Selection of audio materials in college music education courses based on hybrid recommendation algorithm and big data

Tianjiao Li
Shandong Management University, Jinan City, Shandong Province, China

Corresponding author and e-mail: Tianjiao Li, 14438120160072@sdmu.edu.cn

Abstract. Aiming at the problem of audio material selection in college music education courses and to solve the problem of students' low enthusiasm caused by improper selection of audio materials, this paper adopts the hybrid recommendation algorithm combining big data, personalized recommendation algorithm based on Collaborative Filtering Recommendation Algorithm (CF). Big data is used to obtain user evaluation matrix, and Pearson correlation coefficient is used to calculate the similarity between users, so as to form the nearest neighbor set, obtain the nearest neighbor set of target users, and generate the user-based recommendation set. At the same time, a questionnaire is set up to obtain the real evaluation scores of each user on the audio materials. 20% of the questionnaire data are used for model testing and 80% for model training. The accuracy of the square root error is measured, and the prediction score obtained by the model is compared with the real score. It is found that the mean RMSE value of the model adopted in this paper is 0.3813, which is at least 2.564% higher than that of similar models, and has a higher accuracy. Meanwhile, the algorithm is relatively simple, providing a reference for audio selection of college students in courses.

1. Introduction
Music lessons can cultivate people's aesthetic ability and cultivate people's sentiment. Music education can also cultivate the patriotic spirit of college students, and at the same time, intellectual education is helpful [1]. At present, the music education of college students is getting more and more attention, but there are some problems in the existing music education of college students, such as the improper choice of audio in the music education courses of college students, not combining the preferences of college students, resulting in the low enthusiasm of college students to attend classes. In order to alleviate this problem, computer analysis technology is introduced. Currently, the existing recommendation algorithms are as follows:

In 2020, Zheng Xiaonan designed an improved hybrid recommendation algorithm based on the filling of user preference matrix. Taking user rating data, item information and number of neighbors K as input, the user preference matrix was established, and the predicted score of any unrated item was obtained by taking target user's predicted score as output. The experimental results show that the performance of the hybrid algorithm is improved by 2.7%~3.2%, the accuracy by 4.3%, and the recommended performance by 0.8% compared with the traditional algorithm, showing high practicability and reliability [1]. Dong Miao and Wang Qizong designed a collaborative filtering recommendation algorithm based on matrix decomposition and hierarchical clustering. The distance between two clusters was calculated using the Average Linkage clustering algorithm, and the root
mean square error was used as the data evaluation method. It is found that the NDCG value and RMSE value of this method are 0.076 and 0.392 respectively, which have higher precision compared with similar algorithms [2].

Mao Xiaochen and Gu Jie improved the traditional recommendation model by studying the traditional recommendation model and algorithm analysis and combining the characteristics of all parties. They set up an algorithm analysis based on similarity model and then put forward a recommendation algorithm system for big data detection based on multidimensional similarity using Python programming language. It is found that this method can solve the problem of user sparsity well [3].

Ren Min proposed a personalized analysis idea of big data and discussed the problems in data, algorithm, user, cold start and recommendation diversity of personalized recommendation of big data, and proposed a personalized recommendation algorithm that could combine big data with content-based collaborative filtering. She thinks that this algorithm can be applied to the education industry to promote the development of personalized education [4].

Although recommendation algorithm has been widely used in the Internet, but now used in the field of education is less, combining with the data and recommendation algorithm, under the background of big data, based on the content of the personalized recommendation algorithm and the collaborative filtering recommendation algorithm combining the hybrid recommendation, first through the feature extraction of respondents demand, establish a matrix decomposition method for college students of music preferences recommend model, finally, the result of the filtered music big database for final judgment by student evaluation model. Further optimize the selection method of audio materials for music education courses for college students and promote the development of music education for college students.

2. Principles and algorithms

Collaborative Filtering Recommendation Algorithm (CF) is rapidly becoming a popular technology in information filtering and information systems. Different from the traditional content filtering based on direct content analysis for recommendation, collaborative filtering analyzes users' interests, finds out the similar (interest) users of the specified users in the user group, and synthesizes the evaluation of these similar users on a certain information to form the system's prediction of the designated user's preference for this information. Its flow chart is shown in Figure 1:
Firstly, user collection \( U = \{ U_1, U_2, U_3, \ldots, U_m \} \) and music material collection based on big data are established \( I = \{ I_1, I_2, I_3, \ldots, I_n \} \), get the user evaluation matrix \( R_{m \times n} \). When \( R_{ij} = 0 \), it means that the user has not evaluated the item.

\[
R_{m \times n} = \begin{bmatrix}
R_{1,1} & R_{1,2} & \cdots & R_{1,n} \\
R_{2,1} & R_{2,2} & \cdots & R_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
R_{m,1} & R_{m,2} & \cdots & R_{m,n}
\end{bmatrix}
\]

(1)

Pearson correlation coefficient is used to calculate the similarity between users to form a nearest neighbor set. The calculation method is as follows:

\[
sim(a, b) = \frac{\sum_{i \in I_a} (R_{a,i} - \overline{R_a})(R_{b,i} - \overline{R_b})}{\sqrt{\sum_{i \in I_a} (R_{a,i} - \overline{R_a})^2} \sqrt{\sum_{i \in I_b} (R_{b,i} - \overline{R_b})^2}}
\]

(2)

\[
sim(i, j) = \frac{\sum_{a \in I_i} (R_{a,i} - \overline{R_i})(R_{a,j} - \overline{R_j})}{\sqrt{\sum_{a \in I_i} (R_{a,i} - \overline{R_i})^2} \sqrt{\sum_{a \in I_j} (R_{a,j} - \overline{R_j})^2}}
\]

(3)

**Figure 1.** Flow chart of collaborative filtering recommendation algorithm.
Where, sim(a,b) is the user similarity between user a and b, and sim(i,j) is the item similarity of item i and j. Ui,j represents the set of users who have scored both i and j items. Ia,b represent the set of items that both user a and user b have evaluated. Then sort the obtained similar sets to get the target user's nearest neighbor set Uk.

The user-based recommendation set is then generated by obtaining the predicted score Pa,i of user a for any unevaluated item i based on the nearest neighbor set Uk of any user a. The user-based recommendation set is calculated by sorting the predicted score according to the size of the predicted score as follows:

\[
P_{a,i} = \frac{\sum_{b \in U_k} R_{b,i} \times \text{sim}(a,b)}{\sum_{b \in U_k} \text{sim}(a,b)}
\]

(4)

Finally, the model was verified, a questionnaire was set up to obtain the real evaluation scores of each user on the audio materials, and then the predicted scores were compared with those obtained by the model. The Normalized Discounted Cumulative Gain (DNCG) was used to simplify the output data, and then the Root Mean Square Error (RMSE) was used to compare and calculate the predicted value and the real value. The calculation method is as follows:

\[
\text{DCG}(L,u) = \sum_{i=1}^{[L]} \frac{r_{ui}}{d(i)}
\]

\[
\text{NDCG}(L,u) = \frac{\text{DCG}(L,u)}{\text{DCG}(L_{ideal},u)}
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (p_i - q_i)^2}
\]

(5)

(6)

(7)

Where L is the audio list in descending order of recommended score, L_{ideal} is the audio list in descending order of real score, u is the user, and i is ranked r_{ui} is the user's rating of the audio ranked as the i. The set of predicted user ratings in RMSE is \{P_1,P_2,...,P_n\}, the corresponding set of actual user scores is \{q_1,q_2,...,q_n\}.RMSR indicates the higher the recommended accuracy.

3. Experimental design

First of all, combining with big data to understand contemporary college students' music selection preferences, and then on the premise of meeting the teaching theme requirements, 20 relevant audio recordings were found in the database and named as I_1~I_20. The respondents were asked to rate the selected audio recordings with a score of 1-10 and a gradient of 0.1. The higher the score, the higher the audio preference. 5 of them were selected as prediction samples, the remaining 15 were taken as samples for preference analysis, and the 5 extracted audio were scored and predicted, and finally the mean square deviation was calculated and compared with the real value. The experimental process is shown in Figure 2.
A total of 160 questionnaires were issued, 154 were recovered and 152 were valid, with a recovery rate of 96.25% and an effective rate of 98.70%. Due to data limitation, only 10 respondents' scores on audio materials I1~I5 were shown randomly, as shown in Table 1.

Table 1. Real score data.

| Respondents | I1  | I2  | I3  | I4  | I5  |
|-------------|-----|-----|-----|-----|-----|
| R1          | 8.4 | 8.6 | 7.5 | 8.6 | 8.9 |
| R2          | 5   | 9   | 8.3 | 7.5 | 9.5 |
| R3          | 9   | 7.5 | 8.2 | 8.6 | 9.3 |
| R4          | 8.5 | 8.5 | 8.6 | 8.9 | 9.5 |
| R5          | 6.8 | 8.4 | 8.5 | 8.4 | 9   |
| R6          | 7.8 | 7.9 | 7.9 | 8.5 | 8.7 |
| R7          | 8.6 | 8   | 8.2 | 8.1 | 9   |
| R8          | 7   | 9.2 | 8.4 | 8.5 | 9   |
| R9          | 8   | 8.8 | 8.3 | 8   | 9.4 |
| R10         | 6.9 | 8.5 | 8.2 | 7.6 | 8.4 |

4. Results and discussion

The data of I6~I20 in the questionnaire results were input into the user evaluation matrix for analysis and training. The trained algorithm was used to analyze the users, and the predicted value of I1~I5 was obtained. The average contrast between the real value and the predicted value is shown in Figure 3:

Figure 3. Comparison of the mean value of the true value and the predicted value.
According to Figure 3, the mean difference between the real value and the predicted value is small and the predicted value is slightly smaller than the real value. There is a large deviation between the predicted mean value of I1 and the real mean value of evaluation due to the occurrence of extreme values in the data. And the predicted value after normalized treatment of the output results was shown in Table 2.

| Respondents | I1  | I2  | I3  | I4  | I5  |
|-------------|-----|-----|-----|-----|-----|
| R1          | 8.4 | 8.3 | 7.1 | 8.5 | 9.0 |
| R2          | 6.0 | 9.0 | 8.2 | 7.4 | 9.5 |
| R3          | 8.7 | 7.4 | 8.1 | 8.5 | 9.0 |
| R4          | 8.6 | 8.6 | 8.6 | 8.7 | 9.7 |
| R5          | 7.2 | 8.3 | 8.1 | 8.3 | 9.2 |
| R6          | 8.0 | 8.1 | 7.9 | 8.4 | 8.8 |
| R7          | 8.4 | 8.2 | 8.1 | 8.1 | 9.3 |
| R8          | 7.5 | 9.0 | 8.4 | 8.4 | 9.3 |
| R9          | 8.4 | 8.7 | 8.5 | 7.8 | 9.1 |
| R10         | 7.3 | 8.4 | 8.2 | 7.4 | 8.7 |

RMSE was then used to compare and calculate the predicted value with the real value, and the calculation results were shown in Figure 3.

**Table 2.** The prediction results of the recommendation algorithm for student ratings.

**Figure 4.** User evaluation results. (a) True evaluation distribution; (b) Predictive evaluation distribution; (c) The calculation results of RMSE value of each respondent.
As can be seen from Figure 4, the values generated from the questionnaire survey are not uniformly distributed, with the characteristics of random distribution and scattered distribution. Conventional methods are difficult to produce more accurate predictions. The mean RMSE value of the CF based on user evaluation is 0.3813. According to the data of Dongmiao Research in Shanghai Xingjian Vocational College [2], the Mean algorithm's RMSE value is about 0.48, the ITEM-based RMSE value is about 0.42, and the MF-based RMSE value is about 0.39, as shown in Figure 4. The relationship between RMSE value and the number of samples is shown in Figure 5.

![Figure 5. Comparison of RMSE values of Mean, ITEM-based, MF-based and hybrid recommendation.](image)

Figure 5 shows that under the background of big data, the hybrid recommendation algorithm combined with content-based personalized recommendation algorithm and CF adopted in this paper has significant advantages, and its accuracy is at least 2.564% higher than other algorithms. Due to the subjective selection of audio frequency, the data collected through questionnaire survey are distributed in scattered points, which is difficult to analyze with traditional methods and low accuracy. Using big data as well as the hybrid recommendation algorithm on the one hand, with the support of big data technology combining sample quantity is big, can more accurately find the general law, improve the algorithm accuracy and moderate, reduce the random sample, on the other hand, the personalized recommendation algorithm based on content and based on collaborative filtering recommendation algorithm combining the hybrid recommendation, increase the prediction accuracy and practicability.

![Figure 6. Relationship between RMSE value and sample data.](image)
As can be seen from Figure 6, the relationship between RMSE value of each method and the number of samples is that with the increase of the number of samples, the RMSE value first decreases and then increases, and it is the minimum value at the number of samples around 150, which may be due to the choice of the adjacent number $K$, which is too large or too low, and will have an impact on the final result. In conclusion, compared with other methods, hybrid recommendation algorithm has higher reliability and accuracy, so it can be used in music education of college students. Such as the big data analysis, forecast the student to the audio evaluation points, get preference matrix, and near to the collection, the choice of music curriculum in the audio provide effective reference, under the premise of school can meet the requirements of teaching, according to the big data and hybrid recommendation algorithm for audio recommended to improve the classroom enthusiasm has certain help. At the same time, it will be possible to build a database of students' audio preferences and update the data in a timely manner, so that the classroom can not only meet the requirements of teaching tasks, but also mobilize students' enthusiasm to the maximum extent.

5. Conclusions

Combining with the large data and recommendation algorithm, under the background of big data, the personalized recommendation algorithm based on the content and the CF recommendation algorithm of combining the hybrid recommendation, establish collaborative filtering algorithm based on user, build user evaluation matrix, according to Pearson correlation coefficient to calculate the user, the similarity between the form near to the collection, production recommended collection, design questionnaire to obtain real value at the same time, 20% is used in the model test, 80% used for model training. The results showed that the mean value of RMSE of the model adopted in this paper was 0.3813, which was at least 2.564% higher than that of similar models, which was of great significance for the audio selection of college students in music courses.

References

[1] Zheng Xiaonan, TAN Qinhong, Ma Hao, Liu Wuqi. Improved hybrid recommendation algorithm based on user preference matrix filling [J]. Computer engineering and design, 2020, 41(10):2784-2790.

[2] Dong Miao, Wang Qizong. Collaborative Filtering recommendation Algorithm based on matrix decomposition and Hierarchical clustering [J]. Electronic Technology and Software Engineering, 2020(17):183-185.

[3] MAO Xiaochen, Gu Jie, Hang Xuelian, Gong Wenbin. Big Data Detection recommendation Algorithm System based on Multidimensional Similarity [J]. China High and New Technology, 2019(16):8-10.

[4] Ren min. Personalized recommendation analysis of big data [J]. Internet of things technology, 2019, 9(11):62-64+67.

[5] Wu Yu. A Review of the Significance of Music Education in China in the first half of the 20th Century [J]. Contemporary Music, 2020(10):136-139.