Asia-Pacific Journal of Information Technology and Multimedia 14(2) 2025: 565 – 584
https://doi.org/10.17576/apjitm-2025-1402-14

Collaborative Filtering Model for Chinese Music Genre Recommendation System

Model Penapisan Kolaboratif untuk Sistem Pencadang Genre Muzik Cina

*Luo Wenhao, Azuraliza Abu Bakar**

Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia

**Corresponding author: azuraliza@ukm.edu.my*

Received 6 March 2025
Accepted 10 September 2025, Available online 15 December 2025

ABSTRACT

The rapidly increasing amount of information can meet the diverse needs of users, but it also faces the challenge of data screening. In the face of massive music resources, users often can't make quick and appropriate choices. Therefore, music recommendation system has become an effective solution tool in this context and has been applied by many large streaming media platforms. However, the current mature music recommendation system still has challenges including lack of personalized recommendation, cold start and data sparsity problems. This article aims to achieve high-precision personalized Chinese music recommendation by using a hybrid scheme of content-based and collaborative filtering algorithms. Our specific objectives are three folds; (i) to propose features and score about Chinese music items to achieve personalized recommendation. (ii) to propose a content-based method to solve the cold-start problem. (iii) to formulate an efficient similarity computation to overcome the data sparsity problem. Four machine learning algorithms were employed: K-nearest neighbors algorithm, Singular Value Decomposition, Latent Factor Model, and Non-negative Matrix Factorization. Experimental results show that the hybrid solution combining content-based and KNN algorithms performs best in addressing cold start, personalization, and data sparsity problems.

Keywords: Chinese music recommendation, hybrid collaborative filtering model, genre base

ABSTRAK

Jumlah maklumat yang semakin meningkat dengan pantas boleh memenuhi keperluan pengguna yang pelbagai, tetapi ia juga menghadapi cabaran penyaringan data. Dalam menghadapi sumber muzik yang besar, pengguna selalunya tidak dapat membuat pilihan yang pantas dan sesuai. Oleh itu, sistem pencadang muzik telah menjadi alat penyelesaian yang berkesan dalam konteks ini dan telah digunakan oleh banyak platform media penstriman yang besar. Walau bagaimanapun, sistem pencadang muzik matang semasa masih menghadapi cabaran termasuk kekurangan pencadang yang dipersonalisasi, permulaan sejuk dan masalah keterlaluan data. Artikel ini bertujuan untuk mencapai pencadang muzik Cina dipersonalisasi berketepatan tinggi dengan menggunakan skema hibrid algoritma penapisan berasaskan kandungan dan kolaboratif. Objektif khusus kami adalah tiga lipatan; (i) untuk mencadangkan ciri dan skor tentang item muzik Cina untuk mencapai cadangan yang dipersonalisasi. (ii) untuk mencadangkan kaedah berasaskan kandungan untuk menyelesaikan masalah cold-start. (iii) untuk merumuskan pengiraan persamaan yang cekap untuk mengatasi masalah sparsiti data. Empat algoritma pembelajaran mesin telah digunakan: Algoritma jiran terdekat K, Penguraian Nilai Tunggal, Model Faktor Terpendam dan Pemfaktoran Matriks Bukan Negatif. Keputusan eksperimen menunjukkan bahawa penyelesaian hibrid yang menggabungkan berasaskan kandungan dan algoritma KNN berprestasi terbaik dalam menangani masalah cold-start, personalisasi dan masalah sparsiti data.

Kata kunci: Pencadang muzik Cina, model penapisan kolaboratif hibrid, asas genre

INTRODUCTION

The application of recommendation technology responds to streaming media's need for filtering overloaded data. Music is one of the popular fields of streaming media. Richter (2023) found that the global music industry has seen eight consecutive years of sustained growth. Digital music accounted for the largest share of global music revenues in 2021, and streaming services alone accounted for 67% of the industry's total revenues. In fact, digital media has become the mainstream channel for people to consume music. According to Global Streaming Statistics presented by Silber (2019), the number of global music streaming subscribers in 2018 was 180.3 million. According to the survey by McCain (2023), ‘Tencent Music’ (China) ranks third with a 13% market share in the world. Over time, users are more likely to fall into the confusion of choices in the huge digital music library. To solve the subjectivity of user preferences and information overload issues, major music streaming platforms have launched recommendation model.

Music recommendation model which is called machine learning application with modern significance. It reduces the search scope and recommendation function. With the mature application of recommendation technology, the existing recommendation systems are divided into Content-based, collaborative and hybrid. In item-based recommendation, content is defined by items and the attributes or characteristics of items are used as the basis for recommendation. Collaborative recommendation can predict the interaction between users and projects in the future. However, there are still flaws in terms of the current application in the survey by Kundu (2020). Past studies by Fayyaz et al (2020), Dahdouh et al (2019) and Kuar and Mohapatra (2021), have highlighted that music recommendation system current challenges are the lack of personalization, cold start problem, data sparsity. In addition, previous studies on Chinese music recommendation models are relatively limited.

The main contribution of this study is the development of a hybrid music recommendation system for Chinese features, leveraging content-based techniques to enhance collaborative filtering algorithms. The model built by best performance algorithm and similarity measurement tool can significantly improve the personalized of the recommendation results, solve the problems of data sparsity and cold start. The paper first contextualizes the music recommendation system, then describes data collection and development, feature engineering, the recommendation model's implementation process and method evaluation, and concludes with a comparative analysis of results. Finally, the discussion is about that this model can effectively solve the problems of lack of personalized recommendation, cold start and data sparsity. This research not only provides a new model design idea for Chinese music recommendation system but also provides reference value for the application of machine learning-collaborative filtering algorithm in the field of Chinese streaming media.

MATERIALS AND METHODS

As shown in Figure 1, the methodology chapter comprises five phases: data collection and development, Chinese music feature extraction, design of the proposed collaborative filtering algorithm, development of the music recommendation method, and method evaluation.

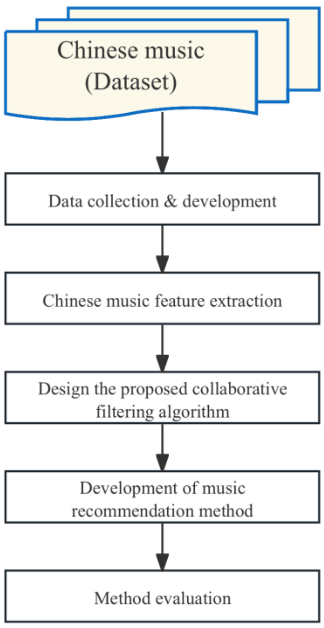

FIGURE 1. Process of Chinese music recommendation

PHASE 1: DATA COLLECTION AND DEVELOPMENT

DATA

The experimental dataset used in this paper contains music item information and user information and behavior records. The songs list data set is based on the existing 3773 playlists on NetEase Cloud Music Platform. The specific content is the relevant attributes and information of Chinese music, including music list name, music name, singer name, music style, music id. User information mainly consists of interest tags, which are collected from questionnaires of 10 users. User behavior records include all operations of these 10 users when using this music recommendation system.

DATA DESCRIPTION

Parsing the 3773 song lists yields a total of 207,922 Chinese songs. In addition to the style definition consistent with the form, these independent song items are also attached with independent information that is only used to introduce specific songs. Music name is one of the most basic attributes of a song. It is also an important channel for users to directly query and obtain target music. Among the song names collected in this data set such as WONIU, JUEJIANG. The Singer name marks the original singer or cover artist. The singer names collected in this data set include well-known Chinese singers such as Jay Chou, Tian Fuzhen, and Li Ronghao. Music style is one of the indicators for effectively classifying music. It can be defined by music rhythm speed, language, expressed emotions, and application scenarios, among other factors. The music styles in this dataset include pop, rock, ballad, jazz, rap, Cantonese, and country. A music ID is a numerical identifier representing a song's unique identity in the music library, which is crucial for platforms to automatically locate, play, and manage songs.

To test the performance of the system, I collected 10 questionnaires from friends via Google forms, which served as data on Chinese music preferences. All survey respondents were Chinese music listeners over 18, with a male-to-female ratio of 7:3; half reported having long used and benefited from music recommendation systems. As shown in Figure 2, User3144, User1277, User7500, and User5443 prefer pop; User1282 prefers rock; User2697 prefers ballad; User4751 prefers jazz; User6915 prefers rap; User6679 prefers Cantonese music; and User5476 prefers country music. Therefore, extracted Chinese music samples with these seven characteristics as research objects to complete the system test. And also invited the above 10 users to personally experience the music system designed for this project and recorded their behaviors - including liked, swipe across, listened and downloaded. It contributes greatly to the evaluation of music score and deriving differentiated recommendation lists.

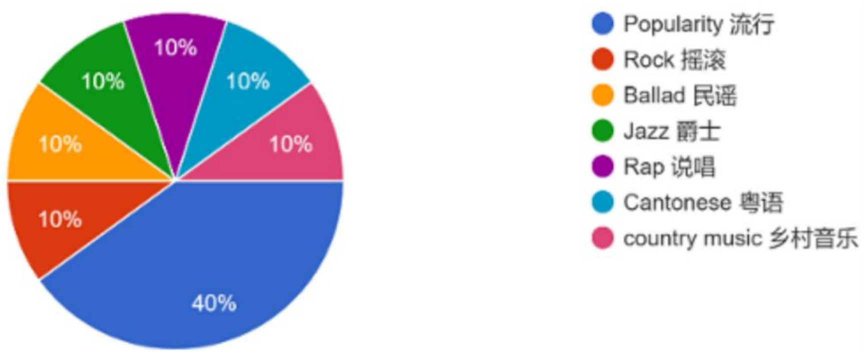

FIGURE 2. Survey of 10 users

The Chinese music with these seven features from 207,922 songs are extracted and counted. As shown in Table 1, the sample size is 3759. 3289 songs with rock tag, 751 songs with ballad tag, 375 songs with ballad tag, 550 songs with Cantonese tag, 712 songs with rap tag, and 631 songs with Countryside tag. According to the statistics, there are 10,067 Chinese music items with the above seven features after data preprocessing.

TABLE 1. The songs sample size for different interest tags

User ID	Interest tags	Sample size of songs
user3144, user1277, user7500, user5443	Popularity	3759
user1282	Rock	3289
user2697	Ballad	751
user4751	Jazz	375
user6679	Cantonese	550
user6915	Rap	712
user5476	Countryside	631
		10,067

DATA PREPROCESSING

To make the dataset more consistent with popular preferences and to reduce memory space, eliminating song lists with <100 subscriptions. Used the regular expression of '[a-zA-Z]' to exclude songs or song lists with English, and units of measurement with unified Arabic numerals. Setting maximum unit comments in the count number is limited to *0,000. Moreover, to ensure the efficient feature engineering, it is necessary to remove data items without comment records. Finally, tag data of the playlist were added to the song's attributes. Setting set of train and test ratio is 75:25.

PHASE 2: CHINESE MUSIC FEATURE EXTRACTION

Basiński et al (2021) found that the most widely used music feature extraction methods are mainly divided into metadata and sentiment analysis of lyrics, usually using feature labels based on genre labels and audio samples. Yılmaz and Scheffler (2023) think that the established attributes of music can be conceptualized and measured to fit the user's music preferences, namely songs' own attributes. The Chinese music features in this study are from songs' own attributes, displayed as list of music style in dataset. Therefore, the Chinese music feature extraction process is equivalent to the dataset extraction process, using the data crawler. Claussen & Peukert (2019) provide guidelines for data crawling and state that Python is the most popular and widely used programming technical in the data crawling process. In this study, the process of crawling is summarized as boundaries for data crawling, identification of observations, obtaining the content. As shown in Figure 3, the process involves three steps: requesting the interface using the required parameters; parsing the JSON format with Python to extract data content (including playlist name, type, ID, introduction, included songs, song names, and song lyrics); and creating a table based on the JSON return fields, then inserting the data into the table to complete data entry.

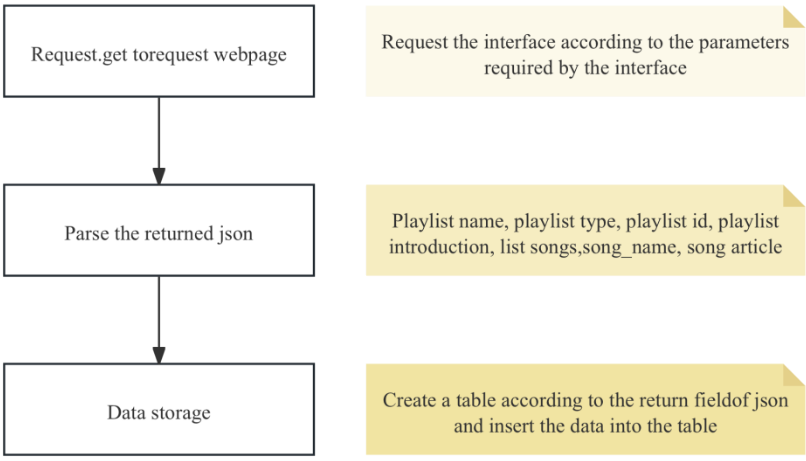

FIGURE 3. Chinese music feature extract process

PHASE 3: DESIGN THE CONTENT-BASED FILTERING MODEL TO SOLVE THE COLD-START PROBLEM

CONTENT-BASED FILTERING METHOD

Nallamala et al. (2020) showed that in content-based filtering, recommendations depend on the user's previous choices or desired elements. Thus, item descriptions and user-oriented profiles are essential in implementing content-based filtering. Specifically, the relationship between the user preference and items' attributes is established in the user-item matrix. This study's content-based filtering mechanism checks the item genre (features) and user interest-matching results. A positive result means the music is consistent with the user's interests, while a negative result means that the music is inconsistent with the user's interests. In theory, content-based filtering performs well in solving cold start problems and offline recommendations. On the one hand, the user-based cold start problem has been solved by filling in at least one interest tag defined by the music genre at the time of registration, which is a common and realized way to collect user information actively. On the other hand, item cold start problem is solved. The adaptation of the same tags allows new items to be recommended immediately.

PHASE 4: DEVELOPING A MUSIC RECOMMENDATION METHOD FOR MUSIC ITEM SIMILARITY CALCULATION TO OVERCOME THE DATA SPARSITY PROBLEM

In this study, collaborative filtering mechanism is to sort the prediction scoring results from high to low, focusing only on the recommendation of items with Predict high-scoring and filtering out irrelevant items with lower scoring results. He (2021) notes that the steps of item-based collaborative filtering are as follows: constructing an item-item score matrix, calculating item similarity, selecting nearest neighbor items, predicting scores, sorting them from high to low, and deriving a recommendation list composed of the top N music items.

MUSIC SCORE

Music score is an important decision-making basis for personalized recommendation, which comes from the different preference conformity and corresponding behavioural response of each user to different music items. In past research, independent and comprehensive user personas were described by quantifying past user behavior data. The interaction with each music item in the past was focused on to measure and select the music genre with the largest proportion as the user feature label. (Wu & Lu, 2025; Dong, 2023) This study chooses to use the current user's likes, swipes, listens, and downloads of a certain music to evaluate the user's level of preference for that music, which is in line with the principle of collaborative filtering recommendation. However, considering the cold start problem caused by the lack of user behavior data in the early stage, this setting integrates music rating rules and expands to consider tag and popularity ratings, in response to the principle of content-based recommendation. Table 2 shows the music scoring rules and corresponding scoring values.

MUSIC RECOMMENDATION ALGORITHM

1. K-nearest neighbors algorithm (KNN)

KNN is a common mode-based method of content-based filtering (CBF). This is based on its ability to cope with the calculate challenges of project datasets with large-scale features and to be simple to implement and efficient. Roy and Dutta (2022) show that KNN function in solving problems of classification and regression and significantly improves their accuracy. In this study, the core recommendation algorithm of k-nearest neighbors algorithm (KNN) in this project, which uses user score data as a support for the derivation of predictive values, and similarity results become the key to determine the recommended items. After the similarity calculation results are derived for sorting by taking the top K music items and using their similarity to the target object as weights, the scores are weighted and summed and finally the results are normalised by the sum of the similarity between these K music items or users and the target. As show in equation 1 according to the study by Schedl (2019), where r_{ui} is determined by the ratings of u for similar items, $N_u^k(i)$ represents the set of the top k items most similar to the current item i in the item set evaluated by User u.

$$r_{ui} = \frac{\sum_{j \in N_u^k(i)} sim(i,j) r_{uj}}{\sum_{j \in N_u^k(i)} sim(i,j)} \tag{1}$$

2. Singular Value Decomposition (SVD)

Presented by Jiang and Chaudhuri (2023), SVD reduces the number of features of the dataset by reducing the spatial dimensions from N dimensions to K dimensions (where K < N). As shown in equation 2, where U and V are orthogonal matrices and Σ is the pair A rectangular diagonal matrix with nonnegative values on the diagonals. A matrix structure is used in collaborative recommendation systems, where each row represents a user, and each column represents an item. The elements of this matrix are the user's ratings of the items. According to Equation 2, A is an m×n utility matrix; U is an m×r orthogonal left singular matrix, representing the relationship between users and Chinese music genres; Σ is an r×r diagonal matrix, describing the strength of each Chinese music genre; and V is an r×n right singular matrix (not diagonal), representing the similarity between items and Chinese music genres. SVD reduces the dimensionality of the utility matrix A by extracting its latent factors, mapping each user and item to an r-dimensional latent space to clearly represent their relationship.

$$A = U\Sigma V^T \tag{2}$$

3. Latent Factor Model (LFM)

In study by Tegene et al (2023), LFM is an efficient feature mapping method. Yu (2020) explained that LFM can identify hidden topics or categories and establish relationships between features through hidden topics or categories. Specifically, interest preferences are obtained based on the user's current preference information and items corresponding to such interests are recommended to the current user. Therefore, the core idea of the LFM algorithm is to set the user-item rating matrix, solve two low-dimensional matrices, and multiply the two low-dimensional matrices to approximately represent the rating matrix. Also rooted in matrix factorization, the Latent Factor Model (LFM) characterizes users and items through factor vectors inferred from user-item rating matrices, serving to identify latent topics associated with these items. The LFM algorithm's formula is therefore presented in Equation 3, where: the m×n rating matrix is denoted as R; the user factor matrix is P (n×f); the item factor matrix is Q (m×f); R[i][j] represents user j's rating of item i; P[j][k] denotes user j's interest level in factor k; Q[i][k] indicates the contribution of factor k to item i; and T(Q) signifies the transpose of matrix Q.

$$R_{UI} = P_U Q_I = \sum_{K=1}^{K} P_{U,K}\, Q_{K,I} \tag{3}$$

4. Non-negative matrix factorization (NMF)

NMF aims to automatically extract hidden patterns from a series of high-dimensional vectors. Zhang (2012) was concluded that NMF has good interpretability based on non-negative constraints, flexibility in the choice of objective function and solution algorithms, multi-channel usage and a solid theoretical foundation. In mathematics, the formulation of Non-negative Matrix Factorization (NMF) is presented in Equation 4. Here, the rank-r matrix H contains r modes, representing the dominant spectral modes in the original data; the weighting matrix W assigns each sample a weight for each spectral pattern, indicating the importance of that pattern to the sample. In this project, matrix A represents user-music item ratings, matrix W denotes user preferences for latent features, and matrix H signifies music items' associations with latent features. These latent feature preferences derive from additional labels beyond the defined interest labels: for instance, if the defined interest label is "pop," the latent features might include "rap" and "rock." The final matrix decomposition thus predicts users' interest in unrated music, with Chinese music items exhibiting stronger interest intensity being more likely to be recommended to specific users.

$$A = WH \tag{4}$$

SYNTAX CODE

For each unrated item "i", first traverse all items "j" that the current user has rated (to prepare for similarity calculation). Obtain the current user's ratings for items "i" and "j", treat these ratings as two vectors, and calculate their similarity "s". Then, multiply this similarity "s" by a weight—specifically, the current user's rating for item "j"—for each unrated item "i". After traversing all rated items, sum the weighted similarities from each iteration and divide by the total similarity to get the predicted value. Once all unrated items are processed, recommend items to users based on the highest predicted values. Below is a logical representation of the syntax code.

MUSIC ITEM SIMILARITY MEASUREMENT

1. Similarity measure

In the study by Liu and Zhang (2023), Pearson correlation coefficient measures the degree of linear correlation between two one-to-one corresponding series. As shown in equation 5, the result of Pearson correlation can directly express the similarity between two music items *a* and *b*. When the result is close to 1, the two music items a and b are very similar. Thus, the system will not push music item *b* to listeners of music *a*.

$$\text{Pearson Correlation} = \frac{n(\sum ab) - (\sum a)(\sum b)}{\sqrt{[n\sum a^2 - (\sum a)^2][n\sum b^2 - (\sum b)^2]}} \tag{5}$$

In the study by Anuzaj and Luma (2022), while cosine similarity measures the cosine of the angle between the two vectors which is determines whether the direction of the vector is consistent or inconsistent in multidimensional space. When the value is used as the judgment index, the cosine results of 1 means that vectors point in the same direction, music a and music b are completely similar. Therefore, the system will recommend music b to the fans of music a.

$$\text{Cosine similarity (a, b)} = \frac{(a, b)}{\|a\|.\|b\|} \tag{6}$$

2. Similarity threshold

The similarity threshold is a great impact on the accuracy and coverage of recommended results. The greater the number of recommended items, the higher the diversity of recommendations, but it will also increase the computational complexity and recommendation time. Therefore, adjustments need to be made based on actual needs and data conditions. This model will determine the similarity threshold according to the number of recommended items N with the best performance.

PHASE 5: METHOD EVALUATION

PRECISION

Gunawardana and Shani (2009), Visa and Salembier (2014) think that precision measures the rate of true positives among all detections, as shown in equation 7. Focus on the items to be recommended, TP + FN is defined as the items to be recommended.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

RECALL

Visa and Salembier (2014) think that recall measure the percentage of detected ground truth annotations, as show in equation 8. It centers on items of interest to the user. Therefore, TP + FN is defined as the set of items in which the user has an interest. A recall value closer to 1 indicates that the recommendation results for positive categories are nearly perfect.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

ACCURACY

Accuracy, as shown in Equation 9, refers to the proportion of correctly predicted samples among all samples. A notable feature of accuracy is that its results may be misleading when dealing with imbalanced category proportions. The closer its value is to 1, the more accurate the predictions of both positive and negative samples.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

COVERAGE

Shinde & Potey (2016) state that coverage enables the assessment of the proportion of available information that can be predicted. As shown in equation 10. Where S_r represents the items number that can be recommended at one time, S_a represents the total number of all items in the database that can match the preference label.

$$Catalogue\ coverage = \frac{S_r}{S_a} \tag{10}$$

ANALYSIS OF RESULTS

PERFORMANCE OF DIFFERENT SIMILARITY MEASURES

As shown in Figure 4, the KNN performance is demonstrated based on Cosine similarity and Pearson correlation similarity in the initial model, that is, the K value is not defined. When precision is selected as the evaluation index, the value of cosine similarity is higher than 70% and its performance is better than Pearson correlation similarity. When recall is selected as the evaluation index, the value of Pearson correlation similarity is higher than 60% and its performance is better than cosine similarity. When accuracy is selected as the evaluation index, the accuracy of cosine similarity is higher than 75% and its performance is better than Pearson correlation similarity. This means that this system has a high accuracy in selecting and recommending Chinese music based on user tags. In summary, Pearson correlation similarity is recommended as the similarity measure when the business goal is to improve recall, while cosine similarity is preferable for enhancing precision or accuracy.

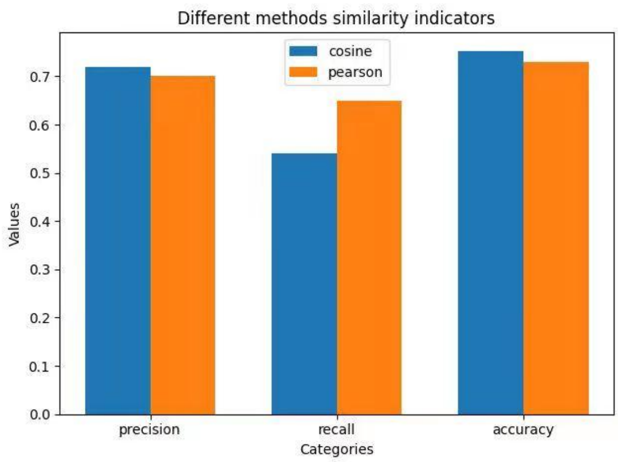

FIGURE 4. Similarity result for Cosine and Pearson

PERFORMANCE OF DIFFERENT RECOMMENDED ITEM QUANTITIES

There are differences in the performance of evaluation results in different recommended items by KNN. As shown in Figure 5, the k value corresponding to the peak result of each evaluation indicator result is the optimal recommendation number. When precision is the evaluation indicator, the number of recommendations with the best effect is up to 25. When recall is used as an evaluation indicator, the best number of recommendations is 20. When accuracy is used as an evaluation indicator, the best-recommended number is 10. When more recommended items need to be satisfied, the performance of precision and recall is better than accuracy. When less recommended items need to be satisfied, accuracy performs better than precision and recall. Therefore, the results of this experiment show that the number of recommended items can be increased (up to k=20 in this study) appropriately when we have recall as a business goal. The recommended items can be increased (up to k=25 in this case) appropriately when we have precision as a business goal. Moreover, the recommended items can be appropriately decreased (up to k=10 in this case) when we have accuracy as the measure. This is consistent with the experimental results of Ezeh (2023): a smaller K may lead to higher accuracy, while a larger K may lead to higher recall.

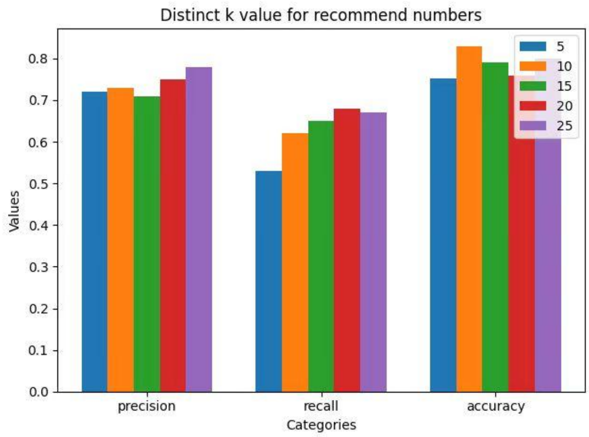

FIGURE 5. Performance evaluation result of different k-items case

PERFORMANCE OF ONE USER UNDER DIFFERENT INTEREST TAGES

Select only one user, name of user3144, as the experimental object to evaluate the performance of different tags in a music recommendation system based on KNN algorithm, confirmed based on the accuracy, recall and coverage. As shown in Figure 6, all recommendation results can cover the user3144 'rock' preferences ratio is 65%, 'ballad' preferences ratio is 78%, 'Jazz' preferences ratio is 69%, 'Cantonese' preference ratio is 74%, 'Countryside' preference ratio is 69%.

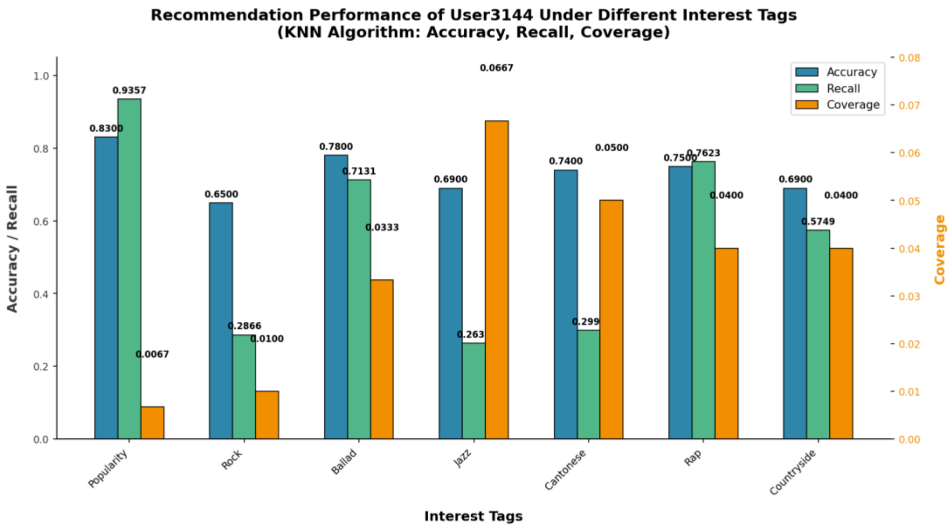

FIGURE 6. Accuracy, recall, coverage result for User3144

From equations 6 and 7, the number of classes predicted to be positive varies with the classification threshold. Recall increase with increasing TP and decreasing FN. Accuracy increase with the increase of TP or TN. Coverage increase as the number of samples that meet user preferences increases. In the study of Foody (2023), it was pointed out that the reason for the large imbalance in coverage is that the greater the deviation of Interest tags from 0.5 in the total sample, the greater the degree of imbalance. Since the number of songs corresponding to each independent Interest tags accounts for a different and smaller proportion in the overall sample, and each coverage value only assesses 1 user preference tag, Coverage is presented in this project A huge imbalance. In general, higher accuracy means lower recall. Because the system will focus more on items that closely match the user's preference and ignore the more marginal related items. Moreover, higher recall may mean lower coverage. This is because the system may focus more on Chinese music items that are relevant to the user's preferences to ensure coverage of the user's interest labels, while ignoring items that are not relevant to the interests in the whole sample. It follows that increasing coverage is unlikely to achieve a parallel increase in accuracy.

PERFORMANCE OF DIFFERENT RECOMMENDATION ALGORITHMS

As shown in Figure 7, the average accuracy of KNN for recommending songs across different users is 0.753; for SVD, it is 0.742; for LFM, 0.587; and for NMF, 0.591. KNN emerges as the optimal algorithm in this music recommendation system, with stronger adaptability to diverse users. Additionally, four users sharing the same "popularity" interest tag exhibited varying performance across the four algorithmic models, indicating that individual user differences influence model performance.

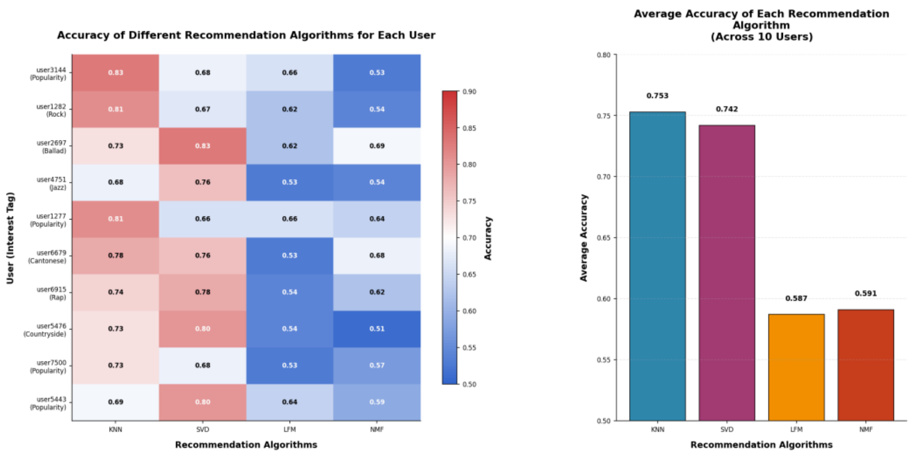

FIGURE 7. Average accuracy results for 10 users

As shown in Figure 8, the average recall of KNN for song recommendations across different users is 0.617; for SVD, it is 0.559; for LFM, 0.404; and for NMF, 0.37. Thus, KNN emerges as the optimal algorithm.

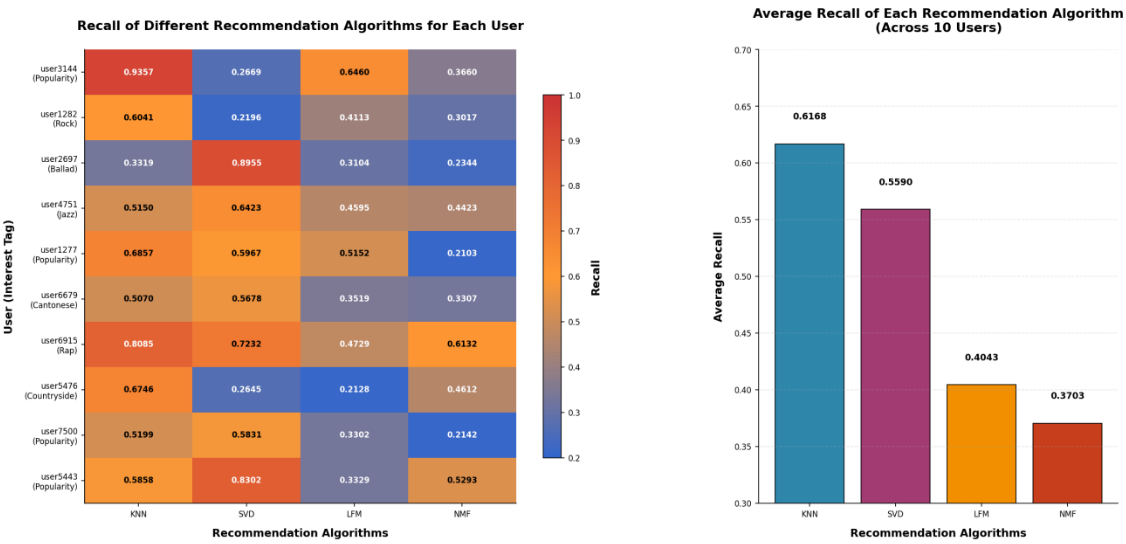

FIGURE 8. Average recall results for 10 users

As shown in Figure 9, the average coverage of KNN, SVD, LFM and NMF is 0.027 for different user music recommendations. In fact, the results of the same average coverage value for users who selected the same tag items are due to the influence of the scoring system based on tag selection, not due to differences in algorithms. Therefore, the coverage values of the four users who all selected popularity as the interest tag are the same based on different algorithms. Moreover, based on the difference in the core idea of matrix decomposition of each algorithm and the influence of the fixed proportion of different tags in the total music library (data set) in the data set, the coverage results of different tags are different. Popularity tag coverage is 0.0067. Rock tag coverage is 0.01. Cantonese tag coverage is 0.05. This is consistent with the results found by Longo (2018) that recommender systems using collaborative algorithms can only recommend a very small fraction of the training set items, which manifests itself as a much lower coverage of collaborative recommenders than random recommenders.

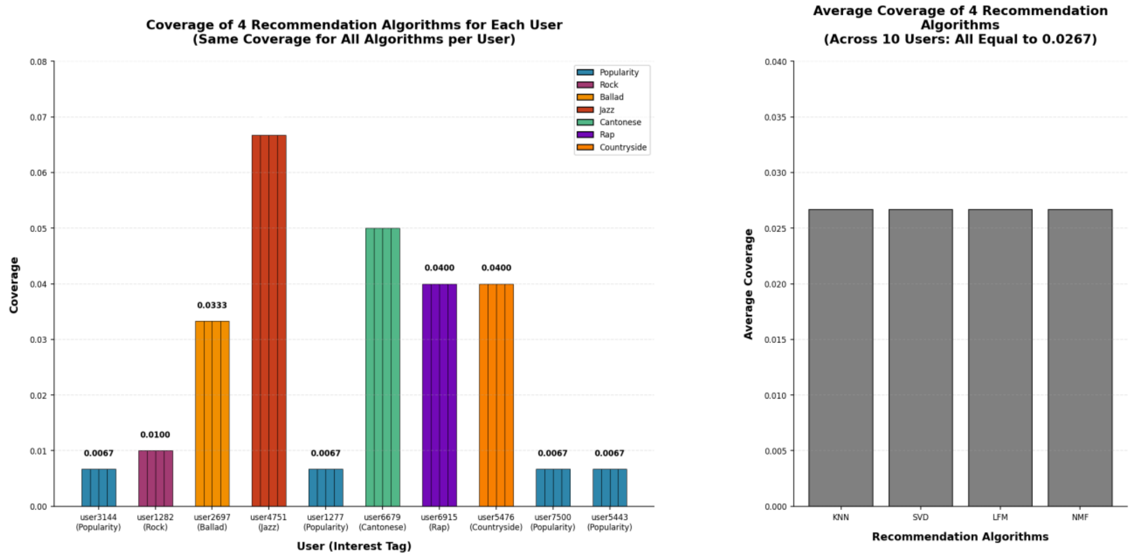

FIGURE 9. Average coverage results for 10 users

DISCUSSION

CHINESE MUSIC TAGS AND SCORES FOR IMPROVED PERSONALIZED RECOMMENDATION

The recommendation system is combined with music user settings to calculate users' ratings of music items through the selection of preference tags. In this paper, seven Chinese music features are listed as user interest tags, including 'popularity', 'rock', 'ballad', 'jazz', 'Cantonese', 'rap' and 'countryside'. In addition, each user's unique behavioral scores, tag scores and popularity scores are included in the total project score calculation equation, resulting in a personalised list of recommendations. As is well known, personalization means that the recommendation results are highly responsive to user preferences, so accuracy can be used as the basis for judging personalized recommendations. According to the experimental results, user3144's accuracy values for the recommendations of the seven different interest labels are different and in the middle to upper level, which means that the recommendations satisfy user3144's personalisation needs. Among them, 83% of the items recommended to user3144 are truly in line with their popularity interests. And in the controlled experiments with different algorithms, the music recommendation system applying 'KNN' algorithm outperforms 'SVD', 'LFM' and 'NMF' with 75.3% accuracy results. Music systems equipped with 'KNN' have better personalized recommendation performance in this study.

CONTENT-BASED METHOD TO SOLVE THE COLD-START PROBLEM

In this study, sufficient item support and matrices help solve the problems of cold start. It's also demonstrated by Thomas and John (2021). In this study, 3773 playlists dataset achieve a huge number of music items support. And through the tag to extract the characteristics of each item as a basis for matching the similarity characteristics of new items, and ultimately solve the problem of item cold start. To solve the user cold start problem, this project uses the preference information filled in by the user during registration for initial personalised recommendations.

The application of KNN improves the accuracy and directly utilizes the relationship between samples, reducing the adverse impact of improper selection of category features on the classification results, that is, it is insensitive to outliers. The experimental results show that accuracy is the best evaluation index when there are few recommended items, and Precision and recall are the best evaluation indexes when there are many recommended items. Furthermore, the recall, accuracy, and coverage results of the KNN algorithm across different interest labels indicate that it can provide more accurate and diverse recommendations. Additionally, comparative experiments with KNN, SVD, LFM, and NMF demonstrate that KNN is the optimal recommendation algorithm in this study.

SIMILARITY COMPUTATION TO OVERCOME THE DATA SPARSITY PROBLEM

In this study, the author used to calculate the similarity between items by Cosine similarity and Pearson correlation similarity. Cosine similarity is better than Pearson correlation similarity under select performance evaluation of precision and accuracy in this study. Pearson correlation similarity is better than Cosine similarity under select evaluation index of recall. As we all know, the accuracy of 'Pearson correlation coefficient' results may be affected by too few common scoring items due to sparse data. However, 'cosine similarity' computing is less sensitive to data sparsity. Even if there are few common scoring items but the spatial vector features are obvious, relatively stable similarity results can be obtained. Therefore, when implement personalized recommendation and solve the problem of data sparsity, cosine similarity is the best choice. Moreover, whether it is based on Cosine similarity or Pearson correlation similarity, the performance evaluation results range from 55% to 75%. This indicates that the KNN-based music recommendation system can accurately recommend Chinese music matching users' preferences, effectively identify related music, and alleviate the data sparsity problem.

REFERENCE

Anuzaj, Y. and Luma, A. 2022. Cosine Similarity – A Computing Approach to Match Similarity Between Higher Education Programs and Job Market Demands Based on Maximum Number of Common Words. *International Journal of Emerging Technologies in Learning (iJET)*, 17(12): 258-268.

Basiński, K., Zdun-Ryżewska, A. & Greenberg, D.M. 2021. Preferred musical attribute dimensions underlie individual differences in music-induced analgesia. Sci Rep, 11, 8622.

Claussen, J. & Peukert, C. (2019). Obtaining Data from the Internet: A Guide to Data Crawling in Management Research. SSRN Electronic Journal, 1-34.

Dahdouh, K., Dakkak, A., Oughdir, L. & Ibriz, A. 2019. Large-scale e-learning recommender system based on Spark and Hadoop. Journal of Big Data, 6(2): 1-23.

Ezeh, A. 2023. Developing Machine Learningbased Recommender System on Movie Genres Using KNN. Stockholm University, department of computer and system sciences.

Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A. & Kashef, R. 2020. Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. Appl. Sci, 10(21), 7748.

Foody, G.M. 2023. Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. PLoS ONE, 18(10): e0291908.

Gunawardana, A. and Shani, G. 2009. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. Journal of Machine Learning Research, 10 (2009), 2935-2962.

He, W. 2021. Interior Design Scheme Recommendation Method Based on Improved Collaborative Filtering Algorithm. Wireless communications and mobile computing, 2021(1), 3834550.

Jiang, H.& Chaudhuri, A. 2023. A new method to improve the efficiency and accuracy of incremental singular value decomposition. Electronic Journal of Linear Algebra,39 (2023): 355-378.

Kuanr, M. and Mohapatra, P. 2021. Recent Challenges in Recommender Systems: A Survey. Progress in Advanced Computing and Intelligent Engineering, 353-365.

Kundu, D.S. 2020. A Survey of Music Recommendation Systems with a Proposed Music Recommendation System. Emerging Technology in Modelling and Graphics. 279-285.

Liu, Y. & Zhang, W. 2023. Design and simulation of precision marketing recommendation system based on the NSSVD++ algorithm. Neural Comput & Applic.

Longo, C. 2018. Evaluation Metrics for Recommender Systems. https://towardsdatascience.com/evaluation-metrics-for [May 18th, 2024]

McCain, A. 2023. 30+ HARMONIOUS MUSIC INDUSTRY STATISTICS [2023]: WORTH, DATA, AND REVENUE. https://www.zippia.com/advice/music-industry-statistics/ [June 16th, 2023)

Nallamala, S.H., Bajjuri, U.R. and Anandarao, S., DurgaPrasad, D. & Mishra, P. 2020. A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems, IOP Conf. Series: Materials Science and Engineering, 981(2020), 1-7.

Richter, F. 2023. Streaming Drives Global Music Industry Resurgence. https://www.statista.com/chart/4713/global-recorded-music-industry-revenues [June 16th, 2023].

Wu, M. & Lu, J. (2025) Design and application of a music recommendation system based on user behavior and feature recognition. systems and soft computing, 7(2025),1-14.

Dong, Y. (2023) Music Recommendation System Based on Machine Learning. Highlights in Science Engineering and Technology, 47,176-182.

Roy, D. and Dutta, M. 2022. A systematic review and research perspective on recommender systems. Journal of Big Data, 9(59), 1-36.

Schedl, M. 2019. Deep Learning in Music Recommendation Systems. Sec. Mathematics of Computation and Data Science, 5(44), 1-9.

Shinde, S.B. & Potey, M.A. 2016. Research Paper Recommender System Evaluation Using Coverage. International Research Journal of Engineering and Technology (IRJET), 3(6), 1678-1683.

Silber, J. 2019. Music Recommendation Algorithms: Discovering Weekly or Discovering Weakly? Muhlenberg College. https://osf.io/6nqyf/download [June 16th, 2023].

Tegene A, Liu Q, Gan Y, Dai T, Leka H, Ayenew M. 2023. Deep Learning and Embedding Based Latent Factor Model for Collaborative Recommender Systems. Applied

Sciences,13(2):726.

Thomas, B. & John, A.K. 2021. Machine Learning Techniques for Recommender Systems – A Comparative Case Analysis. IOP Conf. Series: Materials Science and Engineering, 1085 (2021) 012011.

Visa, G.P. and Salembier, P. 2014. Precision-Recall-Classification Evaluation Framework: Application to Depth Estimation on Single Images. Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, 8689. 648-662.

Yılmaz, T., Scheffler, T. 2023. Song authorship attribution: a lyrics and rhyme based approach. Int J Digit Humanities, 5, 29–44.

Yu, H. 2020. Latent Factor Model for Book Recommendation System ---Taking Douban as an Example. Advances in Economics, Business and Management Research,116(1): 30-35.

Zhang, Z.Y. 2012. Nonnegative Matrix Factorization: Models, Algorithms and Applications. Data Mining: Foundations and Intelligent Paradigms,24. 99-134.

APPENDIX A

Syntax code

```
for Each unrated item i by the current user do
sim     = 0, ratSimTotal = 0
for Each rated item j by the current user do
rating = The current user's rating for item j
Get the ratings of users who rated both i and j, construct two column vectors
similarity = Similarity of two column vectors
ratSimTotal += similarity * rating
simTotal += similarity
end for
The rating prediction for unrated item i is ratSimTotal / simTotal
end for
Sort ratings and recommend to current user from high to low
```

APPENDIX B

TABLE 2. Music scoring rules

Items	Score
Like once	+4
Swipe across once	-8
Same tags	+2
Tag exists but is not the same	-2
Score for a single review	+4
Listened score	+1
Download score	+3

TABLE 3. Music score function

User behavior score	= Number of likes * Single score for likes + Number of swipes * Single score of swipes + Number of listens * Single score of listens + Number of downloads * Single score of downloads
Tag score	= When the song tag and user preference tag are the same, score 2 points =If the song tag and the user preference tag are different, deducted 2points
Popularity score	= Number of comments on the song/Maximum number of comments on the song * Score for a single comment
Total music score	=User behavior score + Tag score + Popularity score

APPENDIX C

TABLE 4. Sample display of Music score result

user_id	item_id	user_id's rating of item_id	timestamp
2	59867	10	1556966041
2	62372	10	1556966041
2	63518	10	1556966041
2	63787	10	1556966041
2	64269	10	1556966041
1	59867	9	1556966042
1	63518	11	1556966042
1	62372	-4.9784	1558152261.8956
1	62372	-4.9784	1558152544.7204
1	62372	-4.9784	1558152639.7094
1	62372	-8.9784	1558152768.1438
1	375180	2.0086	1558153742.7879
1	63787	2.1492	1558157591.3199
1	63787	-1.8508	1558157608.2450
1	178895	2.007	1558157614.1213
1	178895	3.007	1558157711.9966
1	64269	3.0117	1558383675.5071
1	64634	4.2654	1558418269.9617
1	64634	5.2654	1558418271.8304
1	115569	-0.8161	1558429752.6070
1	115569	2.1839	1558429772.8099
1	115569	2.1839	1558429793.5020
1	385973	1.0388	1558431283.5407
1	385973	1.0388	1558433988.9793
1	385973	1.0388	1558434210.5846
1	385973	1.0388	1558434233.7292
1	385973	1.0388	1558434356.2067
1	385973	1.0388	1558434667.6461
1	385973	1.0388	1558435598.9696

Comments	Response
(1) The small and biased user sample (only 10 friends of the author) undermines the reliability of the experimental results.	The goal of this study is to experiment and compare the performance of different algorithms on this Chinese dataset, that is, to evaluate the results. Adding 10 users simply allows each feature to have a corresponding experimental performance. As you can see, the experimental results cover seven interest tags.
(2) Methodological decisions, such as the scoring scheme,are insufficiently justified.	Rating systems are widely used in recommendation algorithm research. The general details focus on user behavior and the fit between content and preferences. To strengthen the validation, I cite the views of Wu & Lu (2025) and Dong (2023). DOI: 10.1016/j.sasc.2025.200274; DOI: 10.54097/hset.v47i.8198
(3) The language and structure of the paper require substantial revision to meet academic standards. Furthermore, key figures referenced in the text are missing, and the interpretation of results is often overstated.	The wording and grammar of the entire article have been professionally revised. The experimental results charts have been exported and replaced with visual images based on the code results.
(4) I encourage the authors to expand the empirical study with a larger and more diverse user base, refine the scoring model with empirical or theoretical support, and significantly revise the manuscript for clarity and academic tone. The paper has potential, but it requires considerable work to reach publishable quality.	The database used in this paper is first-hand data from the crawler, containing over 200,000 Chinese songs. In terms of data scale, it has reached academic value. The reference paper, however, only has a data size of 1,000: https://doi.org/10.3390/app14020855. Furthermore, this is an algorithmic model study, not a systematic one. The experimental users represent only the performance of each feature label, and the average performance of different algorithms in this recommendation model is used to evaluate the differences. Therefore, the number of users is not crucial for this study.