Hindawi
Mobile Information Systems
Volume 2022, Article ID 2730477, 9 pages
https://doi.org/10.1155/2022/2730477

Research Article

Reform and Practice of College Japanese Test Mode Using Big Data Analysis

Ruixin Huo

Tianjin University of Technology and Education, Tianjin 300030, China

Correspondence should be addressed to Ruixin Huo; xieanting@mails.ccnu.edu.cn

Received 8 April 2022; Revised 13 May 2022; Accepted 16 May 2022; Published 18 July 2022

Academic Editor: Liping Zhang

With the acceleration of economic globalization, the demand for Japanese professionals is also increasing. In view of this situation, this paper puts forward a set of college Japanese test systems. A college Japanese teaching model based on big data is established, and an improved clustering algorithm is proposed by comparing correlation analysis and data mining techniques. Aiming at the phenomenon that ACA (ant colony algorithm) cannot guarantee that all data objects are acquired and collected by ants and the same data object can be accessed repeatedly, this paper puts forward some corresponding improvements. The clustering results show that the clustering accuracy of the algorithm is 80%. This proves the rationality and feasibility of dynamically evaluating students' achievements based on clustering algorithm, and it is of great significance to introduce data mining into the teaching environment.

1. Introduction

With the acceleration of educational informatization, educational technology has been integrated into the classroom teaching ecosystem, affecting the balance of the classroom teaching ecosystem along with teachers, students, and classroom environment. The demand for Japanese talent in society is diverse, and many Japanese-related units and businesses require Japanese personnel. In this context, traditional Japanese talent training methods are no longer adequate to meet the demands of the new situation. As a result, the new curriculum reform places a special emphasis on classroom teaching effectiveness, requiring teachers to carefully design and organize classroom teaching in order to provide Japanese students with faster and more convenient learning as well as greater ability. The impact of economic globalization has been amplified in the age of big data. Learning Japanese, as an international language, is critical in this case. Writing instruction is particularly important and closely related to Japanese application among them.

At the moment, Japanese education is primarily conducted using traditional methods, with teachers prioritizing teaching. Students' autonomous learning ability is underutilized, their learning initiative and state are poor, and learning takes a long time with poor results. Kwong et al. believe that the concept of "teaching" should be free of traditional mindset constraints, and that it should not be limited to a single multimedia-assisted classroom teaching. Amirian extracts the text feature variables using natural language processing technology, analyzes the correlation between the two variables, and then determines the beta value of each variable using multiple linear regression to obtain the composition score statistically. Although many literature works have described successful data mining applications, Ghosh et al. pointed out that there are still some challenges associated with data mining and other application fields. Finally, they stated that research into many related technologies should continue in order to advance mining in education.

With the increase of the number of Japanese learners in China, Japanese has become one of the most important foreign languages after Chinese, so Japanese education is developing toward a more diversified direction. Therefore, application-oriented universities must construct a new mode of Japanese talent training and constantly reform the curriculum system. Only in this way can Japanese professionals conform to the trend of the times and better exercise their courage and experience. Through the application of data mining technology, this paper will discuss some problems existing in the management of students' academic qualifications. I hope that the reform of college Japanese test mode can be affirmed and popularized, so that under the current computerized situation, Japanese test can make full use of computer technology and multimedia technology. The research significance of this paper lies in improving the reliability and validity of Japanese test, improving the effect of Japanese test, and making it more effective in evaluating candidates' Japanese ability. The research contributions of this paper are as follows:

(1) At present, there is no examination system that can complete intelligent scoring well and accurately. This paper will use natural language processing technology, calculate the similarity between sentences by using minimum editing distance and cosine vector algorithm, and determine the scores and grammar knowledge points obtained by rule matching, so as to improve the accuracy of automatic correction of subjective questions in advance.

(2) Student achievement evaluation is a kind of educational evaluation, a key link and important content in teaching and teacher management, and the most important information source for teachers' quality evaluation. This paper mainly applies cluster analysis to the evaluation of students' grades, in order to extract the data behind the comprehensive grades, judge and analyze students' comprehensive learning ability through the cluster results, and put forward some suggestions on the related results.

The following sections can be found in the paper: The first chapter introduces the research background and significance and then introduces the main work of this paper. The second chapter mainly introduces the related technologies of the college Japanese test mode. The third chapter puts forward the specific methods and implementation of this research. The fourth chapter verifies the superiority and feasibility of this research model. The fifth chapter is the summary of the full text.

2. Related Work

2.1. Data Mining Research. The most direct purpose of achievement management is to realize the scientific and effective management of students' achievements through the application of related technologies. It is of great significance to use data mining tools to mine students' academic achievements. The LP linear regression analysis method is used to predict students' grades. Through statistics of data, the changing rules are sought, the influencing factors of changes are analyzed, and the fitting relationship is found. Thus, an intuitive change chart and development trend of students' grades are established, the future development of teaching is predicted, and the basis for scientific management teaching is provided. Guan et al. applied the improved data mining algorithm to study the analysis and management of college students' grades and got some internal reasons and conclusions that influenced students' grades, providing reference for the teaching management of colleges and universities [1]. Maggioni et al. used association rule mining to analyze students' grades and mined some rules with certain credibility, which provided a scientific basis for future grade management [2]. Jiao et al. pointed out that although face validity cannot replace experimental validity, it can quickly provide students' and teachers' opinions on the examination [3]. Lin and Hu used the data mining technology to mine the achievement association rules to guide the educational administrators to implement the guidelines and policies [4]. This method can be used to guide education to give full play to subjective initiative. Luo et al. applied data mining technology to the analysis of students' scores and found the correlation between courses, which provided a reference for teaching management and students' course selection [5].

2.2. Research on Clustering Algorithm. The essence of clustering is to divide all the data in the data set into many groups, in which one group must have similar features, while other different groups have different features.

Zhou et al. used a learning rule of competition called the second winner penalty to automatically decide the appropriate number of classes [6]. The idea is that, for each input, the weight of the winning unit will be modified to adapt to the input value, and the second winning unit will be punished to keep it away from the input value. Chen et al. applied PSO, K-means, and hybrid PSO algorithm to four different text files and clustered their data sets. After clustering, through comparative analysis, the clustering result obtained by hybrid PSO algorithm is very compact and takes a very short time [7]. Sun et al. divided clustering algorithms into small data clustering and big data clustering [8]. Small data clustering mainly reflects the basic idea of clustering, while the idea of big data clustering is mainly reflected in several aspects, such as concept, system structure, and architecture. As for the specific implementation algorithm of bottom-level clustering, there is actually no essential difference between it and small data clustering algorithm. Li proposed a fast and adaptive algorithm. This algorithm uses membership degree to determine the similarity of sample points, and it is a fuzzy clustering method based on objective function [9]. Yang et al. proposed a fast clustering algorithm [10] without setting the number of clusters in advance. The algorithm completes the determination of clustering center by transferring two indexes, attraction and attribution, and is relatively suitable for large-scale data sets. Hamzenejad et al. put forward the definition of trajectory data and the mixed regression model of trajectory clustering [11]. Trajectory clustering extends the original isolated point-like distribution of sample points to the queue data on the time axis. Domestic researchers used to call trajectory clustering vividly "spatiotemporal trajectory clustering."

3. Methodology

3.1. Design of College Japanese Test System. Big data not only poses challenges to traditional education, but also makes it easier for teachers to understand and master their students' learning behaviors in real time. Students' behaviors can be digitized during the teaching process, and teachers can use behavior data to identify problems in the teaching mode, which allows them to improve and optimize the teaching mode. Students' learning behaviors will be recorded during the online learning process, forming a large data source that includes each student's mouse click, duration of stay, number of questions asked, number of participants in discussions, and so on. Different people take different paths, and big data can help them find customized solutions. The traditional Japanese major evaluation method is relatively simple, with an emphasis on theory, little application, and an incomplete evaluation method. A single written test can only assess students' basic Japanese knowledge; however, assessing students' practical ability, problem-solving ability, teamwork ability, communication ability, and innovation ability is difficult. We hope that the new assessment method will be able to meet the training and testing of students' abilities. Teachers should become designers, assistants, guides, and responders of students' learning or, in other words, people who serve students.

We can divide the quality structure of students in the comprehensive quality evaluation system into two levels, based on the current situation: obligation level and pursuit level. The most basic requirements for students' ideological and moral cultivation, theoretical study, and physical quality are required courses. The goal of hierarchy is to help students improve their overall quality so that they can develop a noble ideological and moral character as well as a higher level of professional competence. The Japanese examination system is a computer-assisted learning assistant that combines teaching, diagnosis, examination, feedback, and other features. It is built on a system of perfect knowledge structure and a large question bank. Figure 1 depicts the system's fundamental block diagram.

The main functions are to automatically correct errors by the diagnostic engine according to user feedback and to correct subjective problems by using vector space model, matrix singular value decomposition, dynamic programming algorithm and minimum editing distance, and natural language processing technology. The composition is graded automatically, and a user diagnosis report of Japanese knowledge points is formed according to the knowledge points marked in the composition.

3.2. Japanese Intelligent Evaluation. Big data refers to a large network that allows teachers to discover the rules and characteristics of students' learning, prescribe the right medicine, better conform to the trend, improve students' learning behaviors and problems, and track the occurrence and development of teaching with data, forming a virtuous circle. We can provide specific help and personalized learning needs for different learners using learning behavior data mining, as well as discovering different responses of different groups to different knowledge points and different environments. Students are encouraged to learn engineering in an active, practical, and organic way, with a focus on cultivating students' ability in systematic engineering technology, emphasizing students' engineering practice ability, and facing society with no adaptation period.

Teachers are one of the most important aspects of Japanese education. It will be difficult to teach Japanese if teacher quality is poor. As a result, in order to innovate talent training mode, a team of high-quality teachers must be assembled. As a result, application-oriented universities should be able to clarify professors' research directions, allow professors to introduce students to the development direction and academic trend of Japanese occupations through scientific research, promote the combination of teaching and scientific research, and make a connection between scientific research accomplishments and teachers' teaching efforts. Teachers with good research results, according to relevant research surveys, have a higher teaching level and are more likely to be liked by students.

From the perspective of testing, college Japanese composition mainly examines students' comprehensive ability to use Japanese knowledge, including spelling, word collocation, grammar, word selection and sentence structure, understanding of the main idea, design and planning, and rhetorical style. In corpus linguistics, statistics-based processing technology is the main means to obtain various necessary insights from corpus. The basic elements of Japanese are vocabulary and grammar. For subjective problems such as composition, another set of grammar rules based on regular expressions is needed. The system automatically selects the question from the question library according to the set conditions and generates the prototype file of the test paper. Users can download test files from the server to the local computer. At the same time, they can also choose to save them in the question bank for candidates to take the exam online. The program flow chart of this module is shown in Figure 2:

As a tool of educational measurement, test must have good quality to reduce the error of measurement results. Reliability is about the degree to which the measured result deviates from the true value. In measurement, reliability can be defined as the ratio of true score to actual score; namely,

$$a_{xx} = \frac{s_t^2}{s_x^2}. \tag{1}$$

This shows that the greater the proportion of true score variance s_t^2 in actual score s_x^2, the higher the reliability a_{xx}. According to the error equation $s_x^2 = s_t^2 + s_e^2$, (1) can be rewritten as follows:

$$a_{xx} = 1 - \frac{s_e^2}{s_x^2}. \tag{2}$$

It can be seen that if the true score T is close to the actual score X, the s_x^2, s_{-t}^2 will also be close, while the error e, s_e^2 will be small, and the reliability will increase at this time. It can be seen that reliability is a measure of the gap between the measured value and the true value.

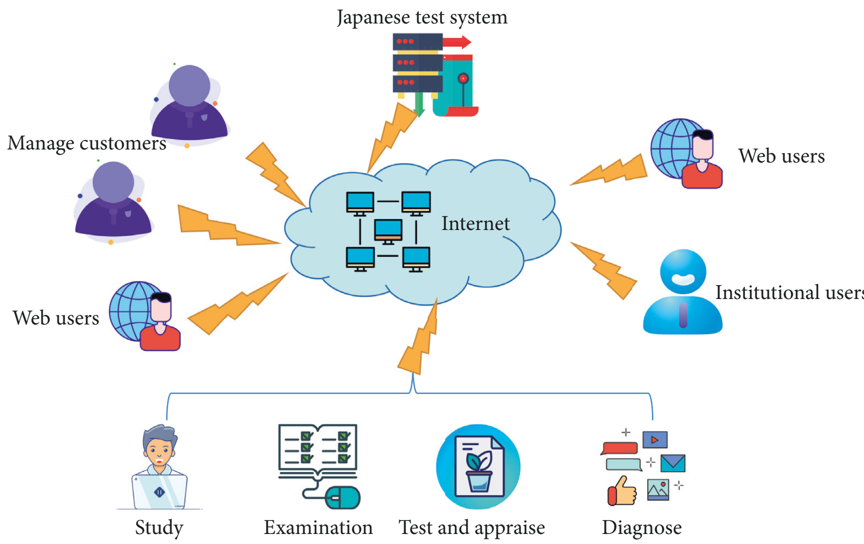

Figure 1: System basic frame diagram.

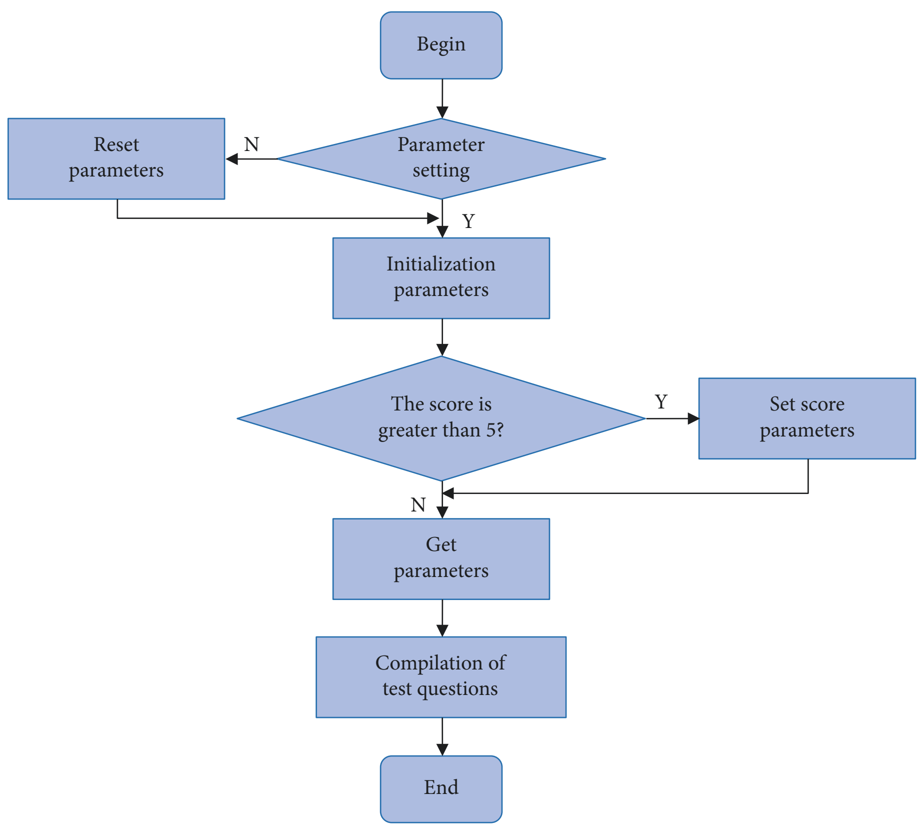

Figure 2: Program flow chart of automatic test paper generating module.

Text similarity is widely used. By setting thresholds to filter the rules and combining the constituent rules in the corpus to make regular matching, we can accurately measure the language abilities such as word placement, vocabulary structure, word selection, and sentence construction. Therefore, it is very important to calculate sentence correlation when diagnosing Japanese composition questions.

$IDF_1, IDF_2, \ldots, IDF_n$ is the weight of each keyword; then, the correlation calculation formula becomes weighted sum [13]; namely,

$$\sum_{i=1}^{n} TF_i \times IDF_i. \tag{3}$$

Let n be the number of times the vocabulary appears in the current sentence, m the total number of sentences in which the vocabulary appears in all other sentences except this sentence, and M the total number of text sentences; then,

$$T_i = n \times \lg \frac{M}{m}. \tag{4}$$

From the above formula, it can be seen that words with more occurrences have larger n values, but such words do not necessarily have higher TF/IDF values. Therefore, the TF/IDF value takes into account the frequency of words and the ability of this word to distinguish different sentences.

After dividing the students' composition into sentences, similarity calculations and regular matching are performed for each sentence and all of the rules in the Atlas, the highest score is calculated, and the corresponding rules are eliminated. After each rule has been exhausted, the composition's final score is given. Given this, the knapsack problem can be transformed into the maximum value of the sum of fully used scores in each rule when the number of rules is limited, and its mathematical expression is as follows.

The objective function is as follows:

$$\begin{gathered} \max f\left(x_1, x_2, x_n\right) = \sum_{i=1}^{n} c_i x_i, \\ \text{s.t.} \begin{cases} \displaystyle\sum_{i=1}^{n} w_i x_i \le p_i, \\ x_i \in \{0,1\}, \quad i = 1, 2, \ldots, n, \end{cases} \end{gathered} \tag{5}$$

where x_i is a 0-1 decision variable, $x_i = 1$ indicates that the rule i matches the sentence successfully, and $x_i = 0$ indicates that the rule i matches the sentence unsuccessfully.

3.3. Cluster Analysis Is Applied to Student Achievement Evaluation. Clustering is the process of categorizing data objects into different classes or clusters, partitioning works on the principle that objects in the same cluster are very similar, whereas objects in different clusters are very different. The classes to be divided in clustering are unknown ahead of time, and class formation is entirely data-driven, indicating that this is an unguided learning method.

Given a data set $X\{x_i | i = 1, 2, \ldots, n\}$, where x_i is a data object, the data set is divided into k groups according to the similarity between the data objects, and the following is satisfied:

$$\begin{gathered} \left\{C_j \mid j = 1, 2, \ldots, k\right\}, \quad C_i \subseteq X,\ C_i \cap C_j = \varnothing, \\ \cup_{i=1}^{k} C_i = X. \end{gathered} \tag{6}$$

Then, the process is called clustering, and $C_i\,(i = 1, 2, \ldots, k)$ is called clustering.

ACA (ant colony algorithm) is a bionic optimization algorithm, which has many advantages and is used to solve the vehicle scheduling problem [12], the color filling problem of connected graph, the salesman problem, and so on. However, ACA has some shortcomings, which will affect the solution of these problems.

In the process of realizing clustering algorithm based on ant foraging principle, we first need to determine the number of target clusters and then select a cluster center for each cluster. The selection of initial cluster center is likely to influence the final clustering effect. The dissimilarity of data objects in the same class is calculated according to the following formula:

$$d_{ij} = d\left(x_i, x_j\right) = \sqrt{\sum_{i=1}^{m} p_i\left(x_{i1}, x_{j1}\right)^2}. \tag{7}$$

The result of clustering is judged, and if the sum of the dissimilarities of all classes is less than the parameter e of clustering end, the clustering is ended [14].

In a two-dimensional grid, ants keep moving, picking up, and dropping down. When ants encounter a data object o_i at position r at some point, we need to calculate its local density by the following expression:

$$f\left(o_i\right) = \begin{cases} \dfrac{1}{s^2} \displaystyle\sum_{o_i} \left[1 - \dfrac{d\left(o_i, o_j\right)}{2}\right], & \text{else}, \\ 0, & f > 0, \end{cases} \tag{8}$$

where $f\left(o_i\right)$ represents the similarity density between this object and other objects around it. $o_j \in \mathrm{Grid}_{s\times s}(r)$. The range area of unit r is $s \times s$.

The lower the average similarity between an object and its neighborhood, the less likely that this data object belongs to this neighborhood [15], and vice versa. The symmetric function Sigmoid is chosen as the probability conversion function according to this principle.

The definition of the probability P_p of a randomly moving mother ant picking up an object without data objects is shown in the following formula:

$$P_p = 1 - \mathrm{Sigmoid}\left(f\left(o_i\right)\right). \tag{9}$$

The definition of $\mathrm{Sigmoid}\left(f\left(o_i\right)\right)$ formula is as follows:

$$\mathrm{Sigmoid}(x) = \frac{1 - e^{-cx}}{1 + e^{-cx}}. \tag{10}$$

It can be seen that if the parameter c is larger, the curve saturation will be faster, which leads to the faster convergence speed of the algorithm.

The flow chart of the improved ACC (ant colony clustering) algorithm is shown in Figure 3.

The basic process of the improved algorithm is described as follows:

(1) Each individual ant in the ant colony is initialized.

(2) All the data objects to be clustered are projected into a plane with a given range; that is, each data object is randomly assigned to a point in the two-dimensional plane, and the coordinate position is (x, y).

(3) At the beginning, each ant is not loaded with any data objects, and the "global memory of historical positions" is set and initialized to be empty.

(4) Calculate the average similarity between data objects. When the ant is not loaded with data objects, its pickup probability is calculated and marked as P_p.

(5) If an object is isolated or the number of objects in its neighborhood is less than a certain constant, mark the object as an isolated point. Otherwise, assign a cluster sequence number to the object and recursively mark its neighborhood objects as the same sequence number.

4. Experiment and Results

The performance of this algorithm is verified by clustering analysis of experimental data and compared with other classical clustering algorithms. The experiment is a comparative test of the effectiveness and performance of the algorithm in this paper. In order to compare the performance of clustering algorithms, the experimental settings are as follows: CPU: Intel Core 2 Duo T5600, 1.83 G; memory: 1.5 GB; hard disk: 100 GB; operating system: Microsoft Windows XP Professional.

Table 1 compares this algorithm, K-means algorithm, and ACA.

As can be seen from Table 1, the clustering analysis results of this algorithm are obviously better than those of K-means algorithm and ACA, the clustering performance is greatly improved, and a better clustering center is obtained. It can be seen that the algorithm in this paper is most commonly used for complex clustering problems with many categories and large amount of data in data mining. The clustering effect of this algorithm is good, and it also overcomes the influence of "noise."

Figure 4 compares the change of the criterion function value with the iteration number of 800 two-dimensional point sets by K-means algorithm and this algorithm, respectively, in 18 categories.

As can be seen from Figure 4 that K-means algorithm is easy to fall into local extremum and difficult to escape. Compared with K-means algorithm, this algorithm converges to the global optimum faster. And the probability is faster, even if the value of the standard function reaches the minimum value. This algorithm is superior to K-means algorithm in terms of iteration times and final convergence value.

Figure 5 compares the standard function values of 15 kinds of clustering analysis of 700 two-dimensional point sets by K-means algorithm and this algorithm 25 times.

As can be seen from Figure 5, the results of K-means algorithm fluctuate greatly during many cluster analysis experiments, while the results of this algorithm have been relatively stable, and the results are better than those of K-means algorithm. After the data is loaded, the ants in the algorithm will move randomly in the 2D plane and then compare the similarity with the surrounding data. If the similarity condition is met, the data will be deleted and reloaded; otherwise, they will move randomly again. The average results obtained from several experiments are shown in Table 2.

It can be seen that when the results are similar, the number of iterations required by the improved algorithm is faster than that of the original algorithm, and the execution time is faster than that of the first one. Because the former may have a local optimal solution, ants cannot pick up data objects. In the unimproved algorithm, cluster deadlock may occur. To sum up, it can be seen that the improved algorithm can improve the convergence speed of the cluster. The algorithm verifies the accuracy of clustering results because the number of class members differs greatly, and the attribute values contain other types of data. The test sample is Car Evaluation Database in UCI machine learning data set. There are four categories in this data set, namely, unacceptable, acceptable, satisfactory, and very satisfactory.

The number of clustered categories is the same as the actual number of categories. The detailed corresponding information of the number of samples to be included in each category and the number of samples from the clustering results is shown in Figure 6.

For 1350 unacceptable categories, the number of wrong samples is 120, accounting for 8.66%. However, the number of acceptable category error samples is 88, accounting for 5.01%. The sample number of satisfactory category errors is 30, with an error rate of 1.66%, and the number of very satisfactory category errors is 35, with an error rate of 1.89. Therefore, the correct clustering rate is about 80%, and the clustering results are in an acceptable high range. The clustering results of the clustering correctness test experiment can verify that the clustering accuracy of the algorithm is above 80%, which shows that the accuracy of the clustering results is high, and it is suitable for the follow-up clustering research. We make statistics on the correct rate and error rate of the following different clustering algorithms and give a comparison chart.

It can be seen from Figure 7 that the grouping results of K-means algorithm, standard ACC algorithm, and enhanced ACC algorithm are divided into three groups. It can be seen that the error rate of K-means algorithm is relatively high, followed by standard ACC algorithm, and the error rate of enhanced ACC algorithm is low. At the same time, comparing the execution efficiency of the three algorithms, we find that the improved ACC algorithm takes less time and has higher execution efficiency. From these two aspects, the improved ACC algorithm shows advantages and can be further studied. For the improvement of ACC algorithm, we only verify it on small data sets, and if conditions permit, we also need to verify the effect of clustering analysis on large data sets.

The classification of students' academic performance over a period of time is referred to as student performance evaluation. Grades have an evaluation function, which is to provide a numerical result by testing students at various levels in order to provide an assessment of students' performance during the learning process. Of course, they also reflect teachers' ability to teach and manage. They are of guiding importance to the school's school-running

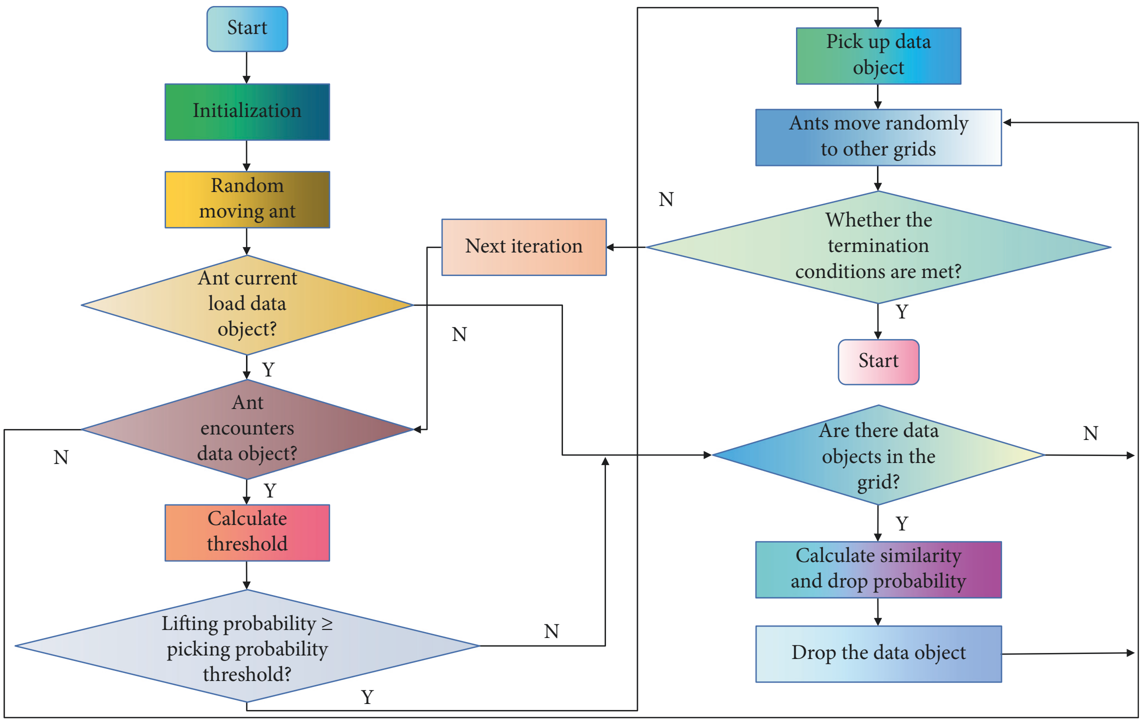

Figure 3: Improved ACC algorithm flow.

Table 1: Comparison of clustering performance of algorithms.

Cluster number	Algorithm in this paper	K-means algorithm	ACA
5	556.32	733.58	604.61
8	367.14	590.25	436.75
12	201.22	533.28	349.02
16	155.48	471.89	218.23

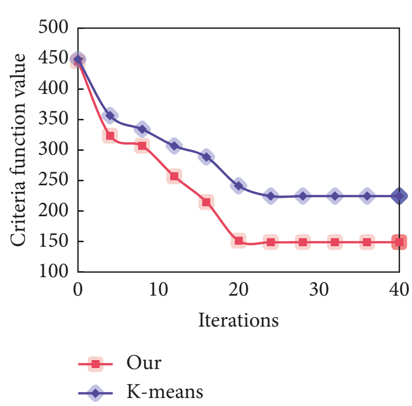

Figure 4: Comparison of the change of criterion function value with iteration number.

mode, talent training mode, and school-running effect in the entire process of students' learning. As a result, performance evaluation is an unavoidable byproduct of teaching activities, and all educational personnel should pay attention to it.

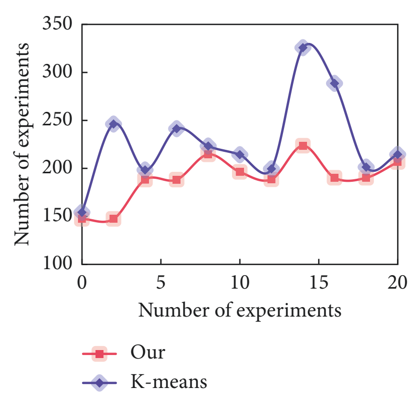

Figure 5: Comparison of the change of criterion function value with the number of experiments.

Table 2: Iteration number comparison.

Algorithm	Iterations	Running time (min)
Before improvement	800	6
After improvement	300	0.7

This paper performs a cluster analysis of student achievement data in the hope of extracting useful information from the data. The end result of grouping is that students' scores are divided into groups. Will the students' study habits, study styles, and class attendance rates be similar in the same group? Our experiments and related research must back up these claims. Then, based on the experimental findings, we should enact appropriate regulations and teach students according to their abilities.

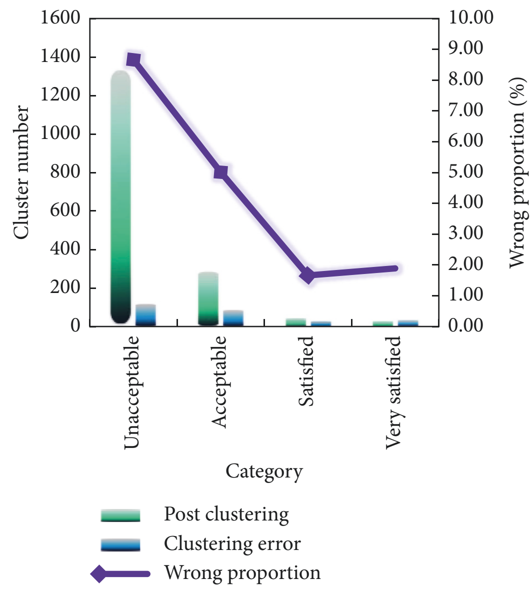

Figure 6: Clustering results.

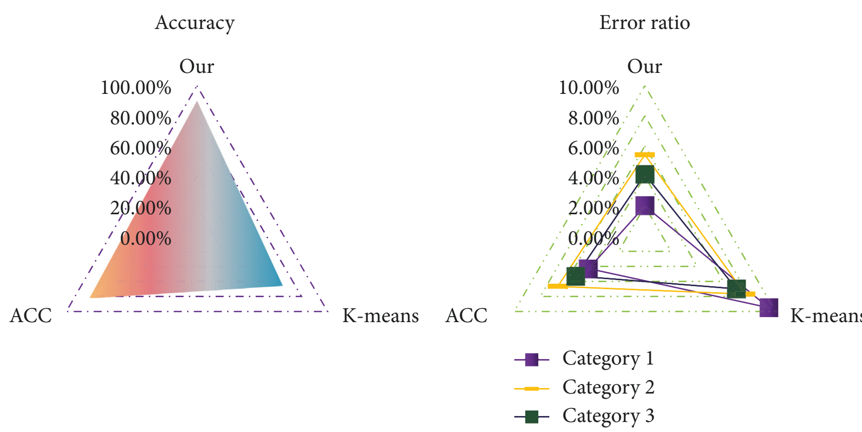

Figure 7: Comparison chart of correct rate and error rate of three algorithms.

Choosing an appropriate algorithm and ensuring the accuracy and integrity of the selected data are critical steps in the data mining process. For cluster analysis, 500 students' Japanese course scores were chosen. We must convert all of the data in the data set into a valid digital form after it has been cleaned. Exam courses and elective courses will appear in some of our score data sets. We must convert their scores into percentage value format because they are divided into pass and fail categories and there are no specific scores. 500 student records are divided into 5 clusters, as shown in Figure 8.

The results of the grouping show that the first type of students' results is generally average, and the results of all subjects are at a medium level, which can be improved in all areas. We can see that these students have a diverse set of interests, which may explain why they do not study some topics in depth. We can encourage these students to explore their passions and hobbies, maximize their potential, and concentrate on developing their own topics. Some students' grades may be relatively high, with more than 90 points, after they study diligently. The students in the fifth category, on the other hand, scored 60-70 points in assembly language programming, data structure, and algorithm or even failed, according to the cluster analysis results. We can see that these students have a weak Japanese foundation. Some questions must be answered in Japanese due to bilingual instruction, which has a significant impact on these students' performance.

5. Conclusions

As economic globalization accelerates, the economic and cultural exchanges between China and Japan have increased the demand for Japanese talent. As a result, application-oriented universities should enhance the innovation of Japanese professional talent training modes and curriculum reform, as well as taking effective measures. The system has auxiliary teaching software that integrates teaching, diagnosis, testing, feedback, and other functions through automatic correction of subjective and

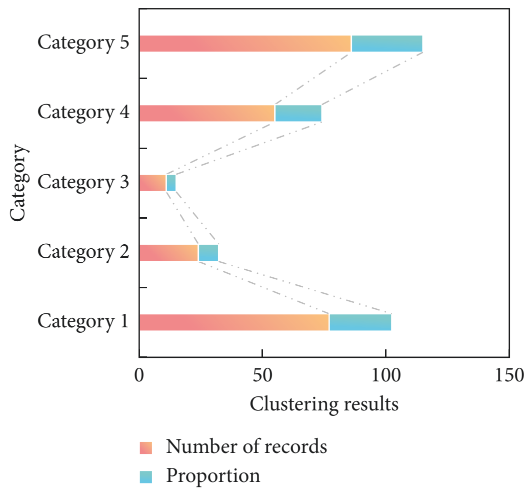

Figure 8: Clustering results of students' achievements.

objective problems and personalized diagnosis design. To record the historical positions of data objects dropped by ants, the "global memory of historical positions" mechanism is introduced, and the improved ACC algorithm is applied to student performance evaluation. The accuracy rate of the clustering algorithm can be verified through the cluster correction test experiment, which shows that the accuracy rate of the clustering result is high, and it is suitable for further cluster investigation. The rationality of cluster analysis is explained based on the actual application.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author does not have any possible conflicts of interest.

Acknowledgments

This study was supported by a study on the structural features of Banana Yoshimoto's works in the literature of the Department of Healing, Item no. SK1704.

References

[1] X. Guan, Y. Fan, Q. Qin, K. Deng, and G. Yang, "Construction of science and technology achievement transfer and transformation platform based on deep learning and data mining technology," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 2, pp. 1843–1854, 2020.

[2] M. Maggioni, "Matrix methods in data mining and pattern recognition," *Journal of the American Statistical Association*, vol. 104, no. 488, pp. 1720-1721, 2009.

[3] G. H. Jiao, J. B. Huang, and H. Huang, "The application of data mining technology in iron and steel industry," *Systems Engineering*, vol. 28, no. 2, pp. 112–116, 2010.

[4] Y. F. Lin and X. P. Hu, "The application of data mining technology in mechanical fault diagnosis," *Key Engineering Materials*, vol. 460-461, pp. 821–826, 2011.

[5] Y. Luo and Y. Xiang, "Application of data mining methods in internet of things technology for the translation systems in traditional ethnic books," *IEEE Access*, vol. 8, pp. 93398–93407, 2020.

[6] H. Zhou, Y. C. Soh, and X. Wu, "Integrated analysis of CFD data with K-means clustering algorithm and extreme learning machine for localized HVAC control," *Applied Thermal Engineering*, vol. 76, pp. 98–104, 2015.

[7] C. M. Chen, D. Pi, and Z. Fang, "Artificial immune K -means grid-density clustering algorithm for real-time monitoring and analysis of urban traffic," *Electronics Letters*, vol. 49, no. 20, pp. 1272-1273, 2013.

[8] J. Sun, W. Liu, and C. Tang, "A novel active equalization method for series-connected battery packs based on clustering analysis with genetic algorithm," *IEEE Transactions on Power Electronics*, vol. 99, p. 1, 2021.

[9] Y. Li, "Research on sports video image analysis based on the fuzzy clustering algorithm," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6630130, 8 pages, 2021.

[10] Z. Yang, L. Wang, and K. Fan, "Exemplar-based clustering analysis optimized by genetic algorithm," *Chinese Journal of Electronics*, vol. 22, no. 4, pp. 735–740, 2013.

[11] A. Hamzenejad, S. J. Ghoushchi, and V. Baradaran, "Clustering of brain tumor based on analysis of MRI images using robust principal component analysis (ROBPCA) algorithm," *BioMed Research International*, vol. 2021, Article ID 5516819, 11 pages, 2021.

[12] J. Miao, Z. Wang, X. Ning et al., "Practical and secure multifactor authentication protocol for autonomous vehicles in 5G," *Software: Practice and Experience*, 2022.

[13] E. Zhu, Y. Zhang, P. Wen, and F. Liu, "Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index," *Neurocomputing*, vol. 363, no. 21, pp. 149–170, 2019.

[14] Y. Tao, Y. Jiang, K. Xia, J. Xue, L. Zhou, and P. Qian, "Classification of EEG signals in epilepsy using a novel integrated TSK fuzzy system," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 3, pp. 4851–4866, 2021.

[15] Y. Ding, X. Zhao, Z. Zhang, W. Cai, N. Yang, and Y. Zhan, "Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, Article ID 5511812, 2022.