al. [11] propose an attention segmental recurrent neural network (ASRNN) that relies on a hierarchical attention neural semi-Markov conditional random fields (semi-CRF) model for the task of sequence labeling. Their model uses a hierarchical structure to incorporate character-level and word-level information and applies an attention mechanism to both levels. This enables their method to differentiate more important information from less important information when constructing the segmental representation. In this paper, the genetic algorithm is a computational model that simulates Darwin's genetic selection and natural elimination of biological evolution. Its ideas are derived from biological genetics and the natural law of survival of the fittest. It is a search algorithm with an iterative process of "survival + detection." The genetic algorithm takes all individuals in a group as the object and uses randomization technology to guide an efficient search of a coded parameter space. Among them, selection, crossover, and mutation constitute the genetic operation of the genetic algorithm; parameter coding, initial group, the five elements of the genetic algorithm, the design of the fitness function, the design of the genetic operation, and the setting of the control parameters constitute the core content of the genetic algorithm [12]. As a new global optimization search algorithm, the genetic algorithm is simple, versatile, and robust. Practical and other outstanding characteristics have been widely used in various fields, achieved good results, and gradually become one of the important intelligent algorithms.

3.1. Basic Principles. Before describing the basic principles, first, use Figure 1 to visually describe the basic process of genetic algorithm.

We are accustomed to refer to the Genetic Algorithm as the traditional GA, which is the process described in figure 1 [13]. Its main steps can be described as follows:

- (1) Coding: GA first expresses the solution data of the solution space as the genotype string structure data of the genetic space before searching. Different combinations of these string structure data constitute different points.
- (2) Generation of initial population: N initial string structure data are randomly generated, each string structure data is called an individual, and N individuals constitute a group. GA starts iterating with these N string structure data as the initial point.
- (3) The fitness value evaluation test: The fitness function indicates the pros and cons of the individual or solution. Different problems have different ways of defining the fitness function.
- (4) Selection: The purpose of selection is to select excellent individuals from the current population, so that they have the opportunity to serve as parents to breed offspring for the next generation. The genetic algorithm embodies this idea through the selection process. The principle of selection is that an individual with strong adaptability has a high probability of contributing one or more offspring to the next generation. The choice realizes Darwin's survival of the fittest principle.
- (5) Crossover: Crossover operation is the most important genetic operation in genetic algorithm. A new generation of individuals can be obtained through exchange operations. The new individuals combine the characteristics of their parents, and exchange reflects the idea of information exchange.
- (6) Variation: Variation first randomly selects an individual in the group and randomly changes the value of a string in the string structure data with a certain probability for the selected individual. Like the biological world, the probability of mutation in GA is extremely low, usually between 0.001 and 0.01. Mutation provides opportunities for the creation of new individuals [14].

3.2. Research on Improvement of Genetic Algorithm. Classification is the most widely used task in data mining. Classification is to make an accurate description or analysis model for each category by analyzing the data in the sample database. The derived model is based on the analysis of the training data set (that is, the data objects whose class labels are known), mining classification rules, and then using the classification rules to classify the records in other databases, to find a concept description of the category, which represents the overall information of this kind of data, that is, the connotation description of the category. The typical methods of establishing classification rules are AQ method, rough set method, genetic classifier, and so on. The connotation description of class is divided into feature

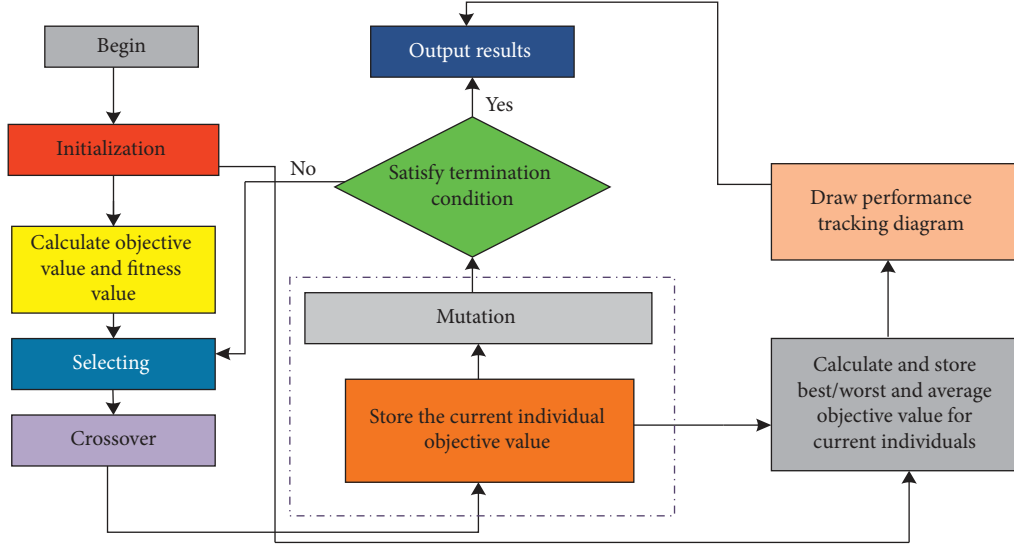


FIGURE 1: The basic flow of genetic algorithm.

description and discriminative description [15]. Feature description is the description of common features of objects in a class. Discriminant description is the description of the difference between two or more classes. Feature description allows common features among different classes. It is usually represented by rules or decision tree pattern. The schema can map tuples in database to a given category. The predicted value of classification model can be discrete (such as judging whether an animal is amphibian or mammal according to its characteristics) or continuous (such as judging a person's salary range according to their education and work experience).

Fuzzy genetic algorithm refers to the introduction of fuzzy control theory into the genetic algorithm, and the genetic algorithm is closer to the optimal solution in the evolution process through fuzzy adjustment of the relevant parameters of the genetic algorithm.

3.2.1. Chromosome Coding. In view of the fixed characteristics of the initial cluster centers, we select fixed-length chromosome coding [16]; that is, the length of the chromosome remains unchanged during the genetic process. According to the previous analysis, in FCM clustering based on genetic algorithm, when chromosomes cross and mutate, the value of each cluster center can be regarded as a whole; that is, each cluster center is regarded as the basic unit of chromosome. This representation method can directly use the number of the cluster centers in the sample set to represent the cluster center, and its direct benefit is that it can shorten the length of the chromosome code.

The text is coded by symbols; that is, the chromosome code is composed of the numbers of the K cluster centers in the sample set. The representation of chromosomes is

$$P = \{p_1, p_2, \dots, p_k\}. \quad (1)$$

Among them, K is the number of clusters, $p_i = 1, 2, \dots, k$ is the number of the sample corresponding to the i -th cluster

center in the sample set, which is a natural number between $[1, n]$, and n is the number of samples.

3.2.2. Population Initialization. Randomly generate K different natural numbers between $[1, n]$, and concatenate these natural numbers to form a chromosome, where n is the number of samples and c is the number of clusters. If the population size is N , then N chromosomes are generated according to the above method.

3.2.3. Design of Genetic Operators. The selection operation has a pivotal effect on the performance of the algorithm. In the evolution of the genetic algorithm, we first adopt the optimal preservation strategy to keep the individuals with the highest fitness in the genetic process, so that they do not participate in the cross-mutation operation, and then use the roulette method, which is determined by the probability distribution corresponding to the fitness function. Individuals in the current group are selected, crossed, and mutated to improve the average fitness of the group [17]. Since it is more appropriate to select elite individuals to account for 3%–6% of the population size, the population size of the algorithm in this paper is 30, so we choose to retain one elite individual.

The detailed flow of the algorithm is shown in Figure 2. When the genetic operation stops, it is necessary to find the chromosome with the highest fitness in the last generation, and its corresponding cluster center matrix P is the optimal solution obtained by the genetic operation. Take this optimal clustering center matrix as the initial clustering center of the FCM algorithm, execute the FCM algorithm, calculate the optimal fuzzy classification matrix, and then determine the optimal clustering division according to the membership division principle.

The specific steps are as follows:

- (1) Calculate the fitness of each chromosome.

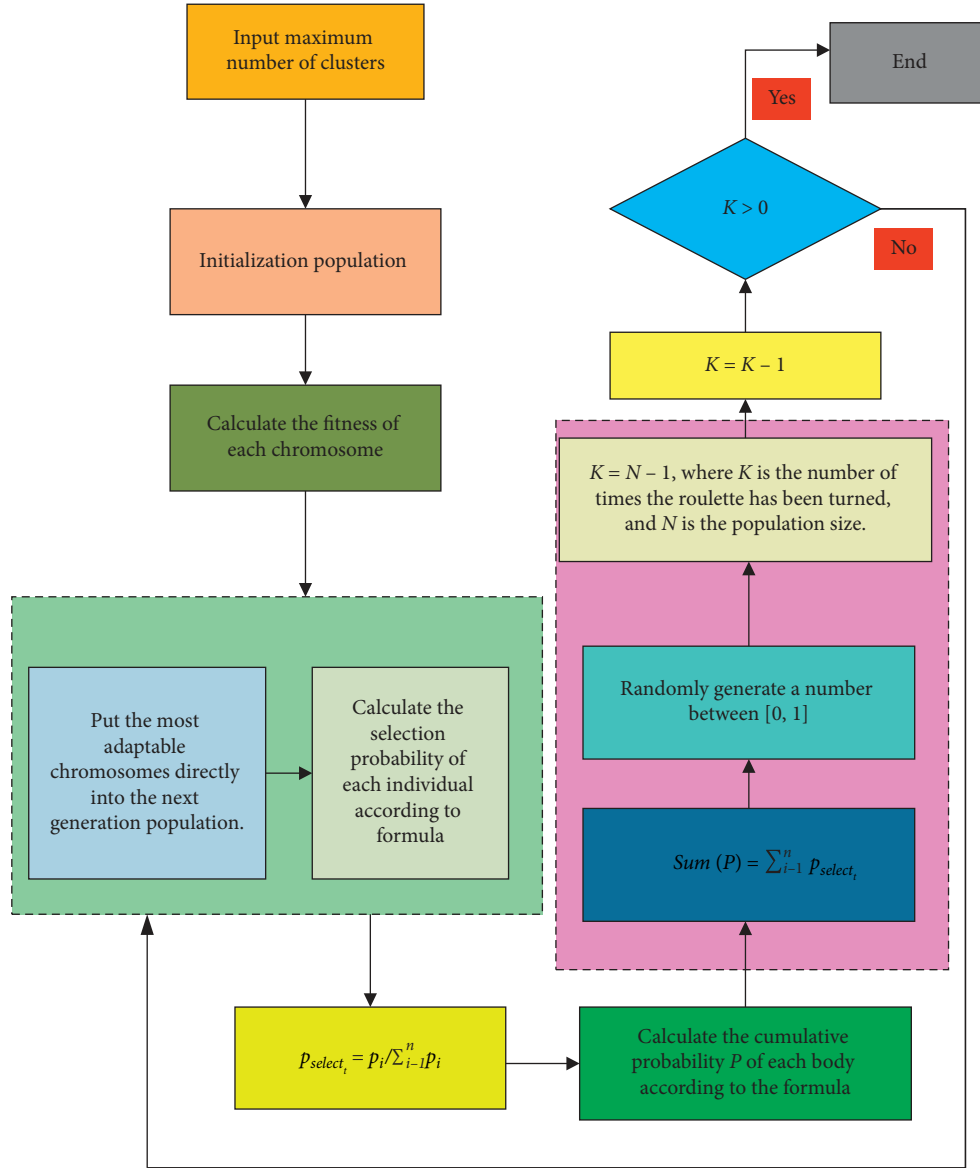


FIGURE 2: Algorithm flow chart of fuzzy clustering algorithm.

- (2) Put the most adaptable chromosomes directly into the next generation population.
- (3) Calculate the selection probability of each individual according to formula (2):

$$P_{select_i} = \frac{P_i}{\sum_{i=1}^n P_i}, \quad (2)$$

where n is the population size, and p is the fitness of individual i .

- (4) Calculate the cumulative probability P of each body according to

$$\text{Sum}(P) = \sum_{i=1}^n P_{select_i}, \quad (3)$$

- (5) $K = N - 1$, where K is the number of times the roulette has been turned, and N is the population size.

- (6) Randomly generate a number ε between $[0, 1]$; if $\varepsilon < \text{Sum}(P)$, then select the first chromosome P_1 ; if else, choose the i -th chromosome P_i that makes $\text{Sum}(P_{i-1}) < \varepsilon < \text{Sum}(P_i)$.

- (7) $K = K - 1$

- (8) If $K > 0$, turn to (6); otherwise, end the selection.

3.2.4. Fitness Function Design. Symbol-encoded chromosome is a chromosome representation method with simple representation and simple genetic operation, and it is easy to understand. At the same time, it can ensure that the search space of cluster centers does not increase with the genetic process, which is conducive to the increase of algorithm efficiency.

For Fuzzy Clustering Algorithm (FCM), the optimal clustering result corresponds to the minimum value of the objective function; that is, the better the clustering effect, the

smaller the number of meshes, and the greater the fitness [18]. Therefore, the fitness function of the individual here can be used to calculate the objective function:

$$f(S, P) = \sum_{k=1}^m \sum_{i=1}^n \lambda_{ik}^m \eta_{ik}(k, p_i). \quad (4)$$

With the help of FCM, the formula can be defined as

$$\begin{aligned} f(S, P) &= F(S, P)^{-1} \\ &= \left(1 + \sum_{k=1}^m \sum_{i=1}^n \lambda_{ik}^m \eta_{ik}(k, p_i) \right)^{-1}. \end{aligned} \quad (5)$$

4. Experimental Simulation and Result Analysis

In order to compare the performance of the traditional fuzzy clustering algorithm and the fuzzy clustering algorithm based on genetic algorithm, we selected sets of standard data; Impact Reporting and Investment Standards (IRIS) data set was selected as the test sample set to compare the convergence speed and optimization degree of each algorithm [19]. The data consists of 120 sample points in a four-dimensional space. The four components of each sample represent the petal length, petal width, sepal length, and sepal width of IRIS. The entire sample set contains three IRIS types: setosa, versicolor, and virginica, each with 50 samples. The first type of IRIS data is well separated from the other two types, and the other two types overlap. This data is often used as standard test data.

For the traditional FCM clustering algorithm, the cluster centers are randomly selected. We can observe its comprehensive clustering effect through clustering many times. Here, we run it 10 times and observe the clustering results. The experimental data is shown in Figure 3.

The data in Figure 3 verifies that the accuracy of the clustering results obtained by the traditional FCM algorithm is not stable enough. It is extremely sensitive to the selected initial cluster centers and thus falls into a local minimum. For example, the result of the third run is not the global optimal solution.

The improved FCM algorithm first optimizes the initial clustering center through an improved genetic algorithm and uses the obtained optimal solution as the initial clustering center of the FCM algorithm to start the FCM algorithm. We also run the improved FCM algorithm 10 times and observe the clustering results. The experimental data are shown in Figure 4.

The data in Figure 4 shows that the FCM algorithm optimized by the improved genetic algorithm can also ensure that the results of each convergence are correct and consistent, avoiding the objective function from falling into a local minimum, and the average number of iterations and total running time are higher than those of traditional ones.

From the comparison in Figure 5, it can be seen that the optimized FCM algorithm has a smaller objective function value than FCM algorithm, and the average number of iterations is less. It can be seen that the optimization of the

initial center can not only avoid falling into the local minimum, but also speed up FCM. In terms of time, although it takes a lot of time to optimize the initial center with genetic algorithm, it takes much less time to optimize the FCM algorithm than the FCM algorithm, and the time used is within the acceptable range of people. Moreover, if the amount of data processed is large, each iteration of the FCM algorithm will take a long time. At this time, reducing the number of iterations may save more time. From this point of view, the time spent optimizing cluster centers is completely worthwhile.

5. Design and Implementation of Student Information Management System

5.1. System Analysis and Design. Among them, the computing method is based on the data density. By calculating the distance of a group of data, cluster analysis can effectively divide these data into several more dense clusters, and the sum of the distances of the data in each cluster to the cluster center is the smallest. After using the cluster analysis technology, in the student performance evaluation, each cluster is a score group, and the data in the center of each cluster is the central score of the score group. Different clusters divide each score group accordingly and give the central score of different score groups correspondingly. These central grades are one of the reference standards for grading students' grades. It can be seen from the above that the score division based on cluster analysis is no longer the absolute score division, but the relative score division. Therefore, the score evaluation of students is more accurate.

The design of the system adopts a structured design method and divides the system requirements into different subfunction modules according to their respective functions. This design method is not only clear in layers and clear in structure, but also convenient for querying errors during design and debugging, and the preparation of programs is conveniently read [20]. Adopting this design method will bring convenience to future maintenance work, and it is easier to realize the system's added functions and improved functions. The system is divided into four modules: student information management, student status information management, performance information management, and reward and punishment information. Each module is subdivided into small modules to implement related functions [21–27]. The detailed function design is as follows, shown in Figure 6.

- (1) Student information management includes the establishment of freshmen's admission personal file management and the inquiry and modification of school student information. The establishment of freshmen files includes the grades of information such as department information, class information, student ID, name, gender, and age, newly assigned to the students. The query and modification of student information refer to the modification of student information, such as the incorrect registration of instant information such as new students, or the

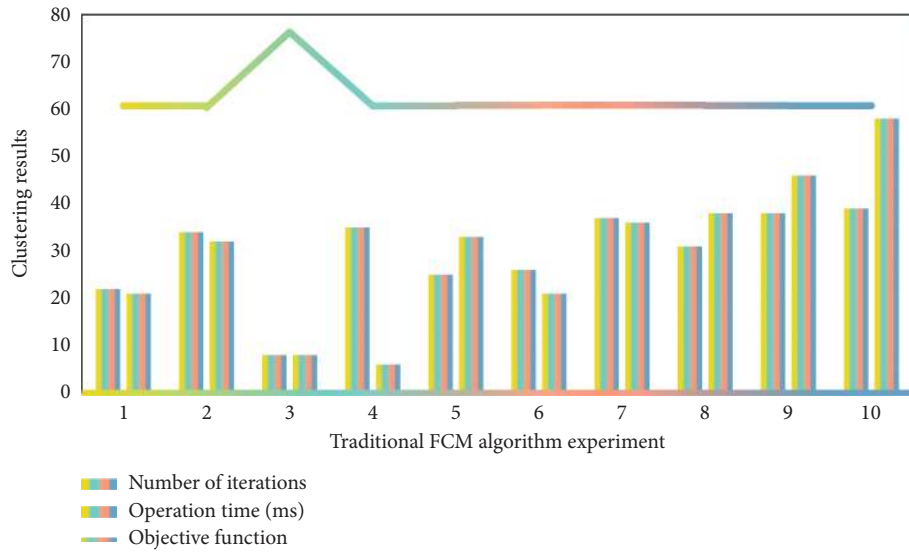


FIGURE 3: Clustering results of traditional FCM algorithm.

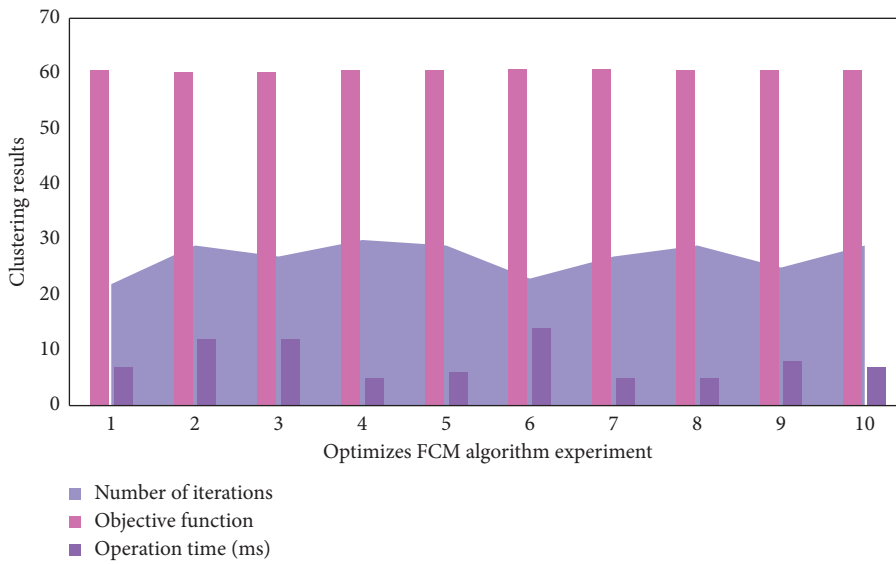


FIGURE 4: Optimization of FCM algorithm clustering results.

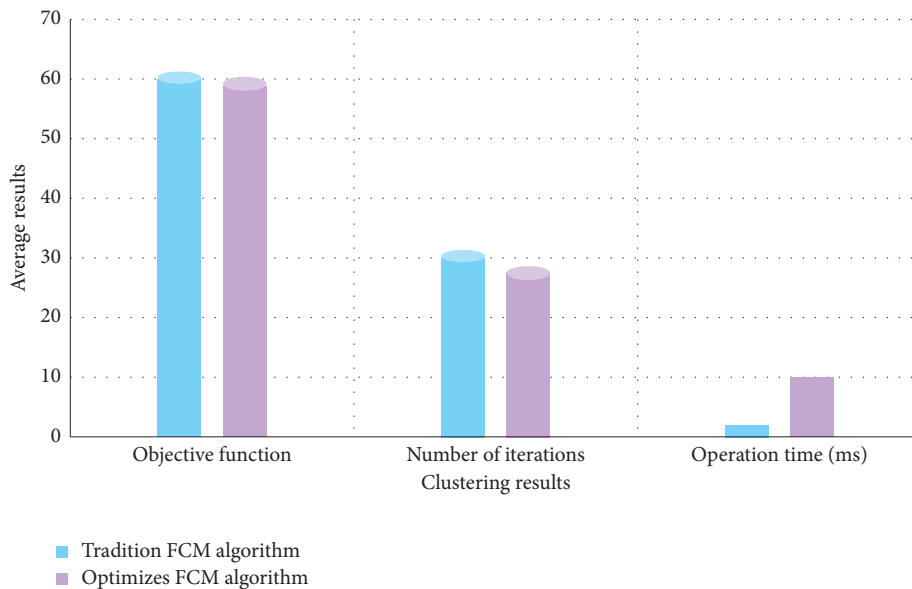


FIGURE 5: Comparison of the average of the clustering results of each algorithm.

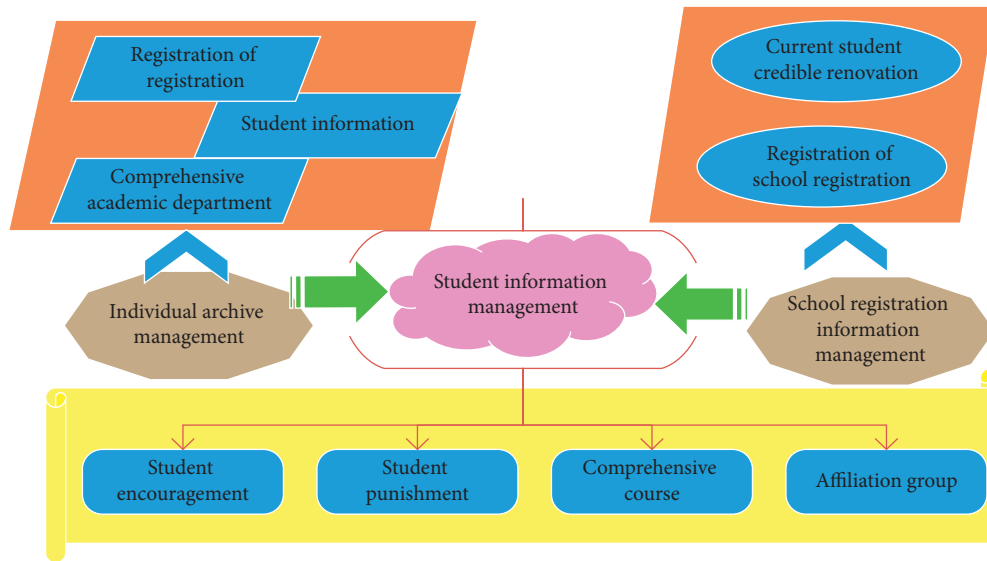


FIGURE 6: System structure block diagram.

change of student information, such as the change of home address and contact information.

- (2) Student status information management mainly refers to the record of student status changes. Student information query modification is mainly to maintain the added student information, including student information modification and student information deletion. The realization of student information maintenance and management is completed by modifying and querying the information in the student basic table [28].
- (3) The management of score information includes the registration and query of scores after each test.
- (4) Reward and punishment information management is mainly to reward students with outstanding learning and punish students with poor performance.

5.2. Analysis of Student Performance Based on Fuzzy Genetic Algorithm Clustering. Student performance is the most important part of the student information database, an important basis for evaluating teaching quality, and an important indicator of evaluating whether students have a good grasp of the knowledge they have learned [29]. Therefore, how to evaluate students' grades scientifically, accurately, and fairly is the work that educators have been studying for many years. With the continuous deepening of the reform of the education system, especially after the credit system teaching management system has become the mainstream, the evaluation method of student performance has developed from the single five-point system and the hundred-point system in the past to the more widely used hierarchical system today.

In order to better explain the application effect of the above-mentioned improved algorithm in the student

achievement data mining system, 180 students' achievements are selected for analysis. The scores of 180 students are divided into five grades (that is, excellent, good, intermediate, pass, and fail); the traditional division: those with 90 points or more are excellent, and those with 80 points or more and less than 90 points are considered as excellent, scores greater than or equal to 70 points and less than 80 points are considered medium, scores greater than or equal to 60 points and less than 70 points are passed, and scores less than 60 points are failed. The results are shown in Figure 7.

Divide the score into five grades (i.e., excellent, good, medium, pass, and fail), and divide it according to the basic k-means algorithm. If the initial cluster center is 75, 60, 65, 75, and 60, the results of the division are shown in Figure 8.

The results are divided into five grades (i.e., excellent, good, medium, pass, and fail). The results of the k-means algorithm based on fuzzy genetic algorithm are shown in Figure 9.

From the comparison of Figures 7 and 8, it can be seen that only one person is excellent in the traditional division method, while the number of excellent people obtained according to the basic k-means algorithm is four. This division is for students with 89.87 points, which is more reasonable. In addition, by comparing Figure 8 with Figure 9, it can be seen that the result of dividing according to the basic k-means algorithm is quite dependent on the initial clustering center, and dividing according to different initial clustering centers will result in different clustering results. If the initial clustering center is not well selected, the clustering result is easy to fall into the local optimum. If you want to get a better clustering result, you should select multiple groups of initial cluster centers for multiple clustering divisions and then compare them to find a better division as the final result. But this method is too dependent on the operator's mastery of the data.

Figure 9 is the clustering results obtained by taking 10 groups of initial cluster centers as 10 subgroups, and then

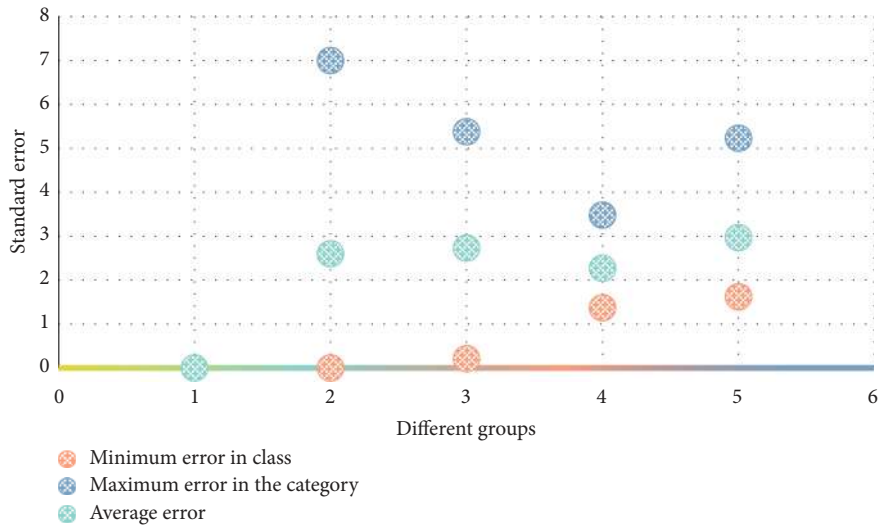


FIGURE 7: Grading students' performance according to the traditional method.

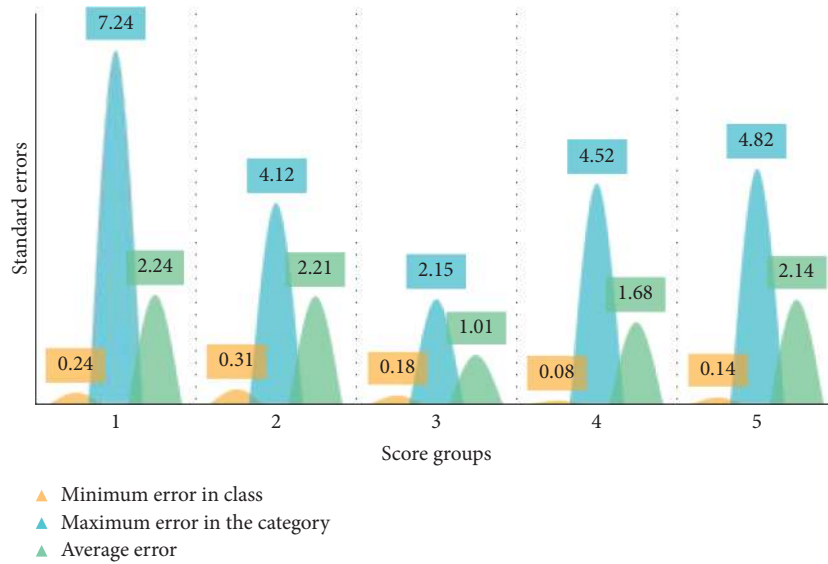


FIGURE 8: The grades of students' grades according to the basic k-means algorithm.



FIGURE 9: The grades of students' grades are divided according to the k-means algorithm based on fuzzy genetic algorithm.

using these 10 subgroups according to the fuzzy genetic algorithm optimization algorithm. By comparing the overall average error of Figures 7–9 (the overall average error of Figure 7 is 1.98, the overall average error of Figure 8 is 2.11, and the overall average error of Figure 9 is 1.94), the overall average error of Figure 9 can be seen. The error is obviously smaller than the other two; that is to say, the clustering result based on fuzzy genetic algorithm is more scientific, fair, and reasonable.

Ullah et al. [30] pointed out that Terahertz-based 6G networks promise the best speed and reliability, but they will face new man-in-the-middle attacks. In such critical and high-sensitive environments, the security of data and privacy of information is still a big challenge. Without privacy-preserving considerations, the configuration state may be attacked or modified, thus causing security problems and damage to data. In their article, motivated by the need to secure 6G IoT networks, an ant colony optimization (ACO) approach is presented by adopting multiple objectives, as well as using transaction deletion to secure confidential and sensitive information. We will work on the security of our method in the next step.

6. Conclusion

As an important part of data mining, cluster analysis has been widely used in various fields. Although various clustering algorithms have been proposed, different algorithms have their own characteristics. Therefore, in practical applications, the best clustering method should be selected or designed according to specific analysis of specific problems. Aiming at the deficiencies of the k-means clustering algorithm, this paper proposes a new idea, combining fuzzy genetic algorithm with an improved k-means algorithm, which, to a certain extent, avoids the sensitivity of the k-means algorithm to the initial clustering center. It is easy to fall into the defect of local optimal solution. The student information system contains a lot of useful information to be explored. Today, when the country is vigorously advocated by science and education, this information is useful for schools to better formulate learning.

Data Availability

The data 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 known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by 2020 fund by Zhejiang Gongshang University: Research on the Function Mechanism of Opinion Leaders in Network Education of Colleges (Approval No. Xgy20017); 2020 fund by Education of Zhejiang Province, Zhejiang Province, China: Research on

the Cultivation Path in Universities by Fusing Complex Situational Data and Social Media (Approval No. Y202045009).

References

- [1] L. H. Li, "Design of college english process evaluation system based on data mining technology and internet of things," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 16, no. 2, pp. 18–33, 2020.
- [2] Y. Zhou, "Design and implementation of book recommendation management system based on improved Apriori algorithm," *Intelligent Information Management*, vol. 12, no. 3, pp. 75–87, 2020.
- [3] G. Li and F. Wang, "Research on art innovation teaching platform based on data mining algorithm," *Cluster Computing*, vol. 22, no. 6, pp. 13867–13872, 2019.
- [4] C. Obeid, I. Lahoud, H. El Houry, and P. A. Champin, "Ontology-based recommender system in higher education," *InCompanion Proceedings of the Web Conference*, vol. 2, no. 3, pp. 1031–1034, 2018.
- [5] E. Sugiharti, S. Firmansyah, and F. R. Devi, "Predictive evaluation of performance of computer science students of UNNES using data mining based on naïve Bayes classifier (NBC) algorithm," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 4, p. 902, 2017.
- [6] S. Jain, R. Garg, V. Bhosle, and L. Sah, "Smart university-student information management system," in *Proceedings of the 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon)*, vol. 17, pp. 1183–1188, Bengaluru, India, August 2017.
- [7] L. I. Sarhan, A. M. Atroshi, and N. S. Ahmed, "A strategic planning of developing student information management system using SWOT technique," *Journal of University of Human Development*, vol. 2, no. 3, pp. 515–519, 2016.
- [8] F. Al-Hawari, A. Alufeishat, M. Alshawabkeh, H. Barham, and M. Habahbeh, "The software engineering of a three-tier web-based student information system (MyGJU)," *Computer Applications in Engineering Education*, vol. 25, no. 2, pp. 242–263, 2017.
- [9] D. Demirkol, C. Seneler, T. Daim, and A. Shaygan, "Measuring emotional reactions of university students towards a student information system (SIS): a Turkish University case," *Technology in Society*, vol. 63, no. 3, p. 101412, 2020.
- [10] X. Shao and G. Purpur, "Effects of information literacy skills on student writing and course performance," *The Journal of Academic Librarianship*, vol. 42, no. 6, pp. 670–678, 2016.
- [11] J. C.-W. Lin, Y. Shao, Y. Djenouri, and U. Yun, "ASRNN: a recurrent neural network with an attention model for sequence labeling," *Knowledge-Based Systems*, vol. 212, Article ID 106548, 2021.
- [12] X. Zhang, V. Venkatesh, and V. Venkatesh, "A nomological network of knowledge management system use: antecedents and consequences," *MIS Quarterly*, vol. 41, no. 4, pp. 1275–1306, 2017.
- [13] A. Whitelock-Wainwright, N. Laan, D. Wen, and D. Gašević, "Exploring student information problem solving behaviour using fine-grained concept map and search tool data," *Computers & Education*, vol. 145, no. 5, Article ID 103731, 2020.
- [14] C. Chan, "Institutional assessment of student information literacy ability: a case study," *CommInfoLit*, vol. 10, no. 1, p. 50, 2016.

- [15] M. C. Chen, S. Q. Lu, and Q. L. Liu, "Uniqueness of weak solutions to a Keller-Segel-Navier-Stokes model with a logistic source," *Applications of Mathematics*, 2021.
- [16] K. Holmes, E. Prieto-Rodriguez, and E. Prieto-Rodriguez, "Student and staff perceptions of a learning management system for blended learning in teacher education," *Australian Journal of Teacher Education*, vol. 43, no. 3, pp. 21–34, 2018.
- [17] J. Kite, T. E. Schlub, Y. Zhang, S. Choi, S. Craske, and M. Dickson, "Exploring lecturer and student perceptions and use of a learning management system in a postgraduate public health environment," *E-learning and Digital Media*, vol. 17, no. 3, pp. 183–198, 2020.
- [18] T. Devasia, T. P. Vinushree, and V. Hegde, "Prediction of students performance using educational data mining," in *Proceedings of the 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, vol. 1, no. 6, pp. 91–95, Ernakulam, India, March 2016.
- [19] A. Bansal, M. Sharma, and S. Goel, "Improved K-mean clustering algorithm for prediction analysis using classification technique in data mining," *International Journal of Computer Applications*, vol. 157, no. 6, pp. 975–987, 2017.
- [20] A. Algarni, "Data mining in education," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 6, pp. 456–461, 2016.
- [21] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Computers & Education*, vol. 113, no. 3, pp. 177–194, 2017.
- [22] S. I. T. Joseph and I. Thanakumar, "Survey of data mining algorithm's for intelligent computing system," *Journal of Trends in Computer Science and Smart Technology*, vol. 1, no. 1, pp. 14–23, 2019.
- [23] J. Wang, Y. Huang, T. Wang, C. Zhang, and Y. H. Liu, "Fuzzy finite-time stable compensation control for a building structural vibration system with actuator failures," *Applied Soft Computing*, vol. 93, Article ID 106372, 2020.
- [24] B. Wang and L. L. Chen, "New results on the control for a kind of uncertain chaotic systems based on fuzzy logic," *Complexity*, vol. 2019, pages, 2019.
- [25] J. Yang, J. Zhang, and H. Wang, "Urban traffic control in software defined Internet of things via a multi-agent deep reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [26] K. Sim, J. Yang, W. Lu, and X. Gao, "MaD-DLS: mean and deviation of deep and local similarity for image quality assessment," *IEEE Transactions on Multimedia*, p. 1, 2020.
- [27] C. Wang, C. Chen, Q. Pei, N. Lv, and H. Song, "Popularity incentive caching for vehicular named data networking," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–14, 2020.
- [28] M. Z. Chowdhury, M. Shahjalal, M. K. Hasan, and Y. M. Jang, "The role of optical wireless communication technologies in 5G/6G and IoT solutions: prospects, directions, and challenges," *Applied Sciences*, vol. 9, no. 20, p. 4367, 2019.
- [29] A. Onan, "Mining opinions from instructor evaluation reviews: a deep learning approach," *Computer Applications in Engineering Education*, vol. 28, no. 1, pp. 117–138, 2020.
- [30] K. W. Ullah, F. Jameel, J. M. Ali et al., "Efficient power allocation for NOMA-enabled IoT networks in 6G era," *Physical Communication*, vol. 39, Article ID 101043, 2020.