International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-11

https://doi.org/10.70917/ijcisim-2026-0141
Article

Application of Data Mining Methods in Analyzing Learning Habits of Chinese Language Chinese Education Students and Improving Teaching Strategies

Xiaoyu Yang *

School of Culture and Media, Zhanjiang University of Science and Technology, Zhanjiang, Guangdong, 524000, China; yangxy202107@163.com

Abstract: This study utilizes multidimensional background information data from 280 students majoring in Chinese language education at a university. Employing data mining and K-means clustering techniques, the study conducts qualitative and quantitative statistical analysis and mining of students' learning data to gain insights into their learning motivations, explore their learning profiles and habits, and uncover the correlations between learning behaviors and academic performance. This research aims to provide theoretical support for teachers in developing teaching strategies and improving instructional practices. Analysis indicates that over 70% of students engage in outdoor activities three or more times, and those who consistently eat breakfast on time and review before class demonstrate superior academic performance. Strict attendance requirements in the classroom are correlated with final course grades. Based on these correlations, teachers can adjust their teaching strategies and methods to enhance instructional effectiveness.

Keywords: K-means clustering; data mining; learning profiles; learning habits

1. Introduction

Chinese language education is an important means of cultivating students' comprehensive understanding of the Chinese language and Chinese literature, as well as preserving and promoting the excellent traditional culture of the Chinese nation [1-2]. In the current educational environment, due to the dry content of Chinese language education and outdated teaching methods, students lack interest in learning, and learning outcomes are not significant [3-5]. This not only requires reforms in educational content and improvements in teachers' teaching abilities but also necessitates a student-centered approach, analyzing students' learning habits to improve teaching strategies and lay the foundation for enhancing Chinese language education [6-8].

Previous analyses of students' learning habits primarily relied on teachers' observations and analyses, which were subject to subjective biases and hindered the implementation of effective interventions. However, with the development of information technology, the application of data mining methods can significantly improve the efficiency and accuracy of analyzing students' learning habits, thereby assisting teachers in refining their teaching strategies [9-12]. Data mining is an interdisciplinary field encompassing knowledge from statistics, machine learning, databases, and other disciplines. In practical applications, data mining methods are often combined with big data technology to help businesses and organizations better utilize their data resources [13-16]. In the analysis of Chinese language education students' learning habits, data mining methods achieve student learning habit analysis through data collection and integration, core analysis techniques, etc., providing references for course design optimization, resource recommendations, and teaching strategy improvements in Chinese language education [17-20].

Reference [21] proposed a student learning performance prediction framework based on behavioral models, aiming to comprehensively describe students' behavioral characteristics. This framework can assess students' mastery of knowledge, provide references for improving teaching strategies, and offer personalized learning services. Literature [22] employs K-Means clustering analysis to examine student behavioral patterns based on learning hours, attendance rates, and tutoring time, with the objective of understanding their impact on learning outcomes. It also highlights the significant role of educational data mining methods in providing personalized education. Literature [23] explores the application of association rule mining in student behavior data analysis, proposing updates to existing mining algorithms to address issues of over-scanning and over-iteration in the original dataset. The study shows that the improved mining algorithm reduces the number of iterations and enhances the efficiency of mining student behavior datasets. Literature [24] highlights the role of learning analytics technology in understanding student behavioral characteristics, proposing the "learning habit mining" method to support students in establishing learning habits, and suggesting intervention measures to enable students to develop learning habits based on evidence from learning logs.

Literature [25] highlights the application and advantages of technology-assisted tools in teaching, while emphasizing the limitations of current research. Based on this, association rule mining and sequence pattern mining methods were employed to analyze the correlation between student learning effectiveness and learning characteristics, effectively providing teaching references for educators and enhancing instructional outcomes. Literature [26] introduces the application of big data in the collection and analysis of student behavioral data, indicating that the extraction of student behavioral characteristics facilitates understanding of students' learning processes and influencing factors, and enables prediction of student performance and adjustment of teaching strategies. Literature [27] aims to analyze effective learning behaviors of students online based on data mining. By constructing a learning behavior mining model and utilizing improved algorithms to mine and analyze learning behavior data, it emphasizes the effectiveness of improved algorithms in student behavior analysis. Literature [28] highlights the important role of learning analytics in improving learning outcomes and introduces commonly used methods in learning analytics. The research findings assist teachers and school administrators in developing learning tools and applying learning analytics to enhance and predict students' academic performance. All the above studies affirm the importance of analyzing student learning behaviors and propose methods such as K-Means clustering analysis, association rule mining, and data mining. These methods facilitate the analysis of students' learning habits and enable teachers to improve teaching strategies based on the deficiencies in students' learning habits.

This paper uses data mining as a tool to analyze the multidimensional background information data of 280 students majoring in Chinese language education at a university, specifically focusing on their academic performance and behavioral habits. Through a questionnaire survey, information was collected on five key indicators: interpersonal relationships, personality traits, family background, social experiences, and health status. The K-Means clustering algorithm was employed to construct a clustering model, uncovering correlations between academic performance and behavioral habits. Based on these findings, recommendations were provided from the perspectives of teaching methods, curriculum design, teaching quality, and decision-making.

2. Related Methods

2.1. Data Collection

First, to obtain multidimensional background information data on students, this study designed a survey questionnaire covering five background domains of students, collecting relevant information from 280 students majoring in Chinese language education at universities. The survey questionnaire included five attributes: interpersonal relationships, personality traits, family background, social experience, and health status.

The five measurement indicators are shown in Figure 1, with each attribute containing three specific measurement indicators. We selected these five attributes because they are generally considered to be important factors influencing students' learning habits. For example, students' interpersonal relationships and social experiences may affect their teamwork abilities and social adaptability. Personality traits may influence their learning styles and motivation. Family background may impact their learning environment and resources. Health status may affect their learning efficiency and endurance. Therefore, information from these five aspects provides us with a comprehensive perspective on students' learning profiles. Additionally, to provide a comprehensive description and evaluation of students' learning habits, we designed a survey questionnaire to statistically analyze students' learning habits. This questionnaire primarily characterizes students' learning habits from five aspects, assessing their learning situations from multiple dimensions.

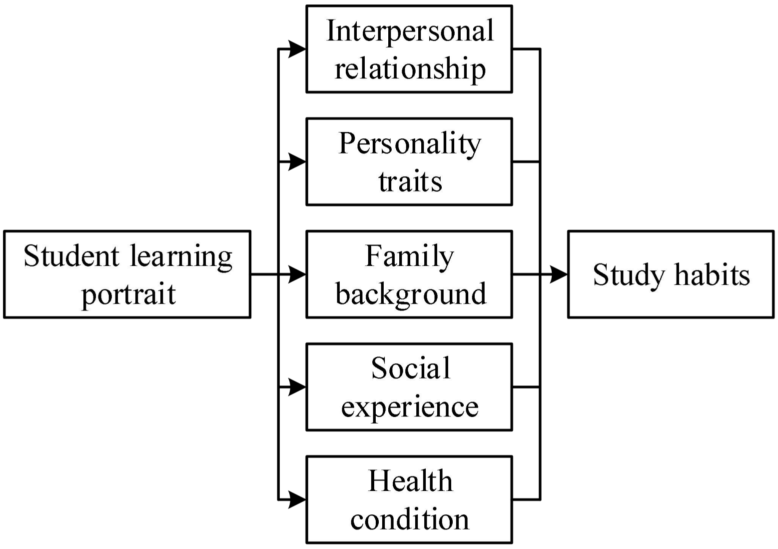

Figure 1 Five measures

2.2. Problem Definition

In this study, we use the type of student learning habits in the student profile as a primary indicator for measuring student learning profiles. The five aspects of background information on students (interpersonal relationships, personality traits, family background, social experience, and health status) mentioned in the statistics are considered as factors influencing student learning habits. We treat the prediction of learning habit types as a classification task, aiming to determine which factors have a higher impact on students' learning abilities. In the experimental design, each student is represented by a five-dimensional feature vector, with each vector containing three metric scores. The label is the student's learning habit type. The learning habit type prediction task is defined as follows:

Problem definition: Given the five-dimensional feature vector of student S , as shown in Formula (1):

$$S = \{s_1, s_2, \cdots, s_n\}, s_i = \{x_1, x_2, x_3\}, n = 5 \tag{1}$$

Among these: s_i represents the indicator scores for the five main factors, and x_i represents the specific indicators under each factor, such as the number of friends in one's social circle, the frequency of social activities, and the score for teamwork ability in interpersonal relationships.

The student's learning habit type is $y \in \{1,2,3,4,5\}$ and corresponds to the five different student learning types $\{A, B, C, D, E\}$. These five types are derived from a learning habit survey questionnaire, with themes including learning time management, learning methods, learning motivation, learning environment, learning pressure, and coping strategies.

The student's learning habit type prediction task can be described as learning a mapping function:

$$F : S \times A \to y \tag{2}$$

Where: A is the mapping matrix, and y is the student's learning habit type. The prediction problem is a classification task, with the goal of predicting the learning habit type of a specific student profile.

The objective of this paper is to identify indicators from various background metrics that are strongly correlated with student learning habits. These indicators will serve as key references for subsequent course design and learning habit enhancement for students.

2.3. Data Mining Methods

2.3.1. Algorithm Flow

The general process of data mining includes: defining the problem, data collection, data preprocessing, data mining, and model representation and interpretation. Defining the problem. The purpose of data mining is to discover knowledge, and clearly defining the objectives of data mining is a solid foundation for successful data mining. Data collection. After clearly defining the problem, relevant data that can solve the problem must be collected.

Data preprocessing. Typically, the collected data is large and unorganized. To ensure the data meets the requirements for data mining, preprocessing operations must be performed. Data preprocessing generally includes data cleaning (filling in missing values, removing duplicate values, etc.), data integration (converting data from different sources and integrating them into a specified format), data selection (removing data that has little or no impact on data mining and filtering out meaningful, relevant data), and data transformation (performing standardized operations on the data). Data mining. This step is the core step in the entire data mining process. Based on the structure and nature of the data obtained through preprocessing, appropriate data mining algorithms are selected to perform data mining and construct corresponding models. For some data mining algorithms, the dataset must be divided into training and testing datasets, with the training dataset used to build the model and the testing dataset used to evaluate the model's performance. Model representation and interpretation. The results obtained from data mining are often a series of data displays. Researchers need to interpret the results manually, remove redundant or irrelevant patterns from the mining results, extract understandable information or patterns, and present them to users in an intuitive and easy-to-understand manner.

2.3.2. K-Means Clustering Algorithm

(1) Definition of K-Means

The K-Means algorithm is a clustering method based on partitioning. This algorithm groups data objects into multiple classes or clusters such that objects within the same cluster exhibit high similarity, while objects in different clusters exhibit significant differences, resulting in dense clusters and sparse inter-cluster relationships. The K-Means algorithm uses object distance as an indicator; the closer two objects are in distance, the higher their similarity. Therefore, each cluster is composed of objects that are close in distance. The main idea of the K-Means algorithm is as follows: first, a constant K is determined in advance, where K represents the final number of clustering categories. Then, K initial points are randomly selected as cluster centers. By calculating the similarity between each sample and the cluster centers, the samples are assigned to the most similar cluster. Next, the centroids of each cluster are recalculated, and this process is repeated until the centroids no longer change, ultimately determining the category to which each sample belongs and the centroid of each cluster.

This paper applies the K-Means algorithm to perform comprehensive clustering of students' behavioral characteristics and academic performance. Based on the characteristics of students within each category, it guides student learning and provides decision-making support for teachers to improve their teaching. Additionally, the categories generated by this algorithm are compared with those obtained from the previous decision tree algorithm to further confirm the association between students' behavioral characteristics and academic performance.

(2) K-Means Workflow

Assuming the input data sample is $S = i_1, i_2, i_3, ..., i_n$, the algorithm workflow is as follows:

a) Clustering begins, and K initial centroids $j_1, j_2, j_3, \ldots, j_k$ are determined.

b) Measure the distance from each sample i_n to each centroid and assign it to the class of the nearest centroid.

c) Recalculate and select new centroids within the already classified classes.

d) Iterate several times until the distance between the new centroids and the original centroids is equal to or less than a specified threshold.

(3) Advantages and disadvantages of K-Means

The K-Means algorithm is simple in principle, easy to implement, suitable for analyzing large-scale datasets, and produces relatively good clustering results. However, it requires the prior determination of the K value, which is very difficult to estimate. In most cases, we do not know in advance how many categories the given dataset should be divided into for optimal results. Additionally, the initial partitioning of the cluster centers has a significant impact on the clustering results; if the initial cluster centers are chosen improperly, effective clustering results may not be obtained [29].

3. Analysis of Student Learning Habits Based on the K-Means Clustering Method

3.1. Descriptive Analysis of Students' Learning Habits

(1) Breakfast Eating Habits

To facilitate statistical analysis and comparison, this study selected the most frequently chosen behavior options from the questionnaires of different student groups for statistical analysis. In terms of breakfast behavior habits, the study selected the option “eat breakfast at a fixed time every day.” As shown in Figure 2, it can be seen that over 60% of all students have the habit of eating breakfast at a fixed time every day, meaning that more than half of the students eat breakfast at a fixed time. Among the various student groups, students in group type “B” have the highest proportion of students with this habit, while students in group type “D” have the lowest proportion. In terms of similarity, the four student groups can generally be divided into two categories: student groups “A” and “B” are similar, while student groups ‘C’ and “D” are similar. The proportions of students in groups “A” and “B” both exceed the overall student proportion, while the proportions of students in groups “C” and “D” are both below the overall student proportion. By comparing the academic performance of students in groups “A” and “B” with those in groups ‘C’ and “D,” it can be observed that there is a correlation between breakfast eating habits and academic performance, with students who consistently eat breakfast at regular times demonstrating superior academic performance.

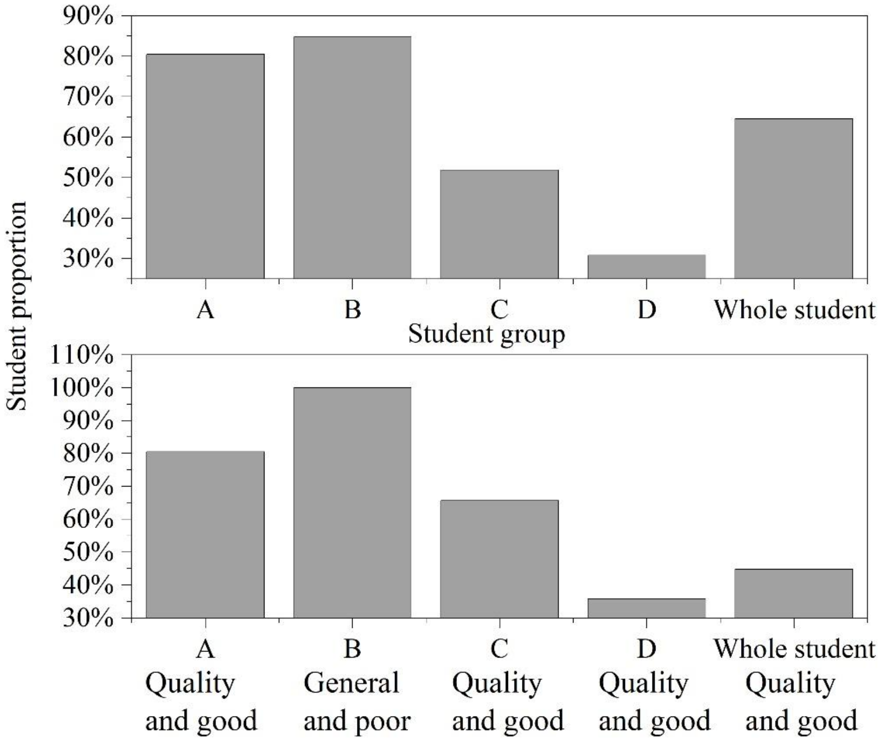

Figure 2 Student group breakfast quality statistics and breakfast behavior profile

(2) Students' activity habits

This paper includes two dimensions in the statistics and analysis of students' activity habits: the number of times they engage in physical activities per week and the duration of these activities. In terms of the number of times they engage in physical activities per week, this paper uses the option of engaging in activities three or more times per week in the questionnaire as the standard for measurement. The following is the distribution of students who engage in activities three or more times per week as a proportion of students of the same type and all students. Students' activity habits are shown in Figure 3.

As shown in the figure, more than half of students in all categories engage in outdoor activities three or more times per week. Among them, students in the “B” category have the highest proportion of students engaging in outdoor activities three or more times per week, while students in the “A” category have the lowest proportion. The proportion of all students engaging in outdoor activities three or more times per week exceeds 70%, indicating that most students are able to relieve themselves from the pressures of academic work through moderate exercise, maintaining physical and mental well-being. Regarding the duration of activities, this study selected the most frequently chosen option in the questionnaire, which was the time segment of “30 minutes to 45 minutes.” Below is the proportion of students from different groups who engage in activities during this time segment. It can be seen that students from groups “B” and ‘D’ have a higher proportion of students who maintain the habit of engaging in activities for “30 minutes to 45 minutes,” indicating that these two groups have a longer duration of physical activity compared to other groups. It can be seen that both types of students have a preference for physical activity. Students in group type “C” have the lowest proportion in this habit. Compared to the overall student population, students in group type “C” maintain a higher frequency of outdoor activities but do not invest significantly in activity duration.

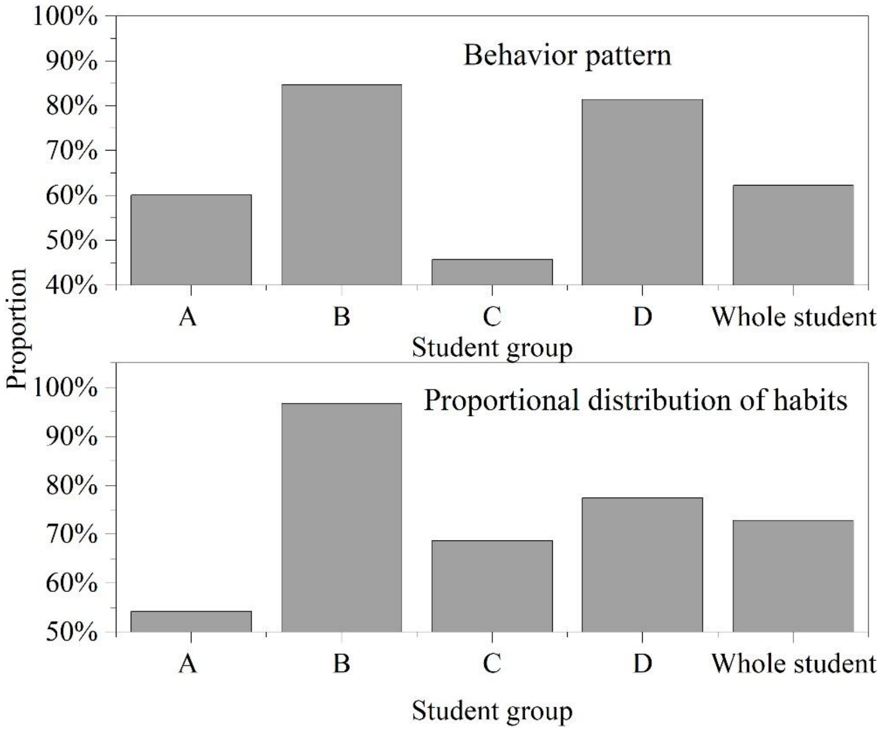

Figure 3 Student behavior habits

(3) Students' pre-class preparation habits

This study's investigation into students' pre-class preparation habits includes two aspects: the duration of pre-class preparation and attitudes toward pre-class preparation. Regarding the duration of pre-class preparation, this study selected the questionnaire options “15–30 minutes” and “over 30 minutes” as the standard for pre-class preparation duration, meaning that students who chose “over 15 minutes” for pre-class preparation time were considered to have the habit of preparing for class for over 15 minutes.

As shown in Figure 4, it can be observed that students do not place a high priority on the duration of pre-class preparation before assignments. Only 44.8% of all students engage in pre-class preparation for 15 minutes or more. However, students in groups “A,” “B,” and “D” do place importance on this habit. Among these habits, students in group type “A” have the highest proportion at 64.5%, while students in group type “D” have the lowest proportion at 46.25%, but both exceed the overall student proportion in this habit. Among the four student groups, students in group types “A” and “B” have a higher proportion of students with a pre-class preparation time of over 15 minutes than students in group types ‘C’ and “D.” It can be seen that students in groups “A” and “B” place greater emphasis on the habit of previewing before assignments. The attitudes of students toward previewing before class also reflect their emphasis on this habit. This study statistically analyzed the choices made by students regarding their attitudes toward previewing before class and compiled the most frequently selected options.

It can be seen that the overall attitude of students toward whether pre-class preparation is helpful for in-class learning is not very positive, with only 36.8% of students believing it is very helpful and 42.7% believing it is sometimes helpful. However, the attitudes of students in groups A and B are completely opposite, with 61.8% and 51% of students, respectively, believing that pre-class preparation is very helpful for in-class learning.

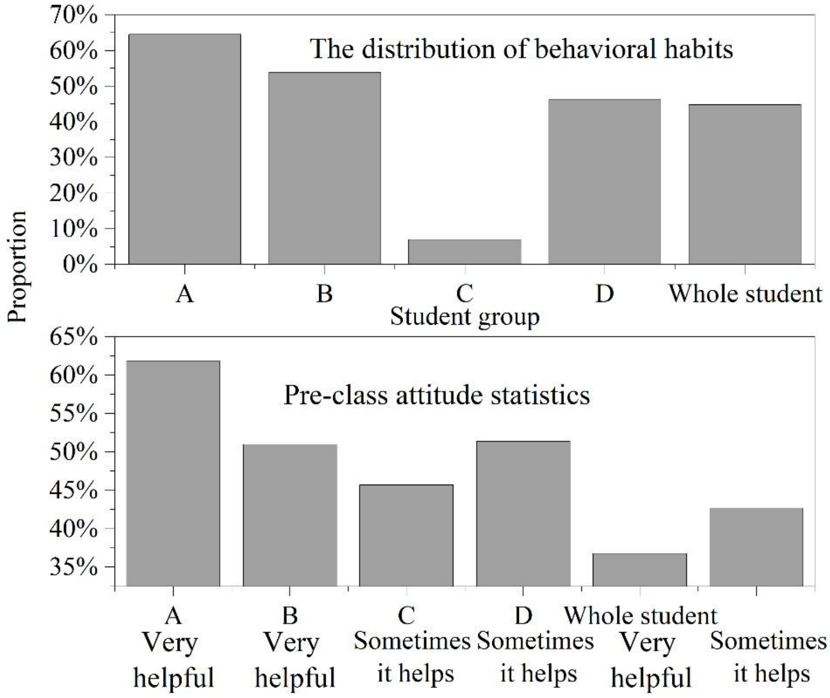

Figure 4 Students' habits

(4) Students' review habits

This study investigated students' review habits in two aspects: the duration of pre-assignment review and post-class review. In terms of the duration of pre-assignment review, this study selected review habits with a duration of 15 minutes or more as the statistical measurement standard. The distribution of students with pre-assignment review durations of 15 minutes or more across different student types is shown in Figure 5.

It can be observed that in terms of after-class review habits, students in group type “A” demonstrate methods that are more effective in understanding and absorbing learned knowledge compared to other student groups. Their proportion within their own group type is also the highest, significantly exceeding the overall student proportion in this after-class review habit category. Students in group type “C” have the simplest methods for post-class review compared to other student groups. This group has fewer habits for pre-assignment review and post-class review. Although their performance is better than that of students in group type “D,” they should still feel a sense of urgency, as there is a possibility of being overtaken.

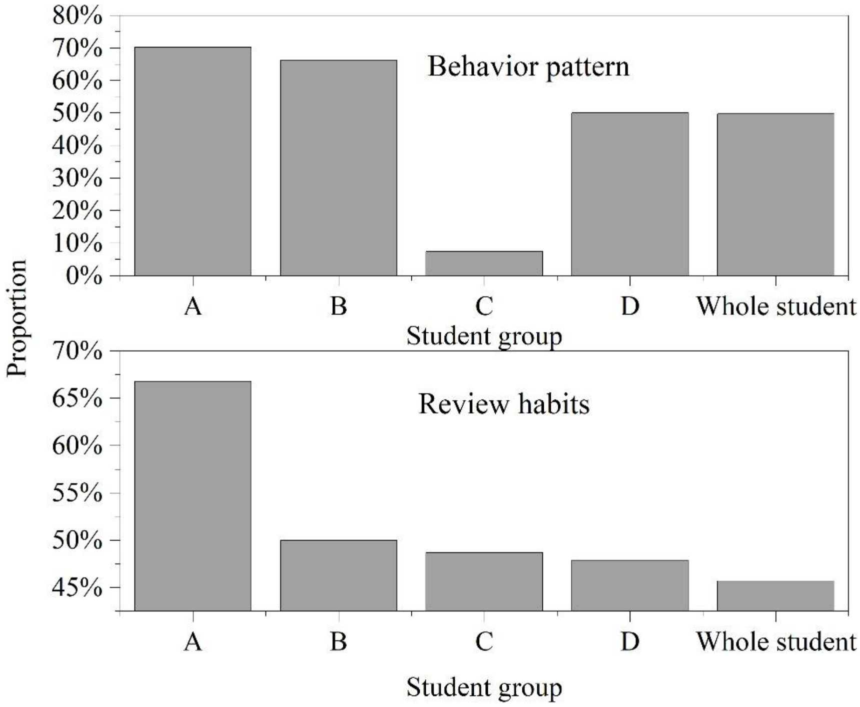

Figure 5 Students review habits

3.2. Cluster Analysis of Students' Learning Habits

A K-means clustering analysis was conducted on four indicators of students' learning habits: class attendance rate, book borrowing volume, number of visits to the library, and duration of study time. When the number of clusters was set to four, the clustering results were optimal. The proportions of students in each cluster and the average values of each indicator for each cluster are shown in Table 1. As shown in the table, the first type of student has a high attendance rate in class, borrows few books, visits the library frequently, and has a long study duration. These students likely attend classes regularly and study in self-study rooms and libraries, but borrow few books from the library, possibly because they prefer to read books in the library without borrowing them.

The second type of student has a relatively low attendance rate, poor classroom attendance, borrows few books, visits the library infrequently, and has a shorter study duration. This indicates that such students often skip classes, lack sufficient effort in their studies, and are among the less academically motivated students in the school. School counselors should closely monitor the academic performance of such students and promptly urge them to strengthen their study discipline and cultivate good study habits. The third type of student has a high attendance rate in class, borrows a significant number of books from the library, visits the library frequently, and spends a considerable amount of time studying. These students are likely to be diligent and hardworking, making consistent efforts both in and out of class, willing to invest time and effort into learning, and possessing good study habits. The fourth type of student has average attendance rates in class, borrows fewer books from the library, visits the library less frequently, and spends less time studying in the self-study room. These students merely attend classes regularly but spend little time studying after class, exhibit low levels of effort in their studies, and invest minimal energy in their academic pursuits. They may belong to the category of students who neglect their studies during regular periods and resort to last-minute cramming before exams. It is recommended that such students cultivate regular study habits and reasonably allocate their study time. Overall, students' attendance rates in class are generally good. The reasons for this may be, on one hand, that teachers strictly enforce attendance requirements linked to final course grades, and on the other hand, that some students may have others swipe their cards for them. Students' book borrowing rates are relatively low, with some students not borrowing a single book in an entire semester. It is recommended that students expand the breadth and depth of their knowledge by increasing their book borrowing rates, as the school library is a treasure trove of knowledge.

Table 1 student learning clustering results

Learning class	Student ratio	Attendance rate	Book borrowing	Entry and exit library times	Learning length	Comparison
1	32.68	95.87	7.8	30.2	154.7	HLHH
2	11.95	86.28	3.5	11.7	85.7	LLLL
3	12.45	92.46	16.4	26.8	133.6	HHHH
4	43.28	90.17	8.5	17.4	117.6	LLLL
Mean		91.87	8.7	22.8	127.8	

3.3. Improvements to Teaching Strategies

3.3.1. Selecting Teaching Methods

In the process of higher education teaching, instructors typically employ one or more teaching methods and approaches to fulfill the instructional objectives of their courses, such as lecture-based instruction, discussion-based instruction, experimental methods, self-directed learning, demonstration-based instruction, and case-based instruction. Different courses may employ different teaching methods, and these methods are typically stored in a teaching management data warehouse. Therefore, data mining techniques can be used to extract useful data from the data warehouse. By applying data mining functions such as regression linear analysis and association rules to analyze the data, instructors can make informed decisions about which teaching methods are suitable for specific professional classes and student groups, optimize teaching strategies to meet student needs, and adopt different teaching methods for different professional classes to address varying instructional content requirements.

3.3.2. Arranging Course Offerings

Currently, there is an issue of unreasonable course settings in universities. For example, students are unwilling to study certain courses, and even if they do study them, they will not be useful in society. Offering such courses is a waste of teaching resources. By utilizing data mining's association analysis and statistical analysis functions, useful rules and information can be extracted from the massive amounts of data in the teaching management data warehouse. This can guide universities to make reasonable course settings that are conducive to improving student performance and employment prospects.

3.3.3. Evaluating Teaching Quality

By utilizing data mining methods to analyze and process the massive amounts of data collected in the teaching evaluation data warehouse, including student academic performance data, behavioral records, and student feedback on teacher instruction, it is possible to promptly obtain evaluation results regarding student learning and teacher instruction effectiveness. This enables timely correction of any negative learning behaviors exhibited by students during the learning process and allows for prompt improvements to issues arising in teacher instruction, thereby achieving the optimal integration of the teacher's guiding role and the student's active role in the learning process.

3.3.4. Decision-Making Student Behavior

By leveraging data mining techniques such as association analysis, clustering analysis, deviation analysis, and evolutionary analysis, valuable data can be extracted from the massive datasets in a management data warehouse. This enables the analysis of students' daily behavioral patterns, the identification of intrinsic connections between various behavioral activities, and the understanding of students' learning status. This enables teachers to help students achieve more effective learning outcomes in a shorter time, helping students avoid detours, thereby enhancing their overall quality and learning abilities, refining their character, and contributing to improving the overall spiritual outlook and learning atmosphere of the school.

4. Conclusion

To investigate the application of data mining methods in analyzing the learning habits of Chinese language education students, this study took 280 students majoring in Chinese language education as the research subjects, collected their multi-dimensional background information data, and employed the K-means algorithm and data mining techniques for big data processing. The aim was to analyze the relationship between college students' behavioral characteristics and learning outcomes, and based on this analysis, to improve teaching strategies and enhance teaching effectiveness. The study found that statistical analysis was conducted on eight questions across four dimensions: breakfast habits, exercise habits, pre-class preparation, and post-class review. These included breakfast habits, breakfast quality, frequency and duration of outdoor exercise, pre-class preparation time, attitude toward pre-class preparation, review time before homework, and post-class review methods. The results showed that students who regularly ate breakfast, engaged in regular exercise, and had habits of pre-class preparation and post-class review tended to perform better academically.

Funding

This research was supported by the Zhanjiang University of Science and Technology + Offline First Class Course Modern Chinese I Project (PJHYLKC-2023669); Higher Education Research Project of Guangdong Higher Education Association + Harmonious Coexistence between Humans and Machines: Discourse Analysis of Chinese Classroom Interaction from an Interdisciplinary Perspective (23GYB102).

References

1. Zang, L., & Chen, D. (2025). A Review of Research on the Application of Case Teaching Method in International Chinese Language Education. International Educational Research, 8(2), p108-p108.
2. Wang, W., & Ruan, J. (2016). Historical overview of Chinese language education for speakers of other languages in China and the United States. Chinese language education in the United States, 1-28.
3. Han, J. (2024). Translanguaging as a pedagogy: Exploring the use of teachers’ and students’ bilingual repertoires in Chinese language education. Applied Linguistics Review, 15(4), 1433-1451.
4. Jiang, X., Li, J., & Chen, C. H. (2024). Enhancing Critical Thinking Skills with ChatGPT-Powered Activities in Chinese Language Classrooms. International Journal of Chinese Language Teaching, 5(1).
5. Akram, T. M., Ijaz, A., & Ikram, H. (2017). Exploring the factors responsible for declining students' interest in chemistry. International Journal of Information and Education Technology, 7(2), 88.
6. Kern, R. (2024). Twenty-first century technologies and language education: Charting a path forward. Modern Language Journal (John Wiley & Sons, Inc.), 108(2).
7. Hariyasasti, Y. (2025). The Influence of Competence, Work Culture on the Performance of Elementary School Teachers in Gunungwungkal District. INTERNATIONAL JOURNAL OF SOCIAL, POLICY AND LAW, 6(1), 37-41.
8. Kaur, J., & Singh, P. (2020). Study habits and academic performance: A comparative analysis. European Journal of Molecular & Clinical Medicine, 7(7), 6161-6166.
9. Zhang, N., Biswas, G., & Hutchins, N. (2022). Measuring and analyzing students’ strategic learning behaviors in open-ended learning environments. International Journal of Artificial Intelligence in Education, 1-40.
10. Chaffee, R. K., Briesch, A. M., Johnson, A. H., & Volpe, R. J. (2017). A meta-analysis of class-wide interventions for supporting student behavior. School Psychology Review, 46(2), 149-164.
11. Gramigna, A., Boschi, C., & Camargo, R. A. (2025). Metacognitive analysis of study habits: Representation and practices of the students. Formazione & insegnamento, 21(3), 1-xx.
12. Sharofitdinovna, N. P., Alamovich, K. A., & Dilovarovna, B. J. (2024). ANALYSIS OF A HEALTHY LIFESTYLE AMONG STUDENTS. Web of Medicine: Journal of Medicine, Practice and Nursing, 2(4), 30-33.
13. Haq, S. U., & Khalil, A. (2022). Identifying common study habits of high achievers of undergraduate students in Pakistan: a thematic analysis. Journal of Development and Social Sciences, 3(2), 986-995.
14. Algarni, A. (2016). Data mining in education. International Journal of Advanced Computer Science and Applications, 7(6), 456-461.
15. Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. IEEE transactions on knowledge and data engineering, 26(1), 97-107.
16. Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. Wiley Interdisciplinary Reviews: Cognitive Science, 6(4), 333-353.
17. Chen, L., Wang, L., & Zhou, Y. (2022). Research on data mining combination model analysis and performance prediction based on students’ behavior characteristics. Mathematical Problems in Engineering, 2022(1), 7403037.
18. Feng, G., & Fan, M. (2024). Research on learning behavior patterns from the perspective of educational data mining: Evaluation, prediction and visualization. Expert Systems with Applications, 237, 121555.
19. Rabelo, A., Rodrigues, M. W., Nobre, C., Isotani, S., & Zárate, L. (2024). Educational data mining and learning analytics: A review of educational management in e-learning. Information Discovery and Delivery, 52(2), 149-163.

20. Kurdi, M. M., Al-Khafagi, H., & Elzein, I. (2018, November). Mining Educational Data to Analyze Students' Behavior and Performance. In 2018 JCCO Joint International Conference on ICT in Education and Training, International Conference on Computing in Arabic, and International Conference on Geocomputing (JCCO: TICET-ICCA-GECO) (pp. 1-5). IEEE.
21. [21] Na, W. (2020). A Data Mining Method for Students' Behavior Understanding. International Journal of Emerging Technologies in Learning, 15(6).
22. Durachman, Y., & Rahman, A. W. B. A. (2025). Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement. Artificial Intelligence in Learning, 1(1), 35-53.
23. Wang, T., Xiao, B., & Ma, W. (2022). Student behavior data analysis based on association rule mining. International Journal of Computational Intelligence Systems, 15(1), 32.
24. Chia-Yu, H. S. U., HORIKOSHI, I., MAJUMDAR, R., & OGATA, H. (2023, December). Learning Habits Mining and Data-driven Support of Building Habits in Education. In International Conference on Computers in Education.
25. Lee, P. J., & Wu, T. Y. (2022). Mining relations between personality traits and learning styles. Information Processing & Management, 59(5), 103045.
26. Wen, Y., Zhang, S., & Yu, M. (2024, November). Research on Data Mining Algorithms for Vocational Education Based on Student Behaviour Analysis. In Proceedings of the 2024 9th International Conference on Intelligent Information Processing (pp. 369-373).
27. Yang, J. (2021, March). Effective learning behavior of students' internet based on data mining. In 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) (pp. 847-850). IEEE.
28. Wibawa, B., Siregar, J. S., Asrorie, D. A., & Syakdiyah, H. (2021, April). Learning analytic and educational data mining for learning science and technology. In AIP conference proceedings (Vol. 2331, No. 1). AIP Publishing.
29. Fengxiang Zhang & Feifei Wang. (2024). Study on abnormal behaviour recognition of MOOC online English learning based on multi-dimensional data mining. International Journal of Continuing Engineering Education and Life-Long Learning,34(1),111-122.