International Journal of Computational Intelligence Systems (2022) 15:32
https://doi.org/10.1007/s44196-022-00087-4

RESEARCH ARTICLE

Student Behavior Data Analysis Based on Association Rule Mining

Tengfei Wang[1] · Baorong Xiao[2] · Weixiao Ma[1]

Received: 28 December 2021 / Accepted: 27 April 2022
© The Author(s) 2022

Abstract
With the advancement of intelligent campus data acquisition technology, student behavioral data are growing in size, variety, and real-time throughput, posing challenges to the storage capacity and computing power of traditional behavioral data analysis methods. The study focuses on the application of association rule mining in student behavioral data analysis. Data collection, storage, computation, and analysis all comprise integral parts of a four-layer data association mining architecture, and the three-step mining process from "data preprocessing" to "finding association rules" to "acquiring relevant knowledge" is described. The existing mining algorithm is updated to address the issues of overscanning of the original dataset and excess iterations. The findings from the case study reveal that the number of iterations in the modified mining algorithm is greatly lessened, effectively improving the mining efficiency of the massive student behavioral dataset.

Keywords Behavior data · Data analysis · Association rule · Mining · Algorithm

Abbreviations

FID	The data serial number.
ACL	Average consumption level
BF	Breakfast frequency
ADOH	Average daily online hours
CMF	Canteen meal frequency
BBA	Books borrowed amounts
AP	Academic performance

1 Introduction

The mass data generated with the continuous development in Internet technology exhibit a discrete and isolated state. It is also difficult to deeply integrate and intelligently process them by computer due to the lack of semantics. Association knowledge represents the relationship between events. Analyzing and refining association knowledge can reveal some potential laws between real-world things and provide guidance for work practice. Association rule mining can implement semantic association between different data sources through data integration and achieve the purpose of comprehensive data sharing, making it convenient for users to further analyze and mine data to access valuable information, and providing effective data support for users' scientific decision making.

Currently, the association rule mining has been widely used in finance, education, transportation, and other fields, and provided scientific data support for decision-making in related fields. Yu and Zhang [1] applied association rule mining to credit classification research and combined it with the feature bagging method to build a new weighting integration model, which provides a reference for loan default risk prediction. Cao et al. [2] used association rules to mine the impact of traffic, environment, and other factors on pavement conditions, to provide scientific and reliable data support for reasonable pavement maintenance. Hu and Guo [3] explored the association rules between $PM_{2.5}$ pollution law and other air pollutants in the pollution season of the urban agglomeration along the Yellow River in Ningxia through Apriori algorithm, suggesting that the pollution control of $PM_{2.5}$ in the urban agglomeration should focus on reducing SO_2. To make full use of the big data about power production safety accidents and incidents, Chen et al. [4] used association rule mining to screen and analyze the key causes for these accidents, which helped improve the efficiency of handling the power production safety accidents and incidents. To improve the accuracy of library information push, Li [5] proposed a

✉ Tengfei Wang
tfwang16@163.com

[1] School of Intelligent Manufacturing and Information Engineering, Shaanxi Energy Institute, Xianyang 712000, China

[2] School of Marxism, Weinan Normal University, Weinan 714099, China

Published online: 19 May 2022

robust association rule algorithm in the field to deal with mass data about book readings and explored the relevance between book information and readers' personalized push. Against the problem of missing or invalid data collected by traffic sensors, Ariannezhad and Wu [6] proposed a systematic method of identifying and describing data error patterns by using association data mining to eliminate the faults of large-scale loop detector. Guo et al. [7] used association rules to mine the specific travel needs of different travel groups, recommending situational route planning in line with individual preferences in the scenic spot.

With the diversification of intelligent data acquisition devices and the gradual increase of data sources and channels, the collected student behavior data are increasing at an alarming rate. These mass data are of high analytical value. Considering the limited storage capacity and backward computing efficiency of traditional data analysis methods, it is difficult to meet the analysis requirements of the current mass student behavioral data. On this basis, this paper introduces the distributed storage and parallel computing technologies into the analysis of student behavioral data, constructs a four-layer data association mining framework including data collection, storage, calculation, and analysis for mass data processing, and defines the three-step mining process from "data preprocessing" to "finding association rules" to "acquiring relevant knowledge", on which basis to improve the existing data association rule mining algorithm. Finally, the mining efficiency of the improved algorithm is tested by mining the association rules of student behavioral dataset.

2 Challenges to and Countermeasures Against Student Behavioral Data Association Mining

2.1 New Features of Student Behavioral Data

With the continuous popularization of university information application systems, mass student information has been created, including personal information, course grade, scientific research information, web browsing records, book borrowing records, campus consumption records, and other behavioral data related to students' campus activities. These data reflect students' status on campus in multiple dimensions, such as students' behavioral habits, learning styles, academic competitions, life trajectories and interpersonal communication. Using the data association mining technology to analyze these mass original data and explore potential value information is conducive to finding out students' behavioral characteristics and behavioral laws and depicting students' portraits. This application also provides important data reference value for the continuous improvement of education and teaching management and service in colleges and universities.

When compared with the traditional student behavioral data, the data collected through the current analysis have the following four new main features: first, the mass data size: as students' environment becomes more and more complex, the amount of data describing students' behaviors is on the explosive increase. Second, diversified data types: as the data acquisition devices become diversified, the traditional single-structured data have gradually evolved into semi-structured or unstructured data in textual, audio, visual, and other forms. Third, rapid data processing: with the continuous development of technology and the continuous change of social environment, data processing, and analysis are accordingly required to be efficient, which entails the implementation of real-time processing to respond to emergency in some fields. Fourth, the implicit data value: although the mass data collected by various devices appear irrelevant on the surface, they contain rich potential value. Using reasonable methods for mining and analysis can provide reliable data support for relevant decision making.

2.2 Shortcomings of Traditional Association Mining Rules

Most of the mass student behavioral data come from different data acquisition devices. Considering the security and efficiency of data processing, distributed database is also the top choice for data storage. Traditional data analysis methods have some shortcomings in analyzing student behavioral data.

First, with the continuous increase of student behavioral data, the standard on computational performance and stability of data processing devices is becoming increasingly stringent, while the traditional forms of data association mining in single-machine mode fall far short of the requirements. Second, as data acquisition devices are diversified in types, heterogeneous data are becoming a commonplace. Traditional data processing methods are mainly dedicated to finding relevant laws by analyzing structured data, while making no breakthrough in analyzing semi-structured and unstructured data [8]. Third, with the continuous change of students' environment, the causal relationship between data presents diversified association forms such as 1-to-1, 1-to-n and m-to-n. However, traditional data mining techniques focus on searching the causal relationship in the analytic process. For the current diversified association forms of data, it is difficult to trace the causal relationship.

2.3 Association Mining of Mass Student Behavioral Data

In the era of data explosion, fast and perspicuous correlation analysis is more practical than causal analysis verified

by strictly controlled experiments. Mass data analysis aims mainly to mine and explore the explicit and implicit association relationships between data by association rules. It can address the defects of traditional single-machine computing and storage capacity, support the processing of multi-source heterogeneous data, and only need association mining to analyze the data association relationship. The operating steps are relatively simple. Therefore, by introducing the distributed storage and parallel computing technologies, a distributed association rule mining framework is built for student behavioral data, the association rule mining process is clarified, the existing mining algorithms are improved, efficient analysis of student behavioral data is implemented, and the useful knowledge contained in student behavioral data is mined, so as to provide data support for campus management to make scientific decisions.

3 Mining Association Rules of Student Behavioral Data

3.1 Association Rule Mining

As a key step of data mining, association rule mining traverses throughout the dataset. First the existing frequent itemsets are found, and then the association rules are constructed according to the correlation between the frequent itemsets.

$R = \{r_1, r_2, \ldots r_m\}$ represents the set of data items, W represents the transaction set, and $F = \{f_1, f_2 \ldots f_m\}$, $(f_i \subset R)$ represents a transaction in W. The unique identification is represented by FID. The sets composed of several data items are represented by A and B, respectively, and the implication in the form $A \rightarrow B$ represents the association rules in W ($A \subset R$, $B \subset R$ and $A \cap B = \emptyset$).

The task of association rule mining is to use data mining algorithm to compare the user preset minimum support (denoted by s) and minimum confidence (denoted by c) according to the given transaction set W, so as to find the qualified association rules. Support indicates the frequency of A and B appearing in W at the same time, that is, the ratio of transactions containing A and B in W to the total transaction in W [9]. (See Eq. (1)):

$$s(A \rightarrow B) = \frac{\text{num}(A \cup B)}{\text{num}(W)}. \tag{1}$$

Confidence indicates the strength of association rules, i.e., the transaction numbers of A and B are simultaneously contained in W, that is, the ratio of the number of simultaneous occurrences of A and B to the number of individual occurrences of A [10]. (See Eq. (2)):

$$c(A \rightarrow B) = \frac{s(A \cup B)}{s(A)}. \tag{2}$$

Here, it is assumed that the user preset minimum support is min s and the minimum confidence is min c. If $s(A \rightarrow B) > \min s$ and $c(A \rightarrow B) > \min c$, then the association rule $A \rightarrow B$ is said to coincide with the circumstance of a strong association rule.

3.2 Framework of Association Rule Mining for Student Behavioral Data

Student behavioral data association mining is based on the distributed storage and parallel computing architecture. For all kinds of multisource heterogeneous student behavioral data, the association rule mining algorithm is used to mine the data items that comply with strong association rules, and meaningful connections are found through analysis.

According to the principle of association rule mining and the characteristics of student behaviors, the framework of student behavioral data association rule mining is designed (see Fig. 1) to comprise data collection, storage, computation, and analysis from bottom to top. Among them, the data collection layer, as the basis of association rule mining, is mainly responsible for the collection of student behavioral data through various network devices, manual investigation, and system log; the data storage layer builds a computing cluster to store the mass multisource heterogeneous student behavioral data provided in blocks by the collection layer, and provides mass data and high-speed data reading and writing services; the data computing layer adopts the distributed computing model MapReduce to implement the calculation and processing of mass data; on the data analysis layer, researchers use the association rule mining algorithm to analyze the computed results and finally acquire useful knowledge.

3.3 Student Behavioral Data Association Rule Mining Process

The purpose of association rule mining is to find strong association rules from the collected mass student behavioral data, and to acquire the relevant knowledge contained in the data through further analysis. It consists of three steps: data preprocessing, finding association rules, and acquiring relevant knowledge (see Fig. 2).

In the data preprocessing stage, the original dataset is processed through data purification, consistency processing, abstract description, and scale compression. With the data required by the association rule mining mode as the standard, data normalization is implemented while outliers are eliminated from the dataset; in the stage of finding association rules, the distributed storage and parallel computing

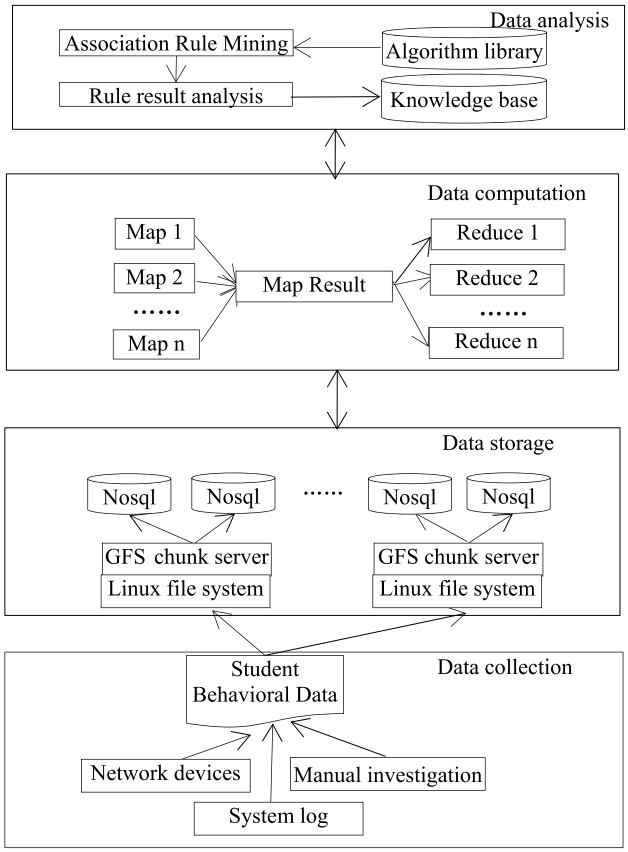


Fig. 1 Framework for association rule mining of student behavioral data

technologies are used to find frequent itemsets according to user requirements, and relevant mining algorithms are used to obtain strong association rules from the frequent itemsets that meet user requirements; the stage of acquiring relevant knowledge is to refine the mining results at the educational level and acquire useful knowledge for reference and decision-making of school's decision-making management.

4 Association Rule Mining Algorithm of Student Behavioral Data and Improvement

Apriori algorithm is a classic data association rule mining algorithm, which has the defect of low efficiency of mining frequent itemsets in a large data size. In order to improve its efficiency, many researchers have optimized the Apriori algorithm in different aspects. Against the low parallelization efficiency of density-based clustering algorithm, Yu et al. [11] divided the algorithm into three stages: data division, local clustering, and global clustering, and formulates the corresponding strategy for algorithm improvement. To improve the mining efficiency of educational big data, Xu and Hoang [12] proposed a random forest reference model with a feature weighting system by analyzing the existing data mining models. Considering that the traditional cluster verification indicators could not correctly handle the increasing dataset capacity, Zerabi et al. [13] proposed two parallel and distributed models using MapReduce framework to implement these indicators. Heidari et al. [14] designed a density-based clustering algorithm to overcome the problem that traditional algorithms could not find clusters with different densities, using MapReduce distributed programming model to split and cluster large datasets. In order to reduce the partition deviation between reducers and give better play to the parallel performance of MapReduce, Wang et al. [15] proposed an incremental data allocation method to divide the mapped data and count their size information at the same time. Considering the operational advantages of multicore CPU, Literature [16] proposed an efficient frequent itemsets mining algorithm based on the MapReduce model, which can mine all frequent K itemsets by converting the block data into a matrix, thereby improving the mining efficiency. The mining process is shown in Fig. 3.

After several case verification, two problems have been identified in the algorithm that need to be addressed:

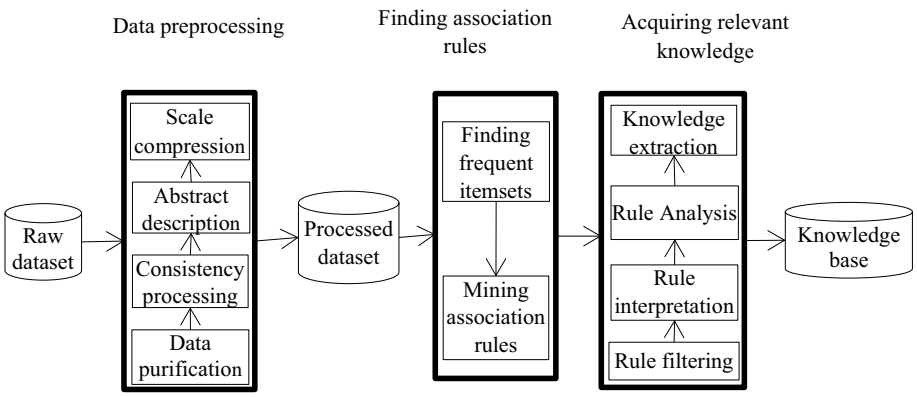


Fig. 2 Process of association rule mining of student behavioral data

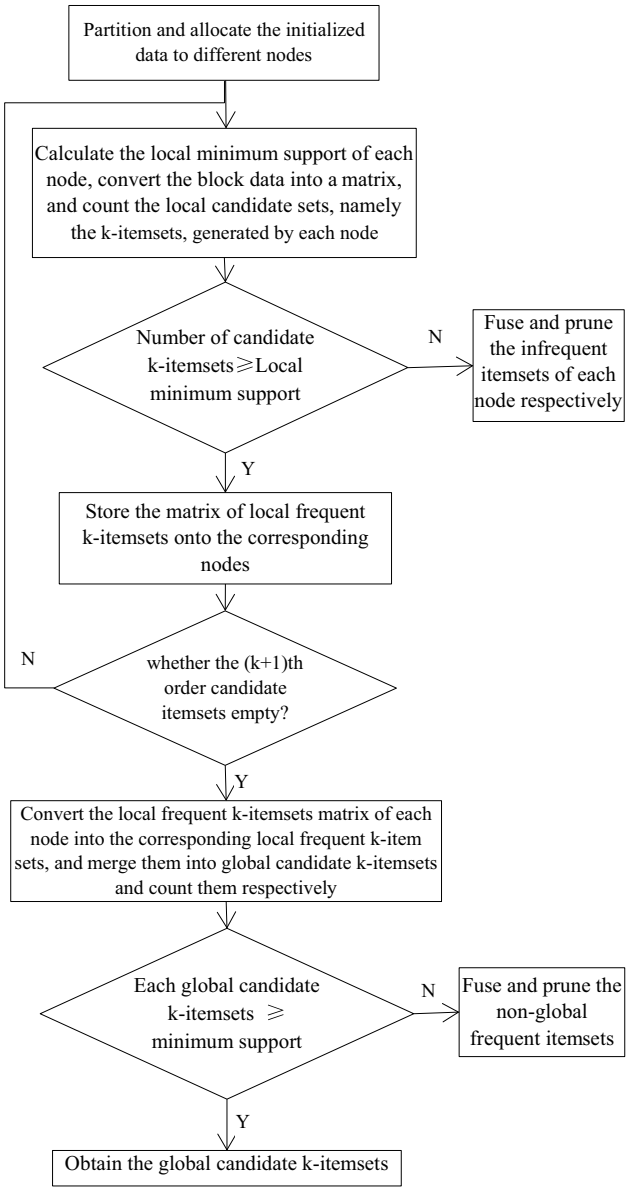


Fig. 3 Algorithm flow of literature [16]

(1) Before getting the final frequent itemsets, the MapReduce function needs to run repeatedly (K iterations). With the increase of iterations, the execution efficiency of the algorithm declines;
(2) When mining association rules, the original dataset needs to be scanned repeatedly. With the continuous augmentation of the original dataset, the mining effect tends to become more and more blurred.

To solve the above problems, this paper makes the following improvement for the parallel association rule mining algorithm. When the MapReduce program is used to mine the potential rules of frequent itemsets, the core scans the original dataset only once; the Map function can be used to obtain all item sets of the dataset including 1 to K items at one time and generate a key value database to store data by key-value pairs, with each key (itemsets) corresponding to a unique value (the number of itemsets supported, uniformly set to 1).

As shown in Fig. 4, by comparison, the improved algorithm flow is found to reduce the number of iterations, and it needs to scan the original dataset and submit the MapReduce task only once, which effectively overcomes the shortcomings of the algorithm proposed in Literature [16].

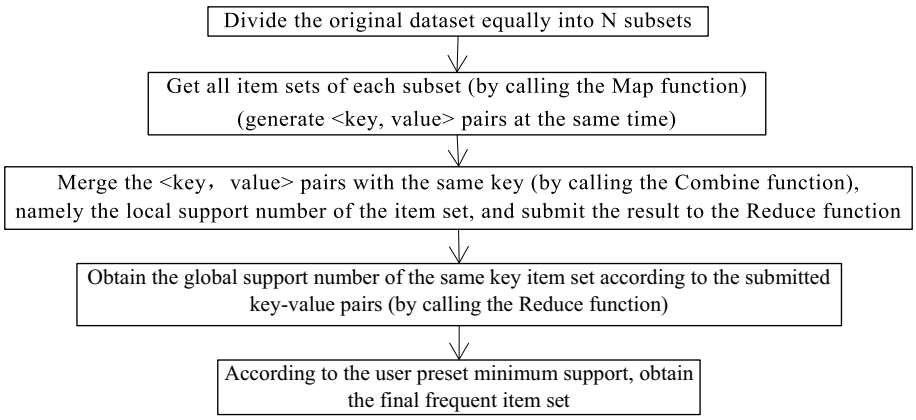
5 Application Case

5.1 Case Background

When the smart campus management system is used to explore the factors affecting students' academic performance, it is required to analyze the correlation between students' campus behavioral data and academic performance. Students' campus behavioral data include:

(1) Average consumption level (ACL). A, B, and C represent low, medium, and high consumption levels, respectively;

Fig. 4 Improved algorithm flow



(2) Breakfast frequency (BF). D, E, and F represent low, medium, and high frequency, respectively;
(3) Average daily online hours (ADOH). G, H, and I represent short, medium, and long hours, respectively;
(4) Canteen meal frequency (CMF). J, K, and L represent low, medium, and high frequency, respectively;
(5) Books borrowed amounts (BBA). M, N, and O represent low, medium, and high amounts, respectively;
(6) Academic performance (AP). *X* and *Y* indicate good and poor performances, respectively.

5.2 Student Behavioral Data Collection and Preprocessing

While collecting the student behavioral data, the data collection system records the specific data of each student on campus. Firstly, the collected data are preprocessed in the format of {FID, ACL, BF, ADOH, CMF, BBA and AP}. FID represents the data serial number. See Table 1 for the preprocessed dataset.

5.3 Data Association Rule Mining and Result Analysis

This paper applies the improved mining algorithm to mine the association rules of the preprocessed student behavioral dataset, setting the minimum support s: 20% and the minimum confidence c: 90%, to obtain all frequent itemsets and filter out the strong association rules based on AP (see Table 2).

By analyzing the filtered strong association rules in Table 2, the following rules can be generated:

Rule 1: Good AP coincides with low ACL of students (with confidence level 96%).

Rule 2: Good AP coincides with high BF of students (with confidence level 95%).

Rule 3: Good AP coincides with low ACL and high BF of students (with confidence level 92%).

By analyzing the above rules, it is found that when studying students' AP, one should pay particular attention to students' ACL and BF, that is, to improve the correlative effect. On the one hand, by analyzing students' ACL, one can understand students' family financial situation and consumption habits, and help students in a well-targeted manner who are ambitious in study but have financial difficulties; on the other hand, one can find out the reasons why students do not eat breakfast by counting the students' BF, so as to foster students' good breakfast habits.

Rule 4: Poor AP coincides with low BBA of students (with confidence level 97%).

Rule 5: Good AP coincides with short ADOH of students (with confidence level 94%).

Rule 6: Good AP coincides with medium BBA and ADOH of students (with confidence level 91%).

By analyzing the above rules, it is found that students' BBA and ADOH also have a certain impact on their AP. On the one hand, universities emphasize students' self-learning ability. In class, the instructor's knowledge is relatively simple given the limited time, and therefore, students need to read a large number of literatures in their spare time after class to improve their understanding of professional knowledge. On the other hand, with the continuous development in information technology, much relevant professional knowledge is available for students to learn on the Internet. Considering the prevalence of a large number of addictive entertainment events such as games and videos on the Internet, it

Table 2 Filtered strong association rules

Records	Rules	Confidence
1	ACL = "A" = = > AP = "X"	96%
2	BF = "E" = = > AP = "X"	95%
3	ACL = "A" ∧ BF = "E" = = > AP = "X"	92%
4	BBA = "N" = = > AP = "Y"	97%
5	ADOH = "I" = = > AP = "X"	94%
6	BBA = "N" ∧ ADOH = "H" = = > AP = "X"	91%
7	CMF = "L" = = > AP = "X"	97%
8	BF = "F" ∧ CMF = "L" = = > AP = "X"	95%
…	…	…

Table 1 Preprocessed student behavioral dataset

FID	ACL	BF	ADOH	CMF	BBA	AP
1	A	F	G	L	O	X
2	B	F	I	L	O	Y
3	A	E	H	K	N	X
4	C	D	H	J	N	Y
5	A	F	G	L	M	X
6	B	E	I	J	O	X
7	C	D	H	K	N	Y
8	B	E	I	K	M	X
…	…	…	…	…	…	…

is necessary to reasonably control students' time online to prevent students against Internet addiction.

Rule 7: Good AP coincides with high CMF of students (with confidence level 97%).

Rule 8: Good AP coincides with high BF and CMF of students (with confidence level 95%).

By analyzing the above rules, it is found that when it comes to BF and CMF, students tend to achieve good AP at high CMF, especially at high BF, with a confidence level of as high as 95%. School managers are suggested to urge students to foster good breakfast habits while improving the dining quality.

5.4 Validation of Association Rules

The Chi-square test is used for multiple classifications of two or more factors to study the correlation and dependency between two variables (presented in the form of a contingency table) [17]. To verify the effectiveness of the association rules, the chi-square value is used to determine the correlation between AP and other factors with the preprocessed results as the sample data. See 3 for the statistical formula.

$$x^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n}x^2[(I-1)(J-1)]. \quad (3)$$

Use MATLAB 7.2. The software calculates the chi-square value (where X_i, $i = 1\ldots 5$; respectively represents ACL, BF, ADOH, CMF and BBA), and further analyzes the relationship between the relevant influencing factors and AP. The results are shown in Table 3.

The results from the Chi-square test showed that students' AP is significantly related to the ACL, BF and CMF, but not with the ADOH and BBA. Generally, modern teaching concepts emphasize the cultivation and development of autonomous learning ability in students. There are many learning resources on the Internet and various auxiliary teaching materials with rich contents, consequently, the ADOH will increase inevitably, and the auxiliary electronic teaching materials are easy to carry and carry out learning. Therefore, students' BBA will be reduced accordingly.

5.5 Algorithm Evaluation

In order to more intuitively test the algorithm's effectiveness, this paper has expanded the data types of students' campus behaviors to increase the size of simulation data. It also takes into full consideration 25 factors including the BF (distinguished as highest, higher, medium, lower, and lowest), ADOH (distinguished as longest, longer, medium, shorter, and shortest), and BBA (distinguished as highest, higher, medium, lower, and lowest) that affect students' AP; at the same time, students' learning styles are distinguished with different campus behavioral data to subdivide their AP into the grades "excellent", "good", "pass" and "fail". Finally, in the processed dataset, 5000, 10,000, 20,000, and 40,000 effective records are randomly selected, respectively, and the improved algorithm proposed in this paper is used to mine the strong association rules between students' campus behavioral data and AP. The mining efficiency is compared with Apriori algorithm and the algorithm in Literature [16], as shown in Fig. 5.

Through the observation and analysis in Fig. 5, the mining efficiencies of the three algorithms are similar for small size of dataset. With the continuous augmentation of the dataset, the mining efficiency of the distributed parallel mining algorithm has been significantly improved compared with the Apriori algorithm, demonstrating the superiority of the distributed parallel technology in processing mass experimental data. On this basis, it can be seen that the improved

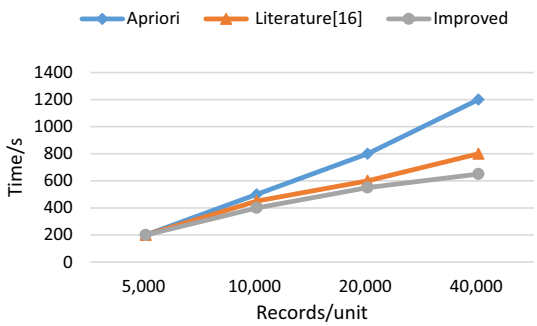


Fig. 5 Comparison in mining efficiency among the algorithms

Table 3 Chi-square test analysis results of sample data

X_i	x^2	Chi-square value comparison	H_0: original assumption	Correlation between AP and X_i
X_1	7.7516	$x^2 = 7.7516 > x^2_{0.05}(2) = 7.38$	Reject	Yes
X_2	11.9781	$x^2 = 11.9781 > x^2_{0.05}(2) = 7.38$	Reject	Yes
X_3	6.0752	$x^2 = 6.0752 < x^2_{0.05}(2) = 7.38$	Accept	No
X_4	9.4302	$x^2 = 9.4302 > x^2_{0.05}(2) = 7.38$	Reject	Yes
X_5	6.3274	$x^2 = 6.3274 < x^2_{0.05}(2) = 7.38$	Accept	No

mining algorithm in this paper is also superior in efficiency to that in the literature [16], lending further support to the feasibility of scanning the original dataset and submitting a MapReduce task once respectively, in improving the mining efficiency of the algorithm.

6 Conclusions

Association rule mining of student behavioral data is an important area of smart campus data analysis. It can expand the methods for smart campus data analysis and deepen the research on student management in the process of smart campus construction. In view of the shortcomings of traditional data analysis methods in the face of mass data analysis, the distributed parallel processing technology has been introduced into student behavioral data association rule mining to construct the framework of distributed student behavioral data association rule mining, the relevant process has been clarified, and the existing mining algorithms have been improved accordingly. The case analysis results show that, the improved mining algorithm has effectively boosted the data mining efficiency. Association rule mining can intuitively reflect the relationship between students' behavioral factors and further analyze the student management knowledge contained in the data, thereby providing an effective basis for campus managers to make sound decisions. In addition, the improved association rule mining algorithm has been applied to different datasets to verify its effectiveness.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by TW, BX and WM. The first draft of the manuscript was written by TW and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Scientific Research Program Funded by Education Department of Shaanxi Provincial Government (Program No. 21JK0579); the Program of Shaanxi Society of Technical and Vocational Education (Program No. 2021SZXYB28); and the Program of Shaanxi Energy Institute (Program No. 20XJZ04).

Availability of Data and Material The data is given in the paper.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to report regarding the present study.

Ethical statement I certify that this manuscript is original and has not been published and will not be submitted elsewhere for publication while being considered by *International Journal of Computational Intelligence Systems*. And the study is not split up into several parts to increase the quantity of submissions and submitted to various journals or to one journal over time. No data have been fabricated or manipulated (including images) to support your conclusions. No data, text, or theories by others are presented as if they were our own. The submission has been received explicitly from all co-authors. And authors whose names appear on the submission have contributed sufficiently to the scientific work and therefore share collective responsibility and accountability for the results.

Consent statement This article does not contain any studies with human participants or animals performed by any of the authors. Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Yu, L.A., Zhang, Y.D.: Weight-selected attribute bagging based on association rules for credit dataset classification. Syst. Eng. Theory Pract. **40**(2), 366–372 (2020)
2. Cao, L., Xu, L., Yang, F., Jia, P.F.: Influencing factors analysis of pavement damage based on mining association rules. Comput. Syst. Appl. **30**(1), 186–193 (2021)
3. Hu, Q.L., Guo, S.: Mining on association rules of atmospheric composite pollutants of urban agglomeration along the Yellow River in Ningxia—a case study of the PM2.5 concentration in the pollution season. J. Ningxia Univ. (Nat. Sci. Ed.) **42**(2), 219–225 (2021)
4. Chen, B.Y., Ding, J., Chen, S.N.: Selection of key incentives for power production safety accidents based on association rule mining. Electr. Power Autom. Equip. **38**(4), 68–74 (2018)
5. Li, X.: Applied research on strong association rules in personalized information push service of smart library. Inf. Sci. **36**(4), 95–99 (2018)
6. Ariannezhad, A., Wu, Y.J.: Large-scale loop detector troubleshooting using clustering and association rule mining. J. Transp. Eng. A Syst. **146**(7), 04020064 (2020)
7. Guo, B., Li, Z.M., Zhang, J., Yu, Z.W.: Cross-modal crowd sourced data for context-based scenic route recommendation. J. Zhengzhou Univ. (Nat. Sci. Ed.) **52**(2), 22–28 (2020)
8. Gao, J.J., Yang, F.: Semi-structured data query optimization algorithm based on swarm intelligence. Comput. Simul. **38**(8), 381–385 (2021)
9. Liang, W.J., Chen, H., Zhao, S.Y., Li, C.P.: A differentially private scheme for top-k frequent itemsets mining over data streams. Chin. J. Comput. **44**(4), 741–760 (2021)
10. Liu, J.Y., Jia, X.Y.: Multi-label classification algorithm based on association rule mining. J. Softw. **28**(11), 2865–2878 (2017)
11. Yu, X., Zeng, F., Mwakapesa, D.S., Nanehkaran, Y.A., Mao, Y.M.: DBWGIE-MR: a density-based clustering algorithm by using the weighted grid and information entropy based on MapReduce. J. Intell. Fuzzy Syst. **40**(6), 10781–10796 (2021)

12. Xu, W., Hoang, V.T.: MapReduce-based improved random forest model for massive educational data processing and classification. Mob. Netw. Appl. **26**(1), 191–199 (2021)
13. Zerabi, S., Meshoul, S., Boucherkha, S.C.: Models for internal clustering validation indexes based on hadoop-MapReduce. Int. J. Distrib. Syst. Technol. **11**(3), 42–67 (2020)
14. Heidari, S., Alborzi, M., Radfar, R., Afsharkazemi, M.A., Ghatari, A.R.: Big data clustering with varied density based on MapReduce. Big Data **6**(1), 1–16 (2019)
15. Wang, Z., Chen, Q., Suo, B., Pan, W., Li, Z.H.: Reducing partition skew on MapReduce: an incremental allocation approach. Front. Comput. Sci. **13**(5), 960–975 (2019)
16. Zhu, K., Huang, R.Z., Zhang, N.N.: Efficient frequent patterns mining algorithm based on MapReduce model. Comput. Sci. **44**(7), 31–37 (2017)
17. Li, M.Z., Ding, Q.X., Wang, Y.P., Lu, T.N.: Performance analysis of management modes of compact disks attached to books in library with chi-square test. Oper. Res. Manag. Sci. **23**(04), 254–257 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.