Construction of Higher Education Management and Student Achievement Evaluation Mechanism Based on Apriori Algorithm

Dongdong Liu
Shangqiu Institute of Technology, Shangqiu 476000, China
Correspondence should be addressed to Dongdong Liu; 1350023002@sggxy.edu.cn
Received 21 May 2022; Revised 23 June 2022; Accepted 24 June 2022; Published 8 August 2022

Academic Editor: Liping Zhang

The Apriori algorithm and DM technology are introduced, examined, and a relationship between bettering teachers’ classroom teaching quality, methods, and means is discussed. An evaluation system for student achievement in higher education is proposed and built on the basis of the Apriori algorithm. The Apriori algorithm is also improved in order to further avoid blind search during mining and increase the effectiveness of frequent itemsets. This paper avoids repeatedly scanning the database and instead reads the entire database into a two-dimensional array, which increases program performance. The algorithm is used to extract useful data from the database using data association mining in the higher education management system. According to the experimental findings, the improved Apriori algorithm’s accuracy can reach 94.81%, which is 10.31% higher than the accuracy of the original Apriori algorithm. The outcomes demonstrate the applicability and reliability of the algorithm model developed in this paper. For managers in teaching and management, it can be a useful reference.

1. Introduction

The core task in college teaching is unquestionably educational administration. It is a crucial and essential link in the management of instruction, and teaching quality is a key component of that chain [1]. The daily management and instruction of colleges and universities have amassed a substantial amount of data, but this data have not been fully utilized [2]. The school’s management of these accomplishments is still in the simple stage of data backup, query, and statistics. At the same time, the school manages students’ grades, and there is a large amount of information about students’ grades that need to be counted. The educational administration system has amassed a wealth of teaching knowledge at this time [3]. The administrators of the academic affairs office have been unable to determine the relationship between courses from students’ grades despite this information having already formed a data warehouse and the growth in enrollment at major universities [4]. The inherent information implied in these data has not been thoroughly examined or thought about in the majority of colleges and universities, either. Most of the data in the management system for educational administration are just stored in databases, and the data processing is limited to simple data insertion, query, and modification, which does not fully utilize information technology. Higher education must immediately address the urgent issue of converting these data into information that can be used to support decision-making for school administrators [5]. According to the argument made in this paper, association rule mining technology can be used to analyze teaching data, identify useful patterns and rules within the data, and provide a foundation for objectively assessing teaching activities and enhancing instruction.

The use and development of DM (Data Mining) technology offer excellent development prospects and solid assistance for resolving issues involving data statistical analysis. The applications of DM technologies in the area of education and teaching are numerous, productive, and difficult. DM is a decision support process that can unearth potential and useful knowledge from a large number of daily accumulated practical application data [6]. Utilizing DM technology to efficiently find the vast educational information already available and give managers helpful
assistance has become more and more crucial. A new data analysis technique called DM is used to extract useful information for human use from the vast amounts of data in the information world [7]. People are becoming more and more interested in extracting useful knowledge from these data in long-term teaching and management as more and more data are amassed. Managers have a basis for decision-making through DM when used in conjunction with the teaching management system currently in use in colleges and universities. Association rules mining in DM technology is the automatic computer mining of numerous associated rules from a large amount of real data. Finding, comprehending, and correctly using these association rules are a crucial part of completing the DM task [8, 9]. Association rules have the advantage of handling variable-length data, supporting indirect DM, and forecasting its calculated consumption [10]. The Apriori algorithm is one of the most important Boolean association rules algorithms for mining frequent item sets. The Apriori algorithm has currently been used successfully in the commercial and financial sectors. This paper develops the mechanism for managing higher education and evaluating student achievement based on the Apriori algorithm. Its innovations are as follows:

1. In this paper, the Apriori algorithm in association rules is introduced to analyze students’ achievement, find the antecedent and successor relationship between courses, and enable administrators to make teaching plans according to market changes and analyze teachers’ teaching achievements, which provides a reference for evaluating teachers’ teaching quality. So as to help the academic affairs office better arrange teaching work and promote the construction of teaching staff.

2. In this paper, the Apriori algorithm in association rules is improved and applied to higher education management and student achievement evaluation. The improved algorithm is used to analyze the teaching evaluation data samples, and the user interaction data records in the database are used to mine frequent itemsets with minimum support and minimum confidence, and valuable data patterns are found from the analysis results. And it can excavate the correlation between courses, and provide guidance for students to choose courses and schools to arrange teaching plans reasonably.

This article will be divided into five sections, with each section’s contents as follows, in accordance with the content it contains and the requirements of the article’s structure: The introduction, which is the first section, primarily explains the context and importance of the subject and presents the research innovation and organizational structure of this paper. The second section is a companion piece. This section examines the management of higher education and the assessment of student achievement in China. And the work and research for this paper are provided. The method section is in the third section. This section analyses the Apriori algorithm and makes it better. And presented a model for higher education management and a system for gauging student achievement; the particular implementation strategy is described in detail. The empirical study is covered in the fourth section. This section conducts an empirical analysis to confirm the evaluation system’s objectivity and logic. Summary and outlook for section 5. This chapter primarily summarizes the key contributions and innovations made during the paper’s final stages and identifies the direction that future research should take.

2. Related Work

Jabbar et al. proposed an improved algorithm based on the AprioriTid algorithm; at the same time, using the improved algorithm, the factors affecting the performance were mined, thus providing a basis for improving the quality of teaching [11]. Wu and Gao proposed an improved Apriori algorithm to solve the problem of the low efficiency of the Apriori algorithm [12]. The algorithm organically integrates the logic and operation of 0-1 coding and mapping, effectively improves the efficiency of the Apriori algorithm, and ensures the accuracy of algorithm results. It is an effective improvement method. Xian et al. found the association rules between courses by analyzing the students’ achievement database; through the analysis of students’ achievements, the relationship between teachers’ educational background, professional titles, and teachers’ teaching effect is extracted, and certain results have been achieved [13]. Aifeng et al. applied DM technology to teaching evaluation. Through the analysis of these data, it was found that the teaching effect of college teachers is related to the age, professional title, education, etc., of teachers [14]. Wu used the association rule algorithm in DM to study the factors that may affect student achievement [15]. Ye analyzes the teaching data samples, finds the existing relationships and rules, plays a guiding role for education and teaching activities, provides reasonable and scientific decision support for teaching management, and puts forward suggestions for further improvement of the system [16]. Pinter et al. pointed out the shortcomings of traditional qualitative analysis, quantitative analysis, and standard deviation methods in teaching evaluation, applied association rule mining in DM to teaching evaluation, and found out the teacher’s classroom teaching effect and teachers’ teaching process and teaching. Association rules between methods and means [17]. Martinez-Garza et al. avoided the specific and complex switching process between memory and external memory and chose to directly rely on the data access system of the database system itself [18]. Although the efficiency is not necessarily the highest, the implementation steps are relatively simple, and better efficiency can be achieved. Reddy et al., based on the study of association rule mining technology, took the achievement of a professional course as a relational database and applied association rules to the analysis of student achievement to mine the correlation between courses [19].
This paper makes an in-depth study of relevant literature, and puts forward and constructs a model of higher education management and student achievement evaluation mechanism. The improved algorithm is used to analyze the teaching evaluation data samples, and the user interaction data records in the database are used to mine frequent itemsets with minimum support and minimum confidence, and valuable data patterns are found from the analysis results. And it can excavate the correlation between courses, and provide guidance for students to choose courses and schools to arrange teaching plans reasonably.

3. Methodology

3.1. The Design of Apriori Algorithm in Higher Education Management. In DM, association rule mining is an important one. It is the role of association rules to acquire and save a large amount of data by various means and to mine valuable information from mountains of data. The most basic algorithm of association rule mining is the Apriori algorithm [20]. Apriori algorithm creatively uses pruning technology based on support, and the system controls the exponential growth of candidate items. This round-robin method uses \( k \) itemsets to generate \( k + 1 \) itemsets. Apriori algorithm for mining association rules includes two basic steps: first, mining frequent itemsets from the database of things, and then generating association rules based on frequent itemsets. In a large database, there are various relationships among its fields, and these relationships are implied in the data contained in the database. Finding these hidden associations from various complex relationships is the purpose of association rule mining. Apriori discovers all frequent itemsets through multiple scans of the database, and only considers all itemsets with the same length in each scan. Finding frequent itemsets layer by layer iteration is the key and core of the Apriori algorithm, which can be accomplished by pruning and connecting itemsets. The whole transaction data set is scanned and compared with the transaction items with the preset minimum support threshold, and the transaction items with support lower than this threshold are deleted, which is called the pruning step.

The key to the high efficiency of the Apriori algorithm is to generate smaller candidate itemsets, that is, to avoid generating and calculating candidate itemsets that cannot be frequent itemsets as much as possible. In order to realize this, this algorithm makes use of such a basic property that any subset of a frequent itemset must be a frequent itemset, and any superset of an infrequent itemset must also be an infrequent itemset. Missing useful information and not finding rules may all be caused by excessive values; finding useless rules, not finding targets, wasting system resources, affecting execution efficiency, etc., may be caused by too small values [21]. The estimation of users’ goals is the key to these influences. In order to improve the efficiency of searching iterative method layer by layer and generating corresponding frequent itemsets, two important properties of the algorithm help to effectively compress the search space of frequent itemsets: (1) Any subset of a frequent itemset is also a frequent itemset. (2) An infrequent itemset and all its supersets are also infrequent itemsets. However, the Apriori algorithm also has some defects. The time and space of the algorithm are mainly consumed in two aspects: (1) the generation of candidate sets. When the frequent itemsets are long or the minimum support is small, the candidate itemsets generated by the join operation are quite large. (2) Scan the database. Repeatedly and massively scanning the database to generate multiple space-consuming candidate sets is time-consuming and space-consuming. The Apriori algorithm needs to scan the database repeatedly, even though several short pattern searches are converted from FP-growth by the way of connecting suffixes, it is still difficult to apply in large-scale databases [22]. With the increase in database size, the time and space performance of the algorithm, especially for mining long patterns, is sharply reduced, and a large number of candidate sets are generated. During the implementation of Apriori, the frequent itemsets are first mined from the things database. Then, strong association rules are generated, which are based on frequent itemsets and measured by the minimum credibility and support, and must meet the predetermined threshold requirements. Finally, the expected rules are generated. Based on the frequent itemsets generated in the above steps, the generated rules only include the items of the set. Based on the concept of Chinese rules, there is only one item on the right side of each rule. Figure 1 shows the process of the Apriori algorithm.

Apriori algorithm consists of two steps, namely, connecting step and deleting step. In this paper, based on the Apriori algorithm, an improved Apriori algorithm is proposed. Its basic idea is: to read the whole database into a two-dimensional array at one time, avoid scanning the database many times, and improve the efficiency of program running. Delete “useless” transactions in time. The minimum credibility and minimum support must be satisfied by these rules, which are generated by frequency sets. The same minimum support degree, its predefined and frequent occurrence of itemsets are at least this, and all frequency sets are found out first, which is the basic idea of the algorithm. Compared with the predetermined threshold range, if the number of itemsets in a leaf node is too large, the leaf node is to be converted into an inner node. The Subset function needs to search all candidates included in a transaction \( t \), in order from root node to leaf. The principle is as follows: if it is an inner node, and the method to reach this node is completed through Hash item \( i \), then Hash should process all items after \( i \) in \( t \); and recursively apply this process to the nodes in bucket. If it is a leaf, it is necessary to determine that itemsets in the leaf are included in the transaction \( t \); let it point to the answer set, and add a reference to ensure that it points correctly. If it is the root node, Hash processes all items in the transaction \( t \). Find out the candidate itemsets by scanning transactions, and generate the candidate itemsets, then count the candidate itemsets that meet the conditions, and remove the candidate itemsets that do not meet the conditions according to the minimum support; The generated frequent itemsets are connected with each other to generate new candidate itemsets. Repeat this cycle \( k \) times until the new candidate item set is empty, and finally determine the frequent \( k \) itemset, which is the maximum frequent itemset. When a
transaction does not contain frequent itemsets of length $k$, it must not contain frequent itemsets of length $k + 1$. Therefore, such transactions can be deleted before generating $k + 1$ frequent itemsets to reduce the time of scanning the array next time. Because the deletion of the array takes a lot of time, the operation of deleting the transaction here does not really delete the array element itself but makes a mark to indicate that the record has been deleted.


The database is kept in the academic affairs office of the school, and the academic affairs staff can use the campus network to connect to it and perform online operations like adding, deleting, updating, and querying. Through the network, the school’s academic affairs office can view the summary data and perform various statistical analyses on it. Users can obtain the necessary information from the academic affairs office using a browser, and some public information may be published on the campus network. In order to organize the entire teaching system in colleges and universities and ensure that students are learning the entire academic year or even the entire university course, many courses must be integrated and connected with one another. As a result, the curriculum for students should be organized in a specific manner, and many courses should be related to one another. System management, teaching plan management,
achievement management, student status management, inquiry, and course selection are all functions that the system is capable of executing. A crucial DM step is identifying mining targets and objects. The mining targets should be predictable even though the mining results are unpredictable. It takes a lot of time and effort to collect the data. Data on students’ basic characteristics and academic progress must be gathered in a variety of ways during the teaching process. The structure design of the higher education management model based on the Apriori algorithm is shown in Figure 2.

In order to ensure the upgrade and replacement of the business system in the future, the data warehouse should be designed to protect the original investment. In addition, any topics added in the future cannot affect the construction of the data warehouse and the application of the original database system, so the business system and data warehouse should be relatively independent. From the flow of the algorithm, it can be seen that the algorithm is based on the database, so we must first access the database, and find the relationship between grades and courses in order to obtain frequent sets in teaching management. In a consistent data storage container, the combination of stored data forms a whole, but it originally exists in multiple data sources. This way is data integration. Data integration can unify the ambiguous, ambiguous, or contradictory places in the original data. According to the needs of the situation, it can also combine two or more attributes into one attribute by operation. Select the collected data and establish the corresponding database. The establishment of database is to provide the data needed for mining association rules, and the data of teachers are stored in the database. By setting the parameters, the mining results of association rules can be adjusted, which has a certain reference for result analysis. The teaching management system uses the data acquisition and preprocessing module to analyze and gather a large amount of data, extract the data that is useful to the system and make some formatting changes so that the data can be used in the system. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I, Y \subset I, X \cap Y = \emptyset$.

The support degree of the rule $X \Rightarrow Y$ in the transaction database $D$ is the ratio of the number of transactions including $X$ and $Y$ to the number of all transactions in the transaction set, denoted as $\text{Support}(X \Rightarrow Y)$:

$$\text{Support}(X \Rightarrow Y) = \frac{|\{T : X \cup Y \subseteq T, T \in D\}|}{|D|}. \quad (2)$$

The confidence of rule $X \Rightarrow Y$ in the transaction set refers to the ratio of the number of transactions containing $X$ and $Y$ to the number of transactions containing $X$, denoted as $\text{Confidence}(X \Rightarrow Y)$:

$$\text{Confidence}(X \Rightarrow Y) = \frac{|\{T : X \cup Y \subseteq T, T \in D\}|}{|\{T : X \subseteq T, T \in D\}|}. \quad (3)$$

In the transaction set, the confidence of rule $A \Rightarrow B$ is $c$, if the probability that $D$ contains both transactions $A$ and $B$ is $c$; this is the conditional probability $P(B \mid A)$, see the following formula:

$$\text{Support}(A \Rightarrow B) = P(A \cup U),$$

$$\text{Confidence}(A \Rightarrow B) = \frac{P(B \mid A)}{P(B)}. \quad (4)$$

The occurrence of itemsets $A$ is independent of the occurrence of itemsets $B$ if

$$P(A \cup B) = P(A)P(B). \quad (5)$$

The appearance of itemset $A$ depends on the appearance of itemset $B$. The correlation between the occurrence of $A$ and $B$ is measured by

$$\text{Corr}_{AB} = \frac{P(B \mid A)}{P(B)} \times \frac{P(A \mid B)}{P(A)} = \frac{P(A \cup B)}{P(B)} \times \frac{P(A \cup B)}{P(A)} \quad (6)$$

If $\text{Corr}_{AB} < 1$, there is a negative correlation; if $\text{Corr}_{AB} = 1$, there is no correlation; if $\text{Corr}_{AB} > 1$, there is a positive correlation. The above formula is equivalent to

$$\text{Corr}_{AB} = \frac{P(B \mid A)}{P(B)} = \frac{P(A \mid B)}{P(A)}. \quad (7)$$

This is also known as “lifting” of association rules.

In this paper, data integration, data selection, and discretization are mainly used for data preprocessing. First of all, the professional course scores of students in the same session of a major are randomly selected from the database of teaching system scores, and stored in the Access database. Secondly, delete the incomplete records in the Access database, and divide the scores into three segments: 0–59, 60–79, and 80–100. Each segment is represented by discrete letters, and each subject is represented by numbers. In the educational administration management system, data storage is usually comprehensive and trivial, and problems such as outliers and null values are inevitable. These useless factors must be removed in the cleaning process to ensure the authenticity of the data analysis. If the data defect affects the normal execution of the data algorithm, it should be dealt with accordingly. If this kind of data loss has little impact on the discovery task, it can be ignored, and it is not necessary to adopt the data loss processing method. The predefined data flow and control flow tasks are included in the toolbox window. Control flow, data flow, event handler, and package explorer are included in the middle view pane. The control flow view provides a design environment in which the use of toolboxes and items related to the control flow can be used to build the control flow. Starting with improving the analysis efficiency, this paper puts forward an improvement scheme to reduce the calculation amount of the Apriori algorithm. It is mainly carried out from two aspects: on the one hand, reducing the times of scanning the database, changing the data storage mode, using a two-dimensional array to store all the data in the database, and deleting useless and redundant information in time to reduce spatial redundancy; On the other hand, based on the Apriori property, the candidate set is reduced, and the amount of calculation is reduced, thus improving the efficiency of the algorithm.
4. Result Analysis and Discussion

The application of educational administration management systems in colleges and universities has greatly reduced the labor force, and reduced a lot of unnecessary human errors so that a large amount of data information can be preserved for a long time and fully utilized. By mining the useful information accumulated in the educational administration management system, the school can adjust the relevant settings of unpopular courses and popular courses and guide and encourage students to choose complementary courses when choosing courses. This is conducive to improving the overall quality of students, but also to the rational distribution of teaching resources. This chapter carries out simulation experiments. In this system, a special data server and application server are used, mainly to ensure the security and stability of the system. At the same time, this paper uses high-performance PHP technology to build dynamic web pages. PHP is an embedded HTML language that mixes C, Java, Perl, and PHP-style new syntax. It can not only run under various operating systems and support almost all databases but also has good system security, stability, and scalability. The test environment is shown in Table 1.

In the test, taking students’ grades as a matter also has its different characteristics. On the surface, we generally think that each student’s academic achievement is an item of affairs. Actually, it is not. In this data set, the items of transactions, which can also be said to be the elements of transactions, are the grades of each student’s grades. For the accuracy of the model, this paper selects two evaluation indexes to comprehensively and systematically evaluate the prediction model. Including MSE (Mean squared error) and MAE (Mean absolute error). The calculation formula of each evaluation index is as follows:

\[ \text{MSE} = \frac{1}{n} \sum_{k=1}^{n} (y_k - \hat{y}_k)^2, \]
\[ \text{MAE} = \frac{1}{n} \sum_{k=1}^{n} |y_k - \hat{y}_k|, \]

where \(y_k\) is the actual value and \(\hat{y}_k\) is the model output value.

Table 1: Test environment.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Set up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware environment</td>
<td>Processor: Dikaryon</td>
</tr>
<tr>
<td></td>
<td>Internal storage: 512 MB</td>
</tr>
<tr>
<td></td>
<td>Hard disc: 200 GB</td>
</tr>
<tr>
<td>Software environment</td>
<td>Operating system: Windows</td>
</tr>
<tr>
<td></td>
<td>Running software: MATLAB</td>
</tr>
</tbody>
</table>

In Figure 2, the structure design of higher education management model based on Apriori algorithm is shown. In Figure 3, the MSE comparison of different algorithms is shown. The MAE comparison of different algorithms is shown in Figure 4.
The original data should be as true and accurate as possible, that is, the quality of the original data directly affects the reference value of DM results for decision support. In the Apriori algorithm, every step is taken, the candidate set of that step is established, and the support degree of each candidate set is counted out, and then it is compared with the predefined minimum support degree, so as to sort out and confirm the maximum item set of this step. The accuracy test results of the algorithm are shown in Figure 5.

During the test, three groups of support and confidence are given to test the algorithm. After the comprehensive results are compared, the result data of support degree 1 and confidence degree 0.5 are finally adopted. A comparison of the evaluation accuracy of the algorithm is shown in Figure 6.

The antecedent or antecedent of the association rules in this paper is the achievement, and then the factors related to the achievement are obtained. In this way, filtering can delete many association rules unrelated to the target, and improve the efficiency of the algorithm. In the process of selecting data sets, the complexity and the requirement of universality in data quantity are also considered. Courses are related to one another in a specific order, so taking one course may affect how well you do in another. It is possible to analyze the correlation between courses and help students choose courses and schools set up lesson plans in a reasonable manner by mining students’ test scores in various subjects from the database of teaching system scores and mining useful rules or relationships hidden behind the data from these massive data. The confidence degree is 0.6 and the support degree is 0.2. The association rules are obtained using the enhanced Apriori algorithm. The resulting partial association rules are shown in Table 2.

Through the abovementioned results, it can be concluded that strengthening the study of data structure is helpful to the study of database principles. Other rules can also be analyzed in this way. According to the mining results, users can easily get relevant information between courses, make decisions, and guide students to choose courses, which
is conducive to students’ better study of various courses. The running time of the algorithm is shown in Figure 7.

It can be seen from Figure 7 that the improved Apriori algorithm greatly reduces the number of database scans. Based on logic and mapping operations, it only needs to scan the whole database twice, which greatly improves the efficiency of the algorithm. After analyzing and studying the Apriori algorithm, this paper further improves it to make it more efficient. The experimental results in this chapter show that the accuracy of the improved Apriori algorithm can reach 94.81%, which is 10.31% higher than that of the traditional Apriori algorithm. The results show that this model can be applied to higher education management and student achievement evaluation mechanism, and the algorithm model constructed in this paper has certain reliability and practical value. The system can be used to classify and analyze relevant data so that these data have new value and practical value. In addition, due to the limitation of experimental conditions, there are still some problems in this study. The next step is to improve the efficiency of the Apriori algorithm, so as to better serve the higher education management.

5. Conclusions

Many fields have successfully used DM technology. Applying DM technology to the teaching system in the field of education can address issues with the preexisting teaching system due to the growth in data. The students’ academic success, to some extent, is a reflection of the teaching quality, which is also the constant theme of the teaching work in colleges and universities. As a result, university teachers and administrators have always been interested in finding ways to raise student achievement and teaching standards. This paper conducts a thorough review of pertinent literature, develops a model of higher education administration, and establishes a system for assessing student achievement. The Apriori algorithm is enhanced to further prevent blind search in the mining process and enhance the effectiveness of frequent itemsets. According to the experimental findings, the improved Apriori algorithm’s accuracy can reach 94.81%, which is 10.31% higher than the accuracy of the original Apriori algorithm. The improved Apriori algorithm greatly reduces the number of database scans. Based on logic and mapping operations, it only needs to scan the whole database twice, which greatly improves the efficiency of the algorithm. The results show that the algorithm model constructed in this paper has certain reliability and practical value. In addition, due to the limitation of experimental conditions, there are still some problems in this study. The next step is to improve the efficiency of the Apriori algorithm, so as to better serve the higher education management.

Data Availability

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

Conflicts of Interest

The author does not have any possible conflicts of interest.

References


