Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 1, pp. 313–326, DOI 10.12694/scpe.v25i1.2297

THE EVALUATION OF ETHNIC COSTUME COURSES BASED ON FP-GROWTH ALGORITHM

RUI XU*

Abstract. In order to make full use of the accumulated curriculum data of Folk costume and dig out useful information from it, so as to provide useful information for curriculum teaching, the article proposes three general functions based on the requirement analysis, and pre-processes the completed grade data of ethnic costume students in 4 academic years, analyzes these data by FP-growth algorithm to understand the situation of association rules between different courses, and through K-means++ algorithm The clustering analysis of students with different levels of achievement is carried out and the results are validated by examples. In the algorithm performance analysis, the performance of FP-growth algorithm is better, the average absolute error of FP-growth algorithm is always smaller than that of Apriori algorithm; When the support degree is 20%, the running time of FP-growth algorithm is 0.4s, which is 0.4s less than that of Apriori algorithm. when the number of calculation nodes is 5, the running time of FP-growth algorithm and the accuracy of the K-means++ algorithm were higher than that of the K-means algorithm. In the Iris dataset, the accuracy of the K-means++ algorithm was 91.05%, which was 8.94% higher than that of the K-means algorithm. When mining the course grade data, the confidence level of the obtained association rules was even higher, even up to 97.15%. The standardized test score for the second group of students was 0.960. The course evaluation method used in the article was more objective and the accuracy of the data analysis was higher, providing valuable reference information for teachers' teaching.

Key words: Ethnic dress; Course evaluation; FP-growth; K-means++ algorithm; Association rules

1. Introduction. The development of technology has not only brought convenience to people's lives, but has also facilitated the development of education. Computer technology is used to improve the quality of teaching and learning, allowing teachers to gain a deeper understanding of their students and to deepen their knowledge with the help of computer technology. Education-related systems have been developed to make the work of education more convenient. And by running the system for a long time, a lot of data is accumulated in the database. This data is not fully utilized. How to make the best use of the data and uncover the latent useful information is the problem that needs to be solved. By applying improved association rules, the running time of the recommender system is reduced and better quality rules are obtained [1]. The efficiency of tyre quality data analysis is improved by mining and analyzing abnormal data through improved Frequent-Pattern growth (FP-growth) algorithm [2]. The K-means algorithm based on principal component analysis was used for clustering analysis of handwritten digital datasets, and the algorithm had better performance [3]. The FP growth algorithm, as an association analysis method, has good performance in frequent itemset data mining. Therefore, facing the problem of data mining of folk costume curriculum, in order to understand the learning situation of students, this paper analyzes the association rules of folk costume curriculum performance data, adopts FP growth algorithm, and uses improved K-means algorithm to cluster the performance data, hoping to mine useful information. The research is divided into four parts. The first part is a literature review, which introduces the research status of domestic and foreign scholars on curriculum teaching, FP growth algorithm, and K-means++algorithm. The second part constructs the teaching evaluation system of folk costume course through FP growth algorithm and K-means++algorithm, and preprocesses the data. The third part analyzes the algorithm performance and application effectiveness. The fourth part summarizes the research methods and points out the research prospects, shortcomings, and future research directions.

2. Related Work. The development of technology has not only brought convenience to people's lives, but has also facilitated the development of education. Computer technology is used to improve the quality of teaching and learning, allowing teachers to gain a deeper understanding of their students and to deepen their

*Academy of Arts Jinling Institute of Technology Nanjing 211169, China (lele1220mao@163.com)

knowledge with the help of computer technology. Education-related systems have been developed to make the work of education more convenient. And by running the system for a long time, a lot of data is accumulated in the database. This data is not fully utilized. How to make the best use of the data and uncover the latent useful information is the problem that needs to be solved. By applying improved association rules, the running time of the recommender system is reduced and better quality rules are obtained [1]. The efficiency of tyre quality data analysis is improved by mining and analyzing abnormal data through improved Frequent-Pattern growth (FP-growth) algorithm [2]. The K-means algorithm based on principal component analysis was used for clustering analysis of handwritten digital datasets, and the algorithm had better performance [3]. The FP growth algorithm, as an association analysis method, has good performance in frequent itemset data mining. Therefore, facing the problem of data mining of folk costume curriculum, in order to understand the learning situation of students, this paper analyzes the association rules of folk costume curriculum performance data, adopts FP growth algorithm, and uses improved K-means algorithm to cluster the performance data, hoping to mine useful information. The research is divided into four parts. The first part is a literature review, which introduces the research status of domestic and foreign scholars on curriculum teaching, FP growth algorithm, and K-means++algorithm. The second part constructs the teaching evaluation system of folk costume course through FP growth algorithm and K-means++algorithm, and preprocesses the data. The third part analyzes the algorithm performance and application effectiveness. The fourth part summarizes the research methods and points out the research prospects, shortcomings, and future research directions.

Liu T et al. optimized the error back propagation (BP) algorithm by improving the particle swarm algorithm and applied the improved algorithm to multimedia courseware evaluation, which has higher prediction accuracy than the algorithm before the improvement [4]. Hideya et al. used topic modeling to collect course evaluation-related data for analysis. After validation, the labels used were found to be valid and able to display the thematic proportions and distribution of course information through visual information [5]. Kazanidis I et.al. applied the overlay algorithm to e-course learning assessment to better assess course quality and facilitate teachers' understanding of teaching and learning [6]. Kazanidis I Bz A et al. analyzed the impact of collaborative learning on students outside the classroom by having them form their own learning groups for a pharmacogenomics course and showed that collaborative learning had a greater impact on students' abilities [7]. Wu X addresses the problem that the "online and offline" teaching mode can improve teaching efficiency, mixes English teaching and civics teaching, and integrates "online and offline" for integrated teaching, analyzes the teaching method of hybrid teaching and proposes reform [8]. The experimental results show that the model has higher recommendation accuracy and lower prediction error [9] on foreign language teaching [10]. Yang Z designed a data mining system based on the Apriori algorithm in order to mine accurate information from the database, and the experiments showed that the system could quickly get effective information, which could effectively improve the quality of education teaching and promote the modernization process of education [11].

Novita R et al. designed a data mining system with the FP-Growth algorithm to determine the relationship pattern of the related topics of the Quran, and the experimental results showed that the system can quickly implement the relationship pattern of the Quran's associated topics and promote the understanding of the Quran [12]. Jiye et al. designed a data mining system with the FP-Growth algorithm in order to improve the detection of irregular behaviors in power management systems, an FP-Growth algorithm detection model was designed, and the experimental results showed that the model can effectively detect behaviors that cause security problems in regular user behaviors and can ensure the security of power association information systems [13]. liu Y et al. In order to improve the dynamic of intelligent rail shuttle trolley scheduling accuracy, a model based on genetic algorithm and K-Means++ algorithm was designed, and the experimental results showed that the model can reduce the system imbalance caused by the CNC system failure and meet the greater production demand [4]. Ye J et al. designed an improved cluster analysis algorithm in order to improve the detection accuracy of UAV FM models by constructing an experimental environment, and the results showed that the algorithm proposed in the study can achieve the detection of UAVs and guarantee a certain accuracy rate, which has good application prospects [15]. Cheng S et al. designed a K-means++ algorithm model in order to improve the performance of neural network models, and evaluated the learning results of the network using the confusion matrix, and the experimental results showed that the model determined a more compact number of nodes in the hidden layer [16]. Zhang G et al., in order to study the impedance adjustment techniques of artificial

Table 2.1: Comparison of Research Contents

Reference	Research Contents
Ref. [4]	Multimedia courseware evaluation
Ref. [5]	Visualization of course information topics
Ref. [6]	Electronic Course Learning Assessment
Ref. [7]	Influence of extracurricular learning cooperation on students of Pharmacogenomics
Ref. [8]	Strategies for Combining Ideological and Political Education with Online and Offline Teaching in English Curriculum
Ref. [9]	Propose a recommendation model for online offline courses
Ref. [10]	Intelligent Online Video Corpus in Teaching
Ref. [11]	University teaching management evaluation combining Big data and data mining
Research	The Application of Data Mining and Clustering Based on Students' Course Score Data in Folk costume Course Evaluation

belt grinding, obtained arm impedance adjustment data based on its relevant data, and obtained the relevant controller through core parameter learning and other processes. During this period, the K-means++algorithm was used to estimate the manual grinding impedance. The results show that the proposed method has a good application effect and can effectively estimate the impedance of manual grinding [17].

To sum up, there is little research on the analysis of students' course scores in the course evaluation, and the data of course scores are not fully utilized. Therefore, the article explores the data of folk costume course scores to understand the students' learning situation. In view of the good performance of FP growth algorithm and K-means++algorithm in data mining, it is applied to the data mining analysis of folk costume courses.

Table 2.1 presents several directions of instructional research. From the data used in the research, there is relatively little literature on conducting research based on student course performance data. Therefore, the unlike cited literature, researching and exploring the relationship between student course performance data and course evaluation can achieve the automation of course evaluation, improve evaluation efficiency and objectivity.

3. Evaluation of Ethnic Costume Courses based on FP-growth Algorithm.

3.1. Course evaluation data pre-processing and correlation analysis. The folk costume course is a study of the design and production of folk costumes. The assessment of the course is carried out to understand the students' learning so that teachers can understand the students' situation and analyze whether they have reached the criteria for graduation. The evaluation is carried out through quantitative and qualitative analysis , which changes the previous method of evaluation which used only grades. In the ethnic dress course evaluation system, a needs analysis is first required to collect course data and grade data. In the first category of data, the course name of the ethnic costume course, the number of students in the relevant course, the basis of assessment and evaluation that each course has and other relevant information need to be collected; in the latter category of data, the students' names, student numbers and other personal information need to be collected, the completion of homework in the usual study, in the examination and practical operation experiments, the performance of students in the data related to students' performance will be collected. Data collection can be done by manual input or batch import. Users can choose different data collection methods according to their own habits, thus enhancing the user experience. The report enquiry and download function allows users to understand how students are learning in the course. Through the data analysis and decision making function, you can understand the students' performance in learning, as well as the achievement of the index points and graduation requirements, and display these data through the information of graphs and charts, so that teachers can make adjustments to the teaching methods based on these visual data, thus improving the quality of teaching. Based on the requirements analysis: three general functions were identified, namely data collection, data analysis, and report management. The use case diagram for data collection and data analysis is shown in Figure 3.1.

Prior to data analysis, data pre-processing is required. The article uses the grade data of students who

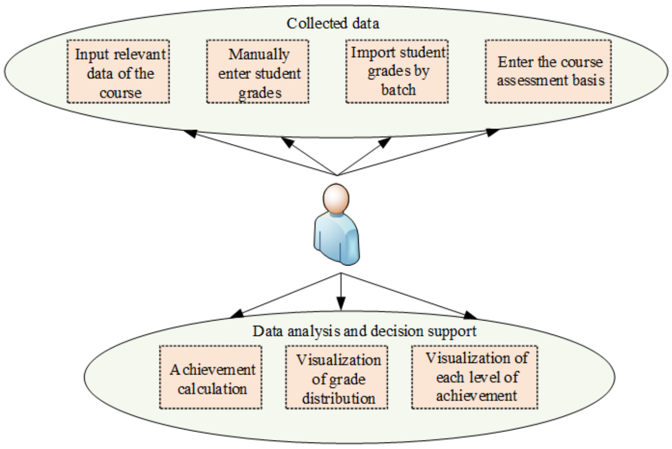


Fig. 3.1: Relevant Use Case Diagrams

completed studies in ethnic dress during the academic year 2016-2020, which is subjected to data cleaning and integration processes, followed by data transformation so that students' exam grades are standardized to lie between [0,1], and finally data statute processing, in which discrete processing results in greater than the course mean data correlation and clustering analysis was then performed, in which the association rule algorithm was applied. In the association rules, the set of items is set as shown in equation 3.1.

$$I = \{i_1, i_2....., i_n\} \tag{3.1}$$

In equation 3.1, represents the term and then sets the transaction term , whose mathematical expression is shown in equation 3.2.

$$D = \{T_1, T_2...., T_p\} \tag{3.2}$$

In equation 3.2,T_p represents a transaction; T_p is the set of i_n and $T \epsilon I$. If $X \epsilon I, X \epsilon \neq \theta, Y \epsilon I, Y \neq \theta$ and $X \cap Y \neq \theta$, there are association rules in the non-empty set X and Y consisting of certain items, and , are the preconditions, outcomes of the relationship. The mathematical expression for the support of this relation is shown in equation (3).

$$sup(X => Y) = \frac{|\{T : X \cup Y \epsilon T, T \epsilon D\}|}{|D|} \tag{3.3}$$

The mathematical expression for the confidence level of the association rule is shown in equation 3.4.

$$Confidence(X => Y) = \frac{|\{T : XUY \epsilon T, T \epsilon D\}|}{|\{T : X \epsilon T, T \epsilon D\}|} \tag{3.4}$$

In equation 3.4, the confidence level represents the probability of X and Y occurring at the same time, and the higher the confidence level, the more reliable the corresponding association rule is. In the case of the item set support $sup(X)$, the formula is shown in equation 3.5.

$$sup(X) = \frac{|\{T : X \epsilon T, T \epsilon D\}|}{|D|} \tag{3.5}$$

In equation 3.5, if $sup(X) \geq min_s up$,$min_s up$ represents the minimum support, then represents the frequent itemset. Since the Apriori algorithm has a high computational overhead and is not conducive to data analysis,

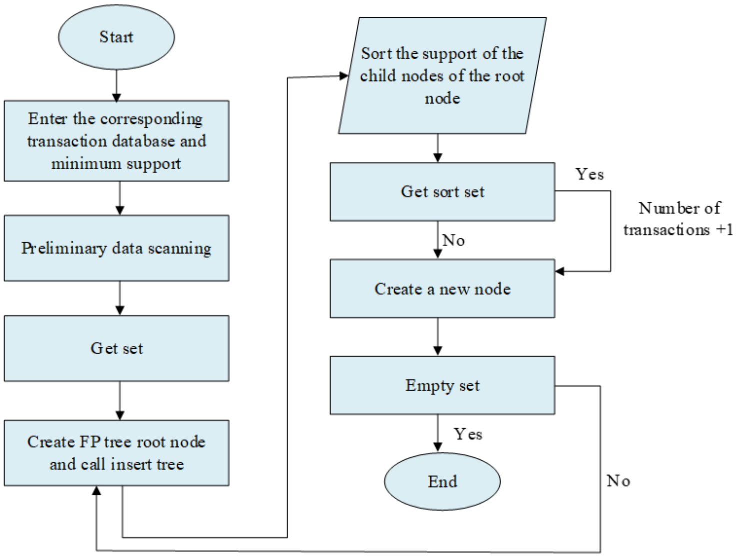


Fig. 3.2: Relevant Use Case Diagrams

the article uses the FP growth algorithm, which mines the entire set of frequent items without generating a candidate set of items. This algorithm can overcome the flaw of multiple scans of transactional databases in the Apriori algorithm. It is worth mentioning that the data processed by the algorithm cannot include subjects that have little relevance to the folk costume course, which will affect the accuracy of the results. Therefore, certain processing of the data is necessary. At runtime, this algorithm first constructs a Frequent Pattern Tree (FP Tree), as shown in Figure 3.2.

In the figure 3.2, the scan of D, which is known through several experiments, and $minsup = 50$, with a minimum confidence level of 0.85, are used to calculate the set of frequent terms, as shown in equation 3.6.

$$F = \{j_1, j_2, ..., j_m\} \tag{3.6}$$

In equation 3.6, F represents the set of frequent items and j_n represents the frequent items. The support number of frequent items, which represents the number of times the item appears in D , is calculated and sorted in descending order to obtain the frequent items table L . The FP-Tree root node is then created and marked by null. It performs a second scan of D and in T and removes the non-frequent items to obtain the frequent items, which are sorted according to L . It creates a branch of T to represent the nodes of the FP-Tree by item herding dust and supports number, where T_n indicates that there is a n transaction T . When linking T_1 frequent items, the root node links the first frequent set, and then links the frequent items in the same way that the latter frequent necklace is linked to the previous frequent item. When a shared path exists in the T_2 branch, the number of all nodes in which the path of that segment is located is added by one and the remaining different paths need to be recreated. When all T branches are inserted, the FP-Tree is obtained. Then frequent item mining is performed, as shown in Figure 3.3.

The association rules are mined by calling FP-growth from FP-Tree. In the recursive process, FP-growth calls the first layer of the FP-Tree with a null value for the node, thus obtaining the frequent 1-item set. Then, the recursive call to FP-growth is made on all the resulting itemsets, resulting in a multivariate frequent itemset. As can be seen from the algorithm's operation, the algorithm is divided into two main parts, building the FP-Tree to mine the frequent itemsets. The FP-growth algorithm is then used to analyze the course performance data of ethnic dress students and to mine the association rules that exist in them. According to the association rules mined, teachers can reasonably formulate the professional training plan, so that the teaching quality of this major can be improved.

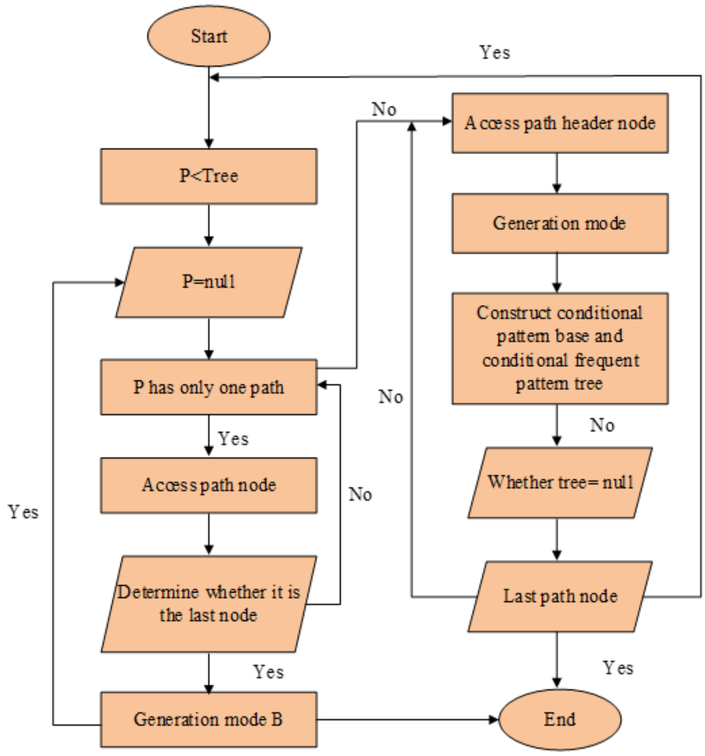


Fig. 3.3: Operation process

3.2. Classification of course performance data based on cluster analysis. The K-means algorithm is also a clustering algorithm that is generally suitable for large-scale data mining, but the number of clusters can affect the results of the clustering analysis. The K-means++ algorithm can make up for the shortcomings of this algorithm, and the flow of the K-means++ algorithm is shown in Figure 3.4.

In Figure 3.4, the first cluster centroid c_1 is randomly selected from the set C and the shortest distance from each point in C to the cluster $d(x_i)$ is calculated using the Euclidean distance formulax_i. represents the first vector$i$ in C . The sum-of-squares operation is performed dx_ion to obtain the sum-of-squares value of x_i $sum(d^2_(x_i))$. The probability of the midpoint of C becoming the next cluster centre is calculated and the relevant formula is shown in equation 3.7.

$$P_i = \frac{d^2(x_i)}{sum(d^2(x_i))} \tag{3.7}$$

In equation 3.7,$P(i)$ represents the probability of being the next cluster centre. It chooses a random number R which lies in the interval [0,1], lets R subtract $P(1)$; When $R - P(1) > 0$, it continues to subtract $P(2)$, before again determining the magnitude of the difference between R and $P(2)$, when it is greater than 0, use the same method, do the subtraction operation until the obtained difference is not greater than 0, stop down, the second clustering centre is the point corresponding to $P(i)$. Starting with the operation to calculate the shortest distance $d(x_i)$, the subsequent steps are repeated to find the optimal K initial clustering centre ck . The distance between the points of C and ck is then calculated, and the nearest ck point to the set C is found and then grouped into the corresponding subset. The mean value of the set is calculated and the calculated mean value is used to update the clustering centres and to perform the error averaging and calculation, the

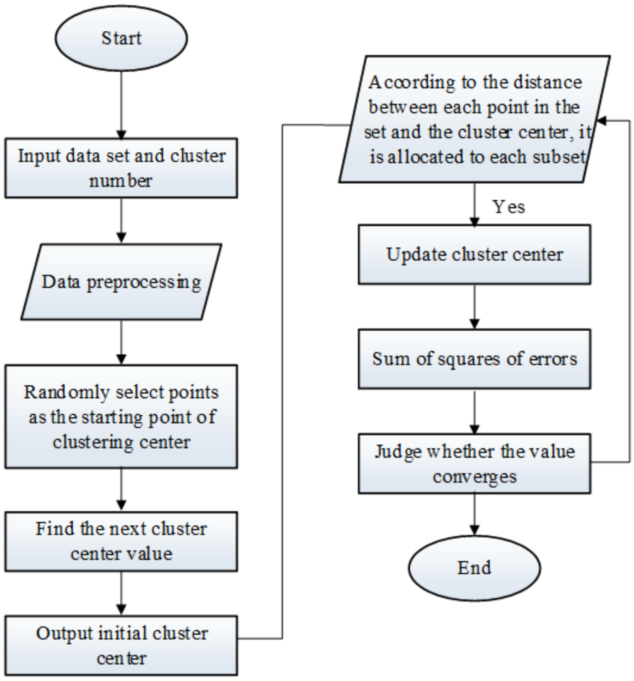


Fig. 3.4: Operation process

corresponding formula is shown in equation 3.8. mi denotes the mean value of C.

$$E = \sum_{i=1}^{k} \sum_{p \epsilon C} |q - mi|^2 \tag{3.8}$$

In equation 3.8,E represents the error sum of squares and the set of points is set to q . This step is repeated from calculating the distance between the points at C and ck and stops when the error sum of squares converges. Applying the K-means++ algorithm to the analysis of student performance data, the numerical class of performance data needs to be discretized before use. After this type of processing, the type of student performance data is transformed into a binary data type, the specific mathematical expression of which is shown in equation 3.9 .

$$\begin{cases} i = (x_{i1}, x_{i2}, ..., x_{ip}) \\ j = (x_{j1}, x_{j2}, ..., x_{jp}) \end{cases} \tag{3.9}$$

In equation 3.9, i and j are the student's performance vectors and the number of attributes of the values is p . The Euclidean distance d(i,j) is one of the commonly used distance metrics and is shown in equation 3.10.

$$d_{ij} = \sqrt{\sum_{k}^{n} (x_{ik} - x_{jk})^2} \tag{3.10}$$

In equation 3.10, d_{ij} represents the Euclidean distance. The Manhattan distance is also a distance metric and its mathematical expression is shown in equation 3.11.

$$d_{ij} = \sum_{k=1}^{n} |x_{ik} - x_{jk}| \tag{3.11}$$

In addition to this, there is the Minkowski distance method, which is calculated using the formula shown in equation 3.12.

$$d(i,j) = \left(\sum_{k=1}^{n} |x_{ik} - x_{jk}|^p \right)^{\frac{1}{p}} \tag{3.12}$$

After discretizing the data related to test scores, this data exists in two states, when the data is 0, it means that the student's score is not greater than the average score; If it is 1, it is greater than the average score. With the help of a binary data column table can reflect the dissimilarity of these two data, while for $d(i,j)$ symmetric binary data, the symmetric dissimilarity $d(i,j)$ is calculated as shown in equation 3.13.

$$d(i,j) = (b+c)/(a+b+c) \tag{3.13}$$

In Equation 3.13, when the values of j and i are both 1 or 0, the corresponding state numbers are a or$d$ respectively; When $i = 0$,$j = 1$ the state number is c ; For asymmetric binary data, the two data states of 0 or 1 are of different importance, and since $i = 1, j = 1$ is more significant than , $i = 0, j = 0$ d is omitted, resulting in the corresponding mathematical expression for the phase difference as shown in equation 3.14.

$$d(i,j) = (b+c)/(a+b+c+d) \tag{3.14}$$

The K-means++ algorithm was then applied to cluster the results of student performance in a course and the target number of clusters was determined $k = 4$. In order to test the performance of the algorithm, an acceleration ratio can be used, which is calculated as shown in equation 3.15.

$$s_m = T_1/T_m \tag{3.15}$$

In equation 3.15, s_m represents the speedup ratio, and the Central Processing Unit (CPU) time to complete the task for m and 1 processor are T_m and T_1 respectively.

4. Analysis of Experimental Results. In order to study the performance of the FP-growth algorithm, the article uses the Apriori algorithm as the comparison algorithm, the number of transactions included in the test data is 8428, by mining test data to reflect the performance of the two algorithms; The CPU used is 2GHZ, to study the average absolute error of the algorithm with different number of iterations, and the running time of the two algorithms with different support, the specific results are shown in Figure 4.1.

As can be seen in Figure 4.1, the mean absolute error of both the FP-growth algorithm and the Apriori algorithm decreases as the number of iterations increases. At 700 iterations, the average absolute error of the FP-growth algorithm is 0.665, while the average absolute error of the Apriori algorithm is 0.725, with the average absolute error of the former algorithm being 0.06 smaller than the average absolute error of the latter. Overall, the average absolute error of the FP-growth algorithm is always smaller than that of the Apriori algorithm. When the support degree was 10%, the running time of the FP-growth algorithm was 1.3s, while the running time of Apriori algorithm was 6.5s. The time difference between the latter algorithm and the former algorithm was 5.2s. With the increase of support degree, the running time of both algorithms decreased continuously, and when the support degree was before 16%, the running time of the Apriori algorithm decreased, the running time of the Apriori algorithm decreased faster than that of the FP-growth algorithm. At 20% support, the FP-growth algorithm runs in 0.4s, which is 0.4s less than the Apriori algorithm, and at 30% and 40% support, the FP-growth algorithm runs in 0.3s and 0.2s respectively. Overall, the FP growth algorithm always has a shorter runtime than the Apriori algorithm at the same level of support. It can be seen that the FP-growth algorithm has a better performance compared to the Apriori algorithm. The performance of the FP growth algorithm was further analysed with the test dataset of webdocs, and study the running time of FP growth algorithm, the Apriori algorithm, the decision tree and the random forest under different numbers of computing nodes, and the running time of the four algorithms under different data set sizes, as shown in Figure 4.2.

In Figure 4.2, the running time of the FP growth algorithm and Apriori algorithm varies with the number of computing nodes. With the increase of the number of computing nodes, the running time of the two algorithms

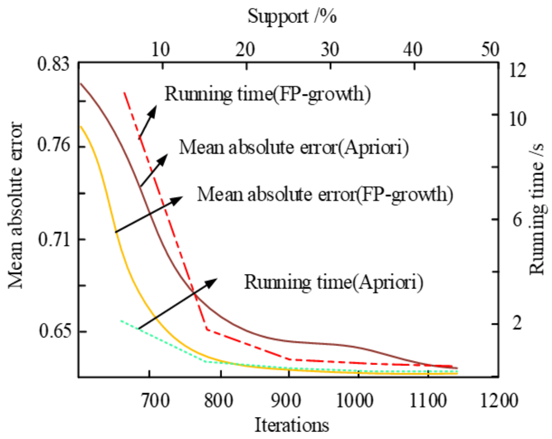
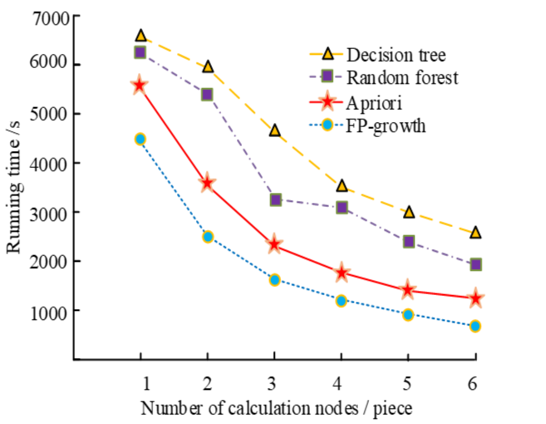
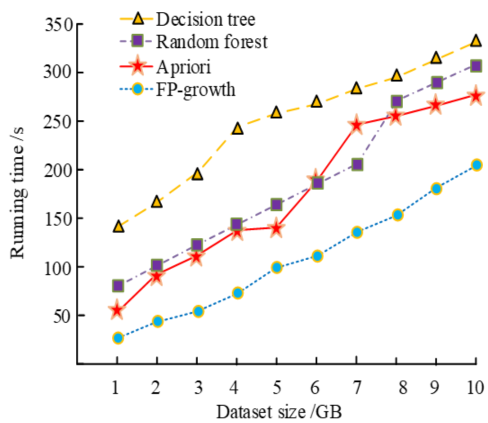


Fig. 4.1: Mean Absolute Error and Running Time of the Two Algorithms



(a) Calculate the running time under the node



(b) Run time at dataset size

Fig. 4.2: Running time of four algorithms

changes, and both are constantly decreasing. When the number of computing nodes is 1, the running time of Apriori algorithm is 5534s, while that of FP growth algorithm is 4476s. When the number of computing nodes is 2, the running time of FP growth algorithm is 2602s, 1082s less than that of the Apriori algorithm. When the number of computing nodes is less than 3, the running time of the two algorithms decreases rapidly; When the number of calculation nodes is greater than 3, the running time of the FP growth algorithm and the Apriori algorithm decreases slowly. When the number of calculation nodes is 4, the corresponding running time of the FP growth algorithm is 1235s, while the running times of the Apriori algorithm is 1727s, which is 492s more than the former. When the number of computing nodes is 5, the running time of the FP growth algorithm and Apriori algorithm is 1000s and 1451s respectively, and the time difference between them is 451s. From the time difference between the two algorithms when the number of computing nodes is 4 and 5, it demonstrates that the time difference between the two algorithms is decreasing. As the data set increases, the running time of the four algorithms increases, and the running time of the FP growth algorithm is relatively minimum. It tests the data sets webdoc1 and webdoc4, and studies the running time and acceleration ratio of the FP growth

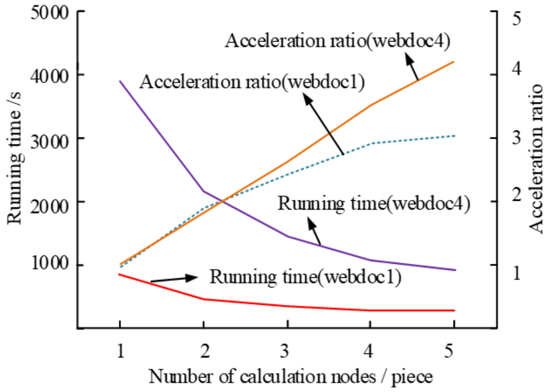


Fig. 4.3: Number of calculation nodes /piece

algorithm under different data sets and different number of computing nodes, as shown in Figure 4.3.

In Figure 4.3, there are differences in the runtime of the FP growth algorithm under the same number of computing nodes in different datasets. As the number of computing nodes increases, the corresponding runtime decreases, and the time difference for mining different datasets becomes smaller and smaller. Overall, the FP growth algorithm has the least runtime when mining dataset webdoc1. When the number of nodes is calculated as 1, the FP growth algorithm takes 854 seconds to mine dataset webdoc1, while it takes 3817 seconds to mine dataset webdoc4. The FP growth algorithm takes much less time to mine dataset webdoc1. When calculating the number of nodes to 2, the FP growth algorithm's mining time for the dataset webdoc4 rapidly decreases, with a running time of 2143 seconds, which is 1732 seconds longer than the algorithm's mining time for dataset webdoc1, while the algorithm's mining time for dataset webdoc1 is 411 seconds. When mining the dataset webdoc1, the FP growth algorithm ran for 264 seconds and 253 seconds when calculating the number of nodes as 4 and 5, respectively. The time difference between the two is relatively small, with a time difference of 11 seconds. In the acceleration ratio of the FP growth algorithm, when the number of computing nodes is not greater than 2 in different datasets mining, the acceleration ratio of the FP growth algorithm in datasets webdoc1 and webdoc4 mining is basically the same. When the number of computing nodes is not less than 3, as the number of computing nodes increases, the acceleration ratio of the algorithm running on the two datasets continues to increase, and the difference in acceleration ratios between the two continues to widen. When the number of computing nodes is 5, the FP growth algorithm has an acceleration ratio of 3.1 on dataset webdoc1, while the algorithm has the acceleration ratio of 4.3 on dataset webdoc4. The former has an acceleration ratio 1.2 less than the latter. Overall, from the performance analysis of the FP growth algorithm, it can be seen that its performance is significantly better than the Apriori algorithm. It studies the accuracy and runtime of the K-means++algorithm and the K-means algorithm under different datasets, as shown in Figure 4.4.

In Figure 4.4, in the accuracy of the algorithms, the accuracy of the different algorithms varied across the datasets, and overall the accuracy of the K-means++ algorithm was higher than the accuracy of the K-means algorithm. In the Iris dataset, the accuracy of the K-means++ algorithm is 91.05%, which is 8.94% higher than that of the K-means algorithm, while the accuracy of the K-means algorithm is 82.11%. The accuracy of the K-means++ algorithm in the datasets Glass and Flame is 83.00% and 95.07% respectively. Overall it can be seen that the two algorithms have the highest accuracy in the dataset Wine, with the K-means++ algorithm having an accuracy of 96.21% and the K-means algorithm having an accuracy of 90.00%. In the running time of the algorithms, the K-means algorithm had the longest running time in the dataset Wine with a time of 6.49s, 1.67s more than the K-means++ algorithm, while the K-means++ algorithm had a running time of 4.82s. In the dataset Flame, the K-means++ algorithm had a running time of 3.12s, 1.67s less than the K-means. The running times of the K-means++ algorithm were 3.32s and 4.01s in the datasets Iris and Glass respectively. It can be seen that the running time of the K-means++ algorithm was less than that of the K-means algorithm;

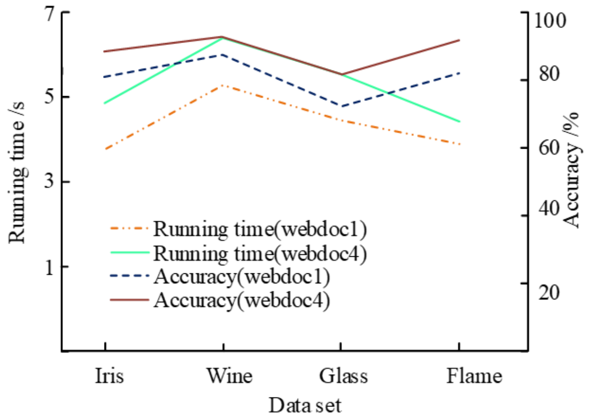


Fig. 4.4: Accuracy and runing time under different data sets

Table 4.1: Association Rules of Some Courses of National Costume Major

Precondition	Result	Confidence Level (%)
National costume culture, northern national folk culture	Appreciation of national costumes	94.47
Foundation of creation, sketch, colour, photography	Art design practice	97.15
Foreign language, computer-aided design	English for art design	89.16

both algorithms had the most running time in the dataset Wine. In addition, through experiments, in the data set Iris, compared with the K-means algorithm, the K-means++algorithm has a smaller Mean absolute error and a larger recall rate. It can be seen that the K-means++ algorithm performs better than the K-means algorithm. The FP-growth algorithm was used to mine the course grade data and the relevant results were obtained as shown in Table 4.1.

As can be seen from Table 4.1, the confidence levels of the association rules vary from course to course. Based on the relationship between the confidence levels, it is known that with the two courses, Ethnic Costume Culture and Northern Ethnic Folk Culture, as prerequisites, the confidence level between them and the course on Ethnic Costume Appreciation as an outcome is 94.47%, indicating that the inference between them is more reliable. Similarly, it can be seen that inferences from the other two course association rules are also more reliable. With the prerequisites of Fundamentals of Creative Writing, Drawing, Colour and Photography, the confidence level of the association rule with the Art and Design Practice course as the outcome was 97.15%. The usual and examination results of the eight students were processed to obtain the corresponding dataset, and sample A6 was used as the initial clustering centre, with the relevant data shown in Table 4.2.

From Table 4.2, it can be seen that there is a difference between the usual and examination results of these students. is the distance from the sample to sample A6, from which the cumulative values of and are known $P(x)$,$P(x)$ as shown in Figure 4.5.

As can be seen in Figure 4.5, the probability of a sample becoming the next cluster centre differs from sample to sample. The probability of sample A1 becoming the next cluster centre is 20.00% and the cumulative value of this probability is also 20.00%; the probability of sample A2 becoming the next cluster centre is greater than that of sample A1 with a probability of 32.50% and the corresponding cumulative value is 52.5%. Among the different samples, sample A6 has the smallest probability of being the next cluster centre of 0, while sample A7 has a probability of being the next cluster centre of 5.00%, which is 2.50% greater than that of sample A5. The probability and cumulative probability of sample A3 being the next cluster centre are 12.50% and 65.00% respectively. The cumulative value of for the A8 sample is 100.00%. The cumulative value of for the first 4

Table 4.2: Association Rules of Some Courses of National Costume Major

Sample serial number	A1	A2	A3	A4
Usual performance	3	4	3	4
Examination results	4	4	3	3
$d(x)^2$	8	13	5	10

Sample serial number	A5	A6	A7	A8
Usual performance	0	1	0	1
Examination results	2	2	1	1
$d(x)^2$	1	0	2	1

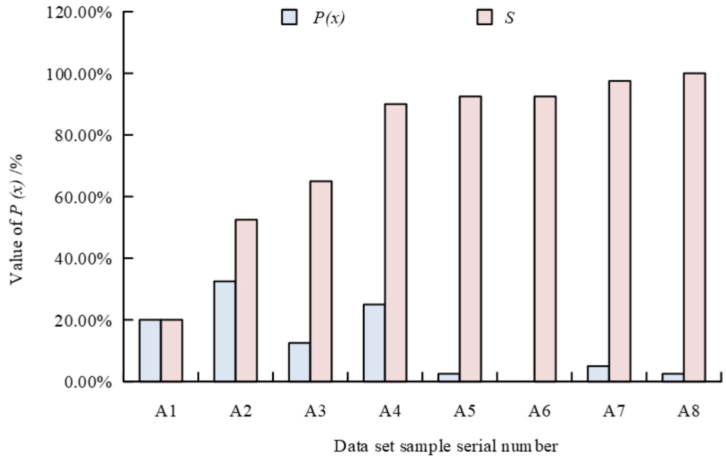


Fig. 4.5: Data set sample serial number

samples was 90.00% and it just so happened that the first 4 samples were farther away from A6 and had a higher compared to the other sample points, validating the theory of the algorithm involved in the study to find different types of students at different test difficulty levels. Using the K-means++ algorithm to cluster discrete art and design practice courses, the relevant results were obtained as shown in Figure 4.6.

In Figure 4.6, students are divided into four categories, with standardized test scores and usual scores for each category. Comparing the standardized scores, it can be seen that students in the first category had good scores in both categories, with a standardized test score of 0.986, while students in the second category had a standardized usual score of 0.715, which was 0.277 higher than the standardized usual score of students in the third category. It can be seen that the application effect of the research method is good.

In the course evaluation of folk costume, when mining the association rules between students' scores, the research did not choose the traditional Apriori algorithm, but carried out the related mining work through FP growth algorithm. By using this algorithm, good mining results were achieved. Some face the problem of frequent pattern mining in large databases and propose Apriori algorithm and FP growth algorithm. After testing the open-source data of Istacart, it was found that compared to the Apriori algorithm, the FP growth algorithm runs faster and has better application effects [20]. It can be seen that selecting an appropriate association rule analysis algorithm is beneficial for improving mining efficiency. The choice of clustering algorithm will affect the clustering effect. Some scholars, in order to improve the effectiveness of hot spot mining for taxi passengers, use K-means++algorithm to carry out relevant clustering work based on the concept of secondary segmentation. By comparing this algorithm with methods such as the K-means algorithm, it was found that

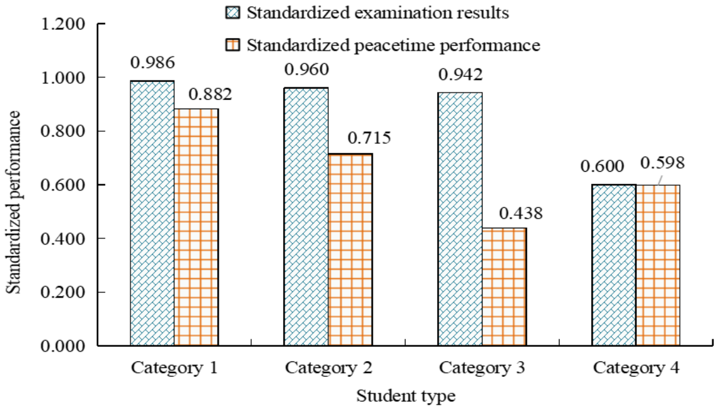


Fig. 4.6: Standardized grades of each category of students

the K-means++algorithm has better clustering accuracy. From this, it can be seen that selecting a suitable clustering algorithm based on the actual situation is beneficial for achieving better clustering results. In the course evaluation of Folk costume, when conducting association rule analysis and cluster analysis on students' course performance data, selecting appropriate algorithms is conducive to improving the effect of folk costume course evaluation, so as to help teachers make better teaching decisions.

5. Conclusion. In order to explore the correlation between ethnic costume courses and understand the learning situation of different types of students, the article uses the FP-growth algorithm to mine the results of courses in ethnic costume, find the hidden useful information in them, and realize the cluster analysis of students with different learning results through the K-means++ algorithm, so that teachers can adjust the learning situation of different types of students according to This allows teachers to adjust their teaching methods according to the learning situation of different types of students, thus realizing personalized teaching. The performance analysis of the algorithm shows that the FP-growth algorithm outperforms the Apriori algorithm, with a running time of 0.4s when the support degree is 20%, which is 0.4s less than the Apriori algorithm. The FP-growth algorithm took the least amount of time to mine the webdoc1 dataset among the different datasets. In the accuracy of the algorithms, the accuracy of different algorithms varied across datasets, with the K-means++ algorithm having a higher accuracy than the K-means algorithm overall. The confidence level of the association rule with the art and design practice course as an outcome was 97.15% with the prerequisites of creative fundamentals, drawing, color and photography. The cumulative value of for the first four samples was 90.00%, while it just so happened that the first four samples were further away from A6 and had higher compared to the other sample points. The first group of students had good scores in both categories, with a standardized test score of 0.986, while the second group of students had a standardized usual score of 0.715. The application effect of the research method is good, and it can be applied to other course evaluations, which is conducive to improving teaching quality. There are still some shortcomings in the research. In the data, there are fewer types of data for student grades, and more types of data should be added to make the analysis more comprehensive. In the data mining analysis, the amount of data involved in mining is relatively small, and it is necessary to increase the amount of data analyzed to further demonstrate the feasibility of the research method. In the future, research can be conducted to add data types and data volumes to improve the applicability of the method.

Funding. The research is supported by: High-level Scientific Research Initiation Project of Jinling Institute of Technology "Study on the Motif of Ethnic Minority Dress Decoration in Northwest China under the Background of One Belt and One Road, Project No. jit-b-201920; General Research Project on Philosophy and Social Sciences in Jiangsu University "A Comparative Study of Chinese Kazakh Clothing on the Silk Road and Central Asian Kazakh Clothing" Project No.2020SJA0533; and National first-class disciplines program funding.

REFERENCES

[1] Hamidi, H. & Hashemzadeh, E. An Approach to Improve Generation of Association Rules in Order to Be Used in Recommenders. *International Journal Of Data Warehousing & Mining.* **13**, 1-18 (2017)

[2] Li, M., Ding, D. & Yi, Y. Data Analysis of Tyre Quality Based on Improved FP-Growth Algorithm. *Zhongguo Jixie Gongcheng/China Mechanical Engineering.* **30**, 244-251 (2019)

[3] Li, B. An Experiment of K-Means Initialization Strategies on Handwritten Digits Dataset. *Intelligent Information Management.* **10**, 43-48 (2018)

[4] Liu, T. & Yin, S. An improved particle swarm optimization algorithm used for BP neural network and multimedia course-ware evaluation. *Multimedia Tools & Applications.* **76**, 11961-11974 (2017)

[5] Hideya, M. Makiko, et al. Analysis of the free descriptions obtained through course evaluation questionnaires using topic modeling. *Educational Technology Research.* **41**, 125-137 (2019)

[6] Kazanidis, I., Valsamidis, S., Gounopoulos, E. & Others Proposed S-Algo+data mining algorithm for web platforms course content and usage evaluation. *Soft Computing.* **24**, 14861-14883 (2020)

[7] Bz, A., Sk, A., Yx, B. & Others A student initiative to implement peer-led study groups for a pharmacogenomics course: Evaluation of student performance and perceptions. *Currents In Pharmacy Teaching And Learning.* **12**, 549-557 (2020)

[8] Wu, X. Research on the Reform of Ideological and Political Teaching Evaluation Method of College English Course Based on "Online and Offline" Teaching. *Journal Of Higher Education Research.* **3**, 87-90 (2022)

[9] Yz, A., Hao, L., Ping, Q. & Others Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization - ScienceDirect[J]. *Neurocomputing.* **415**, 84-95 (2020)

[10] Liu, Y., Han, L., Bo, J. & Others The application and teaching evaluation of Japanese films and TV series corpus in JFL classroom. *The Electronic Library.* **36**, 721-732 (2018)

[11] Zhao, Y. Research on the application of university teaching management evaluation system based on Apriori algorithm. *Journal Of Physics: Conference Series.* **1883**, 012033 (2021)

[12] Novita, R. & Mustakim, S. Determination of the relationship pattern of association topic on Al-Qur'an using FP-Growth Algorithms. *IOP Conference Series: Materials Science And Engineering.* **1088**, 12020-01202 (2021)

[13] Jiye, W. Zhihua, et al. *FP-Growth Based Regular Behaviors Auditing In Electric Management Information System - ScienceDirect. Procedia Computer Science.* **139**, 275-279 (2018)

[14] Liu, Y. & Ling, P. Intelligent RGV Dynamic Scheduling Strategy Based on Greedy Algorithm. *World Scientific Research Journal.* **5**, 278-287 (2019)

[15] Ye, J., Zou, J., Gao, J. & Others A New Frequency Hopping Signal Detection of Civil UAV Based on Improved K-Means Clustering Algorithm. *IEEE Access PP (.* **99** pp. 1-1 (2021)

[16] Cheng, S., Wu, Y., Li, Y. & Others TWD-SFNN: Three-way decisions with a single hidden layer feedforward neural network. *Information Sciences.* **579**, 15-32 (2021)

[17] Zhang, G., Zhang, G., Ni, F. & Others Learning impedance regulation skills for robot belt grinding from human demonstrations. *Assembly Automation.* **41**, 431-440 (2021)

[18] Alcan, D., Ozdemir, K., Ozkan, B., Mucan, A. & Ozcan, T. A comparative analysis of apriori and fp-growth algorithms for market basket analysis using multi-level association rule mining. *Global Joint Conference On Industrial Engineering And Its Application Areas.* pp. 128-137 (2022)

[19] Wang, Y. & Ren, J. Taxi Passenger Hot Spot Mining Based on a Refined K-Means++ Algorithm. *IEEE Access.* **9** pp. 66587-66598 (2021)

[20] Usmani, S., Kamran, S., Zeeshan, M., Islam, N. & Khan, Z. Z."A comparative analysis of apriori and FP-growth algorithms for frequent pattern mining using apache spark". *Proceedings Of International Scientific Research Conference.* (2021)

Edited by: Mudasir Mohd
Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
Received: May 16, 2023
Accepted: Sep 11, 2023