Application of APRIORI correlation algorithm on music education curriculum association rules

Tianjiao Li*
Shangdong Management University, Shandong Province, China
Corresponding author and e-mail: Tianjiao Li, 14438120160072@sdmu.edu.cn

Abstract. Association rules is a research direction of data mining. It can be used in all walks of life by analyzing the correlation of data items in a large amount of data to mine the internal connections between data items. The various professional courses in colleges and universities are independent of each other, and there is a certain sequence of classes, which will also have a certain influence on each other. The learning situation of one course may also affect the learning of other courses. This article uses the APRIORI correlation algorithm to analyze the performance of music education majors, reveals the connections between and within the various course groups, and provides guidance for students' choice and study of professional courses.

1. Introduction
Association analysis is a practical analysis technique, which is to discover the associations or correlations existing in a large number of data sets, thereby describing the laws and patterns of the simultaneous appearance of certain attributes in a thing. The most famous and classic is the APRIORI algorithm, which is used in various scenarios. In the correlation analysis of cigarette physical indicators based on APRIORI association rules, You Liqing excavated the internal laws between physical indicators, machines and work shifts, and provided a reference scheme for the stability control of cigarette physical indicators [1]. Li Fang et al. used the APRIORI algorithm to analyze the company’s shareholder structure and found that there is a certain association law when the company invests in the company [2]. Hangege et al. used the improved APRIORI algorithm to study the quality of meteorological data, and the false detection rate of the data was only 1% [3]. In 2019, Wang Jun proposed the application of association rules in curriculum relevance analysis. He pointed out that in the specific teaching management process, a large amount of original teaching information has been gathered. This original teaching information often contain certain teaching rules and patterns, and it is necessary to make full use of the analysis methods of data mining technology. Taking computer science and technology as an example, he discovered the relevance of professional courses hidden behind the score data based on correlation analysis [4]. In this idea, Huang Chuanteng et al. adopted the method of association rules in the correlation analysis of civil engineering courses. They selected seven representative courses and took the academic performance of the students in the three grades as the research object, explored the correlation between the courses, and provided important guidance for the curriculum system setting and structural adjustment [5]. However, these association analyses are based on these disciplines with strong logic and obvious characteristics of engineering disciplines. There is almost no analysis of some humanities and arts majors. Therefore, this article will use the classic APRIORI algorithm to analyze the correlation of college music professional courses.
2. Principles and algorithms

The APRIORI algorithm is the most basic algorithm for mining association rules, and its core is the recursive algorithm of the two-stage frequency set idea. The APRIORI algorithm mainly includes two aspects: finding frequent itemsets and exploring association rules. The most basic concepts in the implementation of the APRIORI algorithm are support and confidence, support for finding frequent itemsets, and confidence for determining association rules. The support degree indicates the probability of the occurrence of the item set, that is, the proportion of the number of records in the data set that contains the item set in the total data sample. For example, the probability of the item set $X$ and $Y$ occurring at the same time is the number of items containing $X$ and $Y$ and Ratio of all items:

$$Support(X,Y) = P(X,Y) = \frac{\text{number}(X,Y)}{\text{number(all_sample)}}$$

Confidence degree represents the credibility of the rule, that is, the probability that one data appears after another data appears, that is, the conditional probability of the data. For example, the probability that itemset $X$ occurs and $Y$ also occurs is the ratio of the number of items containing $X$ and $Y$ to the number of items containing $X$:

$$Confidence(X,Y) = P(Y/X) = \frac{P(X,Y)}{P(X)} = \frac{\text{number}(X,Y)}{\text{number}(X)}$$

Frequent itemsets are abbreviated as frequency sets, which are all itemsets whose support is greater than the support threshold.

Before executing the algorithm, users set appropriate support and confidence thresholds by themselves. After each iteration, the itemsets whose support degree is greater than the support degree threshold are retained, called frequent itemsets, and the final frequent itemsets generate association rules. The minimum support is the threshold to measure the support, which represents the lowest importance of the item set; the minimum confidence is the confidence threshold, which represents the lowest reliability of the association rule, and the strong association rule means that the minimum support threshold and the minimum confidence threshold are met at the same time.

![Figure 1. APRIORI algorithm flow.](image-url)
3. Experimental results and discussion

3.1. Experimental design

According to the positioning of universities, training talents that meet the needs of the industry and society basically lies in formulating professional talent training programs with clear concepts and ideas. A complete talent training program should include five aspects of training objectives, graduation requirements, curriculum system, syllabus and evaluation system, and has a strict logical relationship from the macro to the micro. Among them, the curriculum system is a link between the previous and the next, which is not only the specific support for the training goals and graduation requirements, but also the clear guidance for the syllabus and evaluation system. Throughout the music education professional curriculum system of various colleges and universities, they all include general education courses, professional education courses and practical courses. Based on the characteristics of the humanities and arts disciplines, the minor language comprehensive curriculum group courses in the general education curriculum, the professional core curriculum group courses in the professional education curriculum, and the professional practice curriculum group courses in the practical courses are all core courses at each stage of the undergraduate course. Curriculum is the framework that supports the professional literacy of music education, and is of decisive significance for the training goals of students' professional literacy.

Therefore, this paper selects 7 core courses of music education majors, and takes the academic performance of 523 students in three grades as the research object, and filters out missed exams, postponed exams, exemptions, violations of discipline, etc. The statistical data results are recorded as data set D, total Contains 523 transactions, and each transaction tk is a 7-item set. Because the comprehensive results of university courses are composed of formative assessment (homework, attendance, classroom performance, etc.) and summative assessment (final exam) in a certain proportion. Jiang Hui and others found that there is no significant linear correlation between the summative assessment scores and formative assessment scores of most courses. In other words, because of the addition of formative assessment, the overall score does not directly reflect the student's mastery of course knowledge. It is easy to understand. On the one hand, formative assessment is aimed at process management, and focuses on examining students’ attitudes and commitment to learning. Students may use other means to search for homework, plagiarize, sign in, etc.) in order to improve their overall performance. Achievement score; on the other hand, in order to make the distribution of comprehensive scores more reasonable, the teacher may artificially adjust the formative assessment scores of students at all levels. Especially for students with academic difficulties, the formative assessment results will greatly improve the overall performance. In order to intuitively measure the level of students' knowledge and ability, this paper chooses the final examination of the summative assessment) as the discrete object of investigation.

In addition, because different subjects, different grades, and different teachers have great differences in the difficulty of writing test papers and reviewing habits, it often results in huge differences in the mean and variance of the course scores, and even some course scores do not conform to the normal distribution. Therefore, it is not possible to simply use the scores to judge the high and low, and the good and the bad. For the same subject, this article uses quartiles for the final exam (final exam) of all students in the same class. After sorting, the top 25% are excellent, the upper 25% are good, the lower 25% are medium, and the last 25%. % is bad. They are represented by 1, 2, 3, and 4 respectively.

Finally, for the convenience of description, use the letters A, B, C, D, E, F for the seven courses of (small language) oral, (small language) synthesis, instrumental music, piano performance, audio-visual ear training, harmony, rehearsal and practice. G said. For example, a student's audio-visual ear training score is ranked in the top 25% of the class, and the student's score is expressed as C1 after data processing. Part of the course scores collected and processed are shown in Table 1:
Table 1. Achievements of some professional courses.

| Number | A | B | C | D | E | F | G |
|--------|---|---|---|---|---|---|---|
| 1      | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| 2      | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3      | 2 | 2 | 1 | 1 | 1 | 1 | 2 |
| 4      | 2 | 2 | 1 | 2 | 1 | 2 | 2 |
| ......  | ...... | ...... | ...... | ...... | ...... | ...... | ...... |

Any two courses I and J courses. Course I starts earlier than J, so there will be a total of 16 student grades for the two courses, as shown in the figure 2:

![Figure 2. Correlation diagram of student performance among courses.](image)

In $\rightarrow J_n$ represents the continuity or inertia of the results between courses, which means that the students in the two courses before and after have achieved the same grade of results. And there is also a kind of $I_1 \rightarrow J_4$ or $I_4 \rightarrow J_1$, which indicates the reversal of the results between courses, which means that the students in the two courses before and after have completely opposite grades. In theory, the closer the connection between the two courses is, the stronger the continuity of knowledge points is and the higher the relevance of the courses is. The stronger the inertia of the same student's test scores, the lower the possibility of reversal, and vice versa. In order to measure the inertia and reversal of student performance, define the inertia index and reversal index as follows:

$$K_{\text{in}} = \sum_{i=1}^{4} (\text{support}(I_i \rightarrow J_i))$$

$$K_{\text{re}} = \text{support}(I_i \rightarrow J_4) + \sup \text{port}(I_4 \rightarrow J_1)$$

3.2. Results and discussion

Table 2 fully shows the results of the correlation analysis among the seven core courses selected in this article.

Table 2. Curriculum correlation analysis results.

| Rule     | Support(%) | Confidence(%) | Rule     | Support(%) | Confidence(%) | Rule     | Support(%) | Confidence(%) |
|----------|------------|---------------|----------|------------|---------------|----------|------------|---------------|
| A1$\rightarrow$B1 | 15.2       | 55            | A1$\rightarrow$C1 | 11.6       | 48.3          | A1$\rightarrow$D1 | 10.4       | 43.2          |
| A2$\rightarrow$B2 | 9.2        | 41.2          | A2$\rightarrow$C2 | 6.8        | 29.3          | A2$\rightarrow$D2 | 4.8        | 20.6          |
| A3$\rightarrow$B3 | 12.4       | 42.7          | A3$\rightarrow$C3 | 6          | 26.3          | A3$\rightarrow$D3 | 7.2        | 27.7          |
| A4$\rightarrow$B4 | 15.3       | 47.2          | A4$\rightarrow$C4 | 12.4       | 47            | A4$\rightarrow$D4 | 12         | 45.5          |
| A1$\rightarrow$B4 | 0.8        | 3.2           | A1$\rightarrow$C4 | 2          | 8.3           | A1$\rightarrow$D4 | 1.2        | 5             |
| A1→B1 | 2.4 | 9.1 | A4→C1 | 2.3 | 10.6 | A4→D1 | 3.6 | 9.6 |
| A1→E1 | 11  | 46.7| A1→F1 | 10.8| 45   | A1→G1 | 12  | 50  |
| A2→E2 | 7.2 | 31  | A2→F2 | 7.2 | 31.0 | A2→G2 | 6.8 | 29.6|
| A3→E3 | 7.2 | 27.6| A3→F3 | 8   | 30.2 | A3→G3 | 10  | 38.5|
| A4→E4 | 10.4| 38.6| A4→F4 | 12  | 45.3 | A4→G4 | 12.9| 48.5|
| A1→E4 | 2.4 | 10.6| A1→F4 | 2.4 | 10.2 | A1→G4 | 5.4 | 10  |
| A4→E1 | 2.6 | 10.1| A4→F1 | 2.8 | 10.6 | A4→G1 | 1.6 | 6.2 |
| B1→C1 | 10.4| 43.3| B1→D1 | 10  | 41.7 | B1→E1 | 10.4| 43.3|
| B2→C2 | 5.6 | 23.6| B2→D2 | 6.4 | 26.3 | B2→E2 | 6.4 | 26.7|
| B3→C3 | 7.2 | 27.7| B3→D3 | 7.2 | 27.6 | B3→E3 | 10  | 27.7|
| B4→C4 | 13.7| 53.1| B4→D4 | 12.9| 50   | B4→E4 | 1.2 | 43.9|
| B1→C4 | 2   | 8.3 | B1→D4 | 2   | 8.3  | B1→E4 | 2   | 8.3 |
| B4→C1 | 2   | 7.8 | B4→D1 | 3.2 | 12.5 | B4→E1 | 4   | 15.6|
| B1→F1 | 9.6 | 40  | B1→G1 | 11.2| 45.6 | C1→D1 | 11.2| 43.6|
| B2→F2 | 6.8 | 28.3| B2→G2 | 8   | 33.3 | C2→D2 | 8   | 36.2|
| B3→F3 | 5.6 | 21.5| B3→G3 | 8.8 | 33.2 | C3→D3 | 5.3 | 25.4|
| B4→F4 | 11.6| 45.3| B4→G4 | 12.4| 48.2 | C4→D4 | 12.3| 42.2|
| B1→F4 | 2.8 | 11.7| B1→G4 | 3.2 | 13.3 | C1→D4 | 1.6 | 8.3 |
| B4→F1 | 2.6 | 10.9| B4→G1 | 1.2 | 4.2  | C4→D1 | 2   | 10.9|
| C1→E1 | 11.2| 46.1| C1→F1 | 10.2| 51.2 | C1→G1 | 13.3| 49.9|
| C2→F2 | 8   | 32.3| C2→F2 | 9.2 | 35.9 | C2→G2 | 8   | 40.1|
| C3→E3 | 5.3 | 22.3| C3→F3 | 7.3 | 35.6 | C3→G3 | 8.5 | 39.6|
| C4→E4 | 13.2| 46.4| C4→F4 | 11.2| 43.3 | C4→G4 | 11.6| 45.2|
| C1→E4 | 2.3 | 6.1 | C1→F4 | 2.3 | 10.2 | C1→G4 | 3.2 | 13.3|
| C4→E1 | 2   | 7.3 | C4→F1 | 3.6 | 12.5 | C4→G1 | 2   | 7.5 |
| D1→E1 | 10.4| 42.3| D1→F1 | 12  | 48.4 | D1→G1 | 11.6| 46.9|
| D2→E2 | 6.3 | 27.8| D2→F2 | 8   | 33.6 | D2→G2 | 6   | 25.4|
| D3→E3 | 6   | 23.4| D3→F3 | 5.6 | 22.2 | D3→G3 | 7.3 | 28.6|
| D4→E4 | 13.3| 50.8| D4→F4 | 12.3| 49.3 | D4→G4 | 12.3| 47.8|
| D1→E4 | 2   | 5.2 | D1→F4 | 0.4 | 1.3  | D1→G4 | 2   | 8.1 |
| D4→E1 | 1.2 | 6.1 | D4→F1 | 2   | 7.7  | D4→G1 | 2.6 | 4.3 |
| E1→F1 | 12.3| 55.4| E1→G1 | 11  | 49.3 | F1→G1 | 13.2| 55.3|
| E2→F2 | 2.4 | 36.3| E2→G2 | 6.2 | 21.3 | F2→G2 | 8   | 33.6|
| E3→F3 | 3.5 | 31.3| E3→G3 | 6.9 | 23.6 | F3→G3 | 9.3 | 39.1|
| E4→F4 | 8.9 | 60.2| E4→G4 | 14.2| 54.3 | F4→G4 | 16.9| 56.4|
| E1→F4 | 0.2 | 1.5 | E1→G4 | 2.3 | 9.8  | F1→G4 | 0.2 | 1.6 |
| E4→F1 | 2.1 | 1.3 | E4→G1 | 0   | 0    | F4→G1 | 0.8 | 2.3 |

The following phenomena and laws can be found:
1) There is a significant phenomenon of solidification of grades among core courses. The solidification of grades generally exists among the selected seven courses, which proves that the music education curriculum system is between the small and medium-language comprehensive curriculum
group courses, the professional core curriculum group courses in the professional education curriculum, and the professional practice curriculum group courses in the practical courses. Close correlation. The solidification phenomenon exists in the groups of good students and poor students. The confidence levels of "excellent" and "poor" in each course are almost all above 40%, and the support and confidence of the two levels of "good" and "medium" are both Less than "excellent" and "poor". In addition, students with medium grades have greater mobility in these two areas, reflecting that these students have the highest perception of the characteristics of the curriculum and the quality of teaching and are more "sensitive".

2) The relevance of the courses within the course group is significantly higher than the relevance between the course groups. The confidence in the inertia of the performance level is high, and the support and confidence of the performance level reversal are very low.

3) The professional core curriculum group courses should pay attention to audiovisual training. It can be seen from Table 2 that the confidence of "audiovisual ear training→harmony" in each grade (In-Jn) is superior to "piano performance-audiovisual ear training" and "instrumental music-harmony".

4. Conclusions

In this paper, seven representative courses are selected from the comprehensive curriculum group courses in the middle and small languages of the music education curriculum, the professional core curriculum group courses in the professional education curriculum, and the professional practice curriculum group courses in the practical courses, and 523 students in three grades are selected. The academic performance is the research object, and the association rule mining algorithm APRIORI is used to deeply discuss the realization process of data mining technology. This article puts forward the concept of course grade inertia and reversal, constructs the inertial index and reversal index, combines the support and confidence between courses, and explores the correlation between courses. It is concluded that there is a relatively significant phenomenon of solidification of grades between core courses, the relevance of courses within the course group is significantly higher than that of courses between courses, and the professional core courses should pay attention to audiovisual training. The analysis results can provide an important reference for scientific curriculum system setting and structural adjustment, accurate control of the core curriculum in the curriculum system, targeted teaching method reform, and academic monitoring and forecasting.

References

[1] You Liqing, Li Xingxu, Li Yongqing, Li Feng, Liu Dan. The correlation analysis of cigarette physical indicators based on APRIORI association rules[J]. Light Industry Science and Technology, 2021, 37(03): 88-91+103.
[2] Li Fang, Guo Jinli, Tan Liming. Association analysis of company shareholders based on APRIORI algorithm[J]. Software Guide, 2021, 20(03): 106-109.
[3] Han Gege, Huang Yanhong, Jiang Nana, Xu Xiaoqing. Research on quality control of meteorological data based on improved APRIORI algorithm[J]. Electronic Testing, 2021(05): 63-64+8.
[4] Wang Jun. The application of association rules in curriculum relevance analysis[J]. Computer Knowledge and Technology, 2019, 15(35): 253-255.
[5] Huang Chuanteng, Pu Shuang, Tang Di, Li Qing. Correlation analysis of civil engineering courses based on association rule mining algorithm APRIORI[J]. China Education Information, 2020(23): 55-58+84.
[6] Jiang Hui, Zhou Ling, Ye Xuefei. Analysis of the relationship between formative and summative assessment results[J]. China Distance Education, 2011(8): 58-62.