





Personalized Recommendation Algorithm in Computer-Aided Music Instruction System

Chun Zhong¹  and Zhengzhao Qiu² 

¹Sports, Art, and Labor Education Center, Zhejiang Shuren University, Hangzhou, Zhejiang 310015, China, chunzhong3@163.com

²Xianlin Middle School, Yuhang District, Hangzhou, Hangzhou, Zhejiang 311122, China, qiuzhengzhao1@163.com

Corresponding author: Chun Zhong, chunzhong3@163.com

Abstract. This article aims to explore and enhance the personalized recommendation algorithm within the Computer-Aided Music Instruction System (CAMIS), ultimately improving user experience and learning outcomes. To accomplish this, we delve into crucial steps, including user portrait construction, music content feature extraction, and the selection and refinement of the recommendation algorithm model. This article provides a personalized discussion on the processing techniques of music signals based on user interest behaviour algorithms. In terms of innovative method behaviour, it has conducted the capture and analysis of music signals. In the process of comprehensively understanding users' needs and preferences, this article uses music signal processing technology to personalize the attributes of music signals. The experimental results show that the algorithm proposed in this paper achieves an accuracy of 96.35%. The diversity and accuracy of music education have significant advantages in terms of algorithm superiority. In the personalized algorithm performance recommendation process, this article further innovates the algorithm's performance in music education.

Keywords: Computer-Aided; Music Teaching System; Personalized Recommendation; User Behavior

DOI: <https://doi.org/10.14733/cadaps.2025.S4.224-236>

1 INTRODUCTION

The pitch features reflect the range of the instrument and the fundamental frequency of the notes. Using the idea of multi-fundamental frequency estimation, a custom parameter filter group is used to extract initial audio features and feed them into a convolutional network to extract pitch features [1]. Under copyright permission, obtain the original audio of real-world performances and corresponding MIDI digital scores on professional platforms. Manually annotating the presence of instruments at each time frame in multi-instrument music is a time-consuming and labour-intensive process [2]. Some scholars have explored effective features for multi-instrument recognition in polyphonic music

and designed extraction methods. To this end, it drew on these frame-level dataset construction methods and further expanded the data volume by building its dataset. Compared some existing sequence alignment algorithms [3]. Then, a dynamic time-warping algorithm based on dynamic programming is used to align the label annotations on the music score with the audio; To ensure accuracy, allow personnel with a certain music background to perform calibration [4]. The dataset with frame-level instrument labels has only appeared in recent years. These datasets only contain a maximum of over 200 songs. Therefore, most publicly available datasets only provide instrument labels at the clip level, marking which instruments exist throughout the entire audio clip that may last for several seconds [5]. This clip-level label does not specify the presence of an instrument for each short frame. The constant Q transform is a special wavelet transform that has been improved and is beneficial for music analysis. It can reflect the energy distribution of each pitch. We use an improved fast calculation method to extract the constant transform matrix of the audio. In multi-instrument recognition in polyphonic music, general time-frequency features may not achieve good recognition results [6]. Therefore, we selected pitch features and a constant Q transformation matrix as input features for the model. The benchmark model proves the effectiveness of pitch features in instrument recognition [7]. The attention mechanism has been widely applied in computer vision, and some scholars have applied it to the "auditory attention" of music signals. The combination of these two features can effectively capture the harmonic structure of music signals, which is reflected in the timbre of the instrument in music. At present, no work has been found to associate pitch with timbre in instrument recognition. The two-level classification model first performs a rough classification of instrument families and then performs a detailed classification of a specific instrument in the corresponding instrument family. We processed the extracted features and constructed three classification models, namely the baseline model, the attention network-based classification model, and the two-level classification model. Hierarchical recognition conforms to basic cognitive logic and also alleviates the unavoidable class imbalance problem in music signals [8]. In addition, the algorithm can predict potential challenges that may arise during the performance, formulate response strategies in advance, and ensure the quality of the performance. Based on this data, algorithms can intelligently adjust rehearsal plans, prioritize weak links, and thus improve rehearsal efficiency [9].

Personalized recommendation algorithms can intelligently recommend suitable composers and their works based on learners' interests, preferences, and learning progress, rather than relying solely on traditional designated composer lists or statistical-based automatic recommendation systems. Some scholars have delved into the diverse information visualization techniques applied to classical composer databases, which not only reveal the macro music world's context during the period of shared practice but also finely depict the micro details of the classical music field in the 20th century [10]. Accurately locating the relative positions of each composer on the map while ensuring the accuracy of distances between paired composers provides a novel perspective for understanding the complex relationships in classical music history. By mining and analyzing data on the personal music influence among composers, this article successfully constructed a network diagram that intuitively displays the interrelationships and influences between music compositions. Furthermore, the study innovatively integrated style influence data with the "ecological" information of composers, constructed similarity/distance matrices between composers, and utilized multidimensional scaling analysis techniques [11]. On this basis, some scholars have also introduced personalized recommendation algorithms in auxiliary music teaching systems, combining this cutting-edge technology with visual analysis of classical music, and expanding the boundaries of music education experiments. This method not only enhances the personalization and interactivity of music exploration but also encourages learners to actively discover the inner charm of composers and their music. Breaking the limitations of traditional rote learning promotes a more autonomous and in-depth music learning experience. By combining information visualization and personalized recommendation algorithms, a new path has been provided for the field of music education, aimed at stimulating learners' broad interest and profound understanding of classical music heritage.

The objective of this study is to delve deeply into personalized recommendation algorithms within CAMIS and attempt to integrate CNN technology. The key questions this study aims to address are:

Firstly, how to construct precise user profiles that fully capture individual student needs? Secondly, how can CNN technology extract features from music content to enhance the precision of the recommendation algorithm? Lastly, how to refine the recommendation algorithm model for more efficient and personalized music recommendations? Resolving these issues is vital for enhancing CAMIS performance and offering valuable insights into the development of music education technology. This research introduces the following innovations:

(1) In this article, an advanced personalized recommendation algorithm is applied to CAMIS, to realize the accurate matching between students' needs and music learning resources.

(2) CNN technology is introduced to extract the deep features of music content and improve the efficiency of the recommendation algorithm.

(3) A user portrait construction method for music learners is proposed, which comprehensively considers multi-dimensional information such as students' learning behaviour, interest preferences and music literacy, to capture students' individualized needs more comprehensively.

(4) A recommendation algorithm model suitable for music teaching is designed and optimized, which combines traditional recommendation algorithms with deep learning technology to achieve more efficient and individualized music recommendation.

(5) The real-time feedback mechanism is introduced into the recommendation system, which can adjust the recommendation strategy in time according to the student's learning feedback and realize a more dynamic and personalized recommendation service.

In the specific research process, firstly, the present situation of CAMIS is reviewed, and its advantages and disadvantages in practical application are analyzed. Then, the basic principles and common methods of personalized recommendation algorithms are discussed. On this basis, we will study the application of CNN in music feature extraction and explore how to combine it with personalized recommendation algorithms to build a more efficient and accurate recommendation system.

2 RELATED WORK

Audio codecs play a crucial role in the processing of digital audio streams, as they achieve efficient compression and decompression by removing information that is difficult for human auditory perception. However, at high compression rates, they may also cause significant damage. In this context, Melchiorre et al. [12] explored the potential of random generators in the generative adversarial network (GAN) architecture for recovering music audio signals. However, in the field of music, especially for the recovery of heavily compressed audio signals, research is relatively scarce and complex, as there is often no unique and clear answer to restoring the original signal. The spectral distribution matrix generated by the constant Q transform contains certain noise, including the noise generated by the error of the constant Q transform processing and the original noise of the music signal. The double tube here provides a frequency reference point where all instruments, including the oboe, still exhibit this pitch shift phenomenon. The purpose of adding oboe tuning is to make the pitch shift of all instruments the same. In computer vision, data augmentation is often achieved by adding noise to images to improve the robustness of network models. These raw noises may include Gaussian white noise and overtone noise with very small absolute amplitudes. These noises can cause the pitch frequency of the actual performance to deviate from standard instruments, and experienced orchestras will use oboes to sing note A * to tune other instruments. In addition, when recording raw audio in the real world, there is often noise caused by instrument tuning errors, as well as noise caused by deviations in instrument sound quality due to the physical environment and performance conditions. Even if there is a pitch deviation, all instruments in the entire orchestra still produce harmonious sounds. However, Pan et al. [13] do not intend to artificially process the spectral distribution matrix of the constant Q transformation. I hope that deep network structures can 'learn' parameters that can extract effective intermediate features from noise interference during parameter training. Therefore, the noise on the spectrogram may play a role in

enriching the diversity of data and bringing it closer to the performance results of the real world. In the auxiliary music teaching system, personalized recommendation algorithms intelligently recommend suitable music works and practice materials by analyzing learners' preferences, skill levels, and learning progress. However, these suggestions are often limited by the quality of audio resources, especially when the audio resources are highly compressed and the sound quality deteriorates, which may affect the learner's experience and learning effectiveness. For beginners, the system can generate clearer and easier-to-understand audio versions. For advanced learners, more refined and expressive restorative effects can be provided to stimulate their deeper exploration and understanding of music.

Given the limitations of collaborative filtering algorithms in accurately matching user preferences, Pei and Wang [14] proposed an innovative personalized music resource recommendation algorithm based on category similarity. And specifically explored its potential applications in auxiliary music teaching systems. This algorithm not only optimizes the accuracy of music recommendation but also further promotes the personalization and effectiveness of music learning. These parameters not only cover users' direct preferences, but also involve their growth trajectory, skill level, and interest changes in the music learning process. Based on these parameters, some scholars have established personalized preference judgment models that can dynamically adjust to adapt to users' constantly changing preferences as the learning process progresses. By analyzing the homomorphic reliability of personalized scores, Peng and Li [15] constructed a joint parameter-matching model for music resources, which can accurately match the current learning needs of users with music resources, ensuring the pertinence and effectiveness of recommended content. By constructing and analyzing knowledge graphs, we can gain a deeper understanding of users' music preference type parameters. In the auxiliary music teaching system, use feature registration algorithms to explore the personalized features of music resources in depth. These characteristics not only include the basic attributes of music (such as style, rhythm, melody, etc.) but also the comprehensive educational significance (such as difficulty level, skill point coverage, etc.). These parameters not only help the system understand the uniqueness of each piece of music but also provide insights into the associations and differences in educational value between different pieces of music. In addition, Tian [16] utilized category similarity features to perform adaptive statistical analysis on music resources, extracting more refined personalized feature parameters. The simulation experiment results show that in the optimal state, the algorithm achieved a minimum satisfaction rate of up to 92.7%, the resource holding level remained stable at over 92%, and the recommendation accuracy reached an astonishing 98.9%. Based on this, the system can generate personalized learning paths and recommendation lists, guiding users to learn music according to the most suitable rhythm and method for themselves. These data not only validate the effectiveness and practicality of the algorithm but also demonstrate its enormous potential for application in auxiliary music teaching systems.

The popularity of music streaming services has profoundly changed the way people learn music, especially with the development of auxiliary music teaching systems, which have pushed personalized recommendation technology to new heights. Given this, Velankar and Kulkarni [17] proposed an innovative music recommendation method that cleverly integrates reinforcement learning mechanisms into personalized recommendation algorithms for auxiliary music teaching systems. Specifically, a recommendation system based on the Q-learning model has been implemented, which dynamically adjusts the recommendation strategy by continuously evaluating the accumulated rewards of users playing and liking similar songs during the session, achieving continuous optimization of the incremental reinforcement learning algorithm. In the context of auxiliary music teaching systems, this personalized recommendation algorithm combined with reinforcement learning has demonstrated unique advantages. Although traditional collaboration and hybrid filtering techniques can effectively recommend popular songs, they often overlook in-depth exploration of music content, dynamic changes in user tastes, and novelty of recommendation results. This is particularly important in music teaching environments that pursue personalized and differentiated learning. This process not only considers users' immediate feedback but also constructs a more refined and dynamic user profile through a comprehensive analysis of implicit and explicit

feedback, accurately capturing subtle changes in users' music tastes. The experimental results show that compared with existing music teaching applications, the system proposed by Xu et al. [18] using the reinforcement learning personalized recommendation algorithm in this paper increased the average interaction time of users by 35%.

3 CAMIS OVERVIEW

This article mainly studies the music source feature extraction and separation algorithm based on deep neural networks, with a focus on the music source separation algorithm based on feature masking. The main research content and work summary of this article proposes a music source separation algorithm based on skip attention mechanism and amplitude spectrum features. Improved Unet using skip attention mechanism and proposed a feature extraction module for amplitude spectrum information to obtain multi-scale feature information. Design and implement an automatic music source separation system. The system can be called the music source separation algorithm, which is designed in this article to separate the music signals uploaded by users automatically. Design a self-attention convolutional block using a self-attention mechanism and design an encoder network based on a self-attention mechanism to fuse the phase spectrum feature information of music signals into the amplitude spectrum. Further integration with the amplitude spectrum information of the music is needed to improve the performance of music source separation while modifying the training method of the model and introducing a time-domain loss function. And display the evaluation indicators of the separated music sources, or users can select music from the music source dataset provided by the system for separate display. By researching current music source separation algorithms, we aim to improve the performance of music source separation algorithms based on deep neural networks. Explored the impact of the masking type and loss function used in the model on its training and separation performance and selected the optimal combination of masking type and loss function. By correcting the phase spectrum of the original music signal and using the feature extraction module designed in this paper to extract phase spectrum feature information.

<i>Feature/Application</i>	<i>Description</i>	<i>Advantages</i>	<i>Disadvantages</i>
System Architecture	Front-end user interface, back-end server, database	Separated design for easy maintenance and scalability	Requires significant technical investment and operational costs
User Management	User registration, login, personal information management	Convenient user management enhances system security	The risk of user information leakage requires strict data protection measures
Music Resource Management	Upload, download, categorization, and retrieval of musical works	Provides a wealth of music resources to meet diverse user needs	Resource copyright issues require legal authorization and usage
Learning Progress Tracking	Records user's learning history, progress, and achievements	Helps users understand their learning situation and adjust learning strategies accordingly	Data accuracy issues require effective data validation and cleaning mechanisms
Personalized recommendation	Recommends musical works and learning resources based on user behaviour and data	Improves learning efficiency, satisfies user's individualized needs	Accuracy challenges in recommendation algorithms require continuous optimization and updating

Interaction and Feedback	Provides an interactive interface, collects user feedback and suggestions	Enhances user engagement, used to optimize system performance	Efficiency challenges in processing user feedback, require establishing an effective feedback-handling mechanism
Technical Challenges	Data management, algorithm optimization, stability, and compatibility	Requires continuous technological innovation and investment	Rapid technological advancements necessitate ongoing updates and upgrades

Table 1: CAMIS characteristics.

Table 1 shows the characteristics of CAMIS. Constructing user portraits necessitates the collection and analysis of diverse data from users during the learning process, including learning progress, duration, browsing history, and collection records, to comprehensively grasp users' learning habits and interest preferences. Music, as a multifaceted art form, encompasses rich characteristic information like melody, rhythm, harmony, and timbre. In the process of extracting music content features, musical works are transformed into a series of digital feature vectors, which precisely depict the attributes of the music, such as style, emotion, and genre. The data mining architecture of CAMIS is illustrated in Figure 1.



Figure 1: Data mining architecture.

Descriptive information (music name, introduction, lyrics) and the audio itself are representative of a piece of music's unique attributes. Notably, the audio feature stands out as the most distinctive and effective information representation. By leveraging this feature, this article aims to distinguish between various music pieces with the utmost precision. Sound signals are converted into image

representations to enhance feature extraction using CNN. To achieve this, Hertz frequency f is mapped to Mel frequency $mel f$.

$$mel f = 2595 \times \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

Using the Mel frequency derived from the Mel mapping formula, human hearing linearly perceives Mel frequency. Specifically, a doubling of the Mel frequency in audio corresponds to a doubling of the perceived tone by human ears. The Mel spectrum leverages human auditory perception characteristics to produce a spectrum, and incorporating this characteristic into neural networks notably boosts the performance of automatic music classification.

An automatic encoder learns the hidden layer representation of input data by reconstructing it through two processes: encoding and decoding. Essentially, an automatic encoder encompasses these two primary processes:

The encoding process involves transforming the input layer into the hidden layer:

$$h = f x = \sigma W'x + b \quad (2)$$

The decoding process entails transforming the hidden layer into the output layer:

$$\tilde{x} = g h = \sigma W^T h + b \quad (3)$$

W, b signifies the weight and bias term of the feature, whereas $\sigma \cdot$ indicates the activation function.

The goal of the automatic encoder is to reconstruct the input layer data at the output layer; thus, its loss function incorporates diverse data forms.

When selecting the recommendation algorithm model, it is necessary to comprehensively consider the characteristics of the data set, the performance requirements of the recommendation system, and the actual needs of users. For example, if the data set is sparse, you can choose a content-based recommendation algorithm, which can make use of the feature information of music works to make recommendations, regardless of the sparsity of the data set. After determining the recommendation algorithm model, it is necessary to optimize and adjust the model to improve the accuracy and efficiency of the recommendation. As shown in Figure 2.

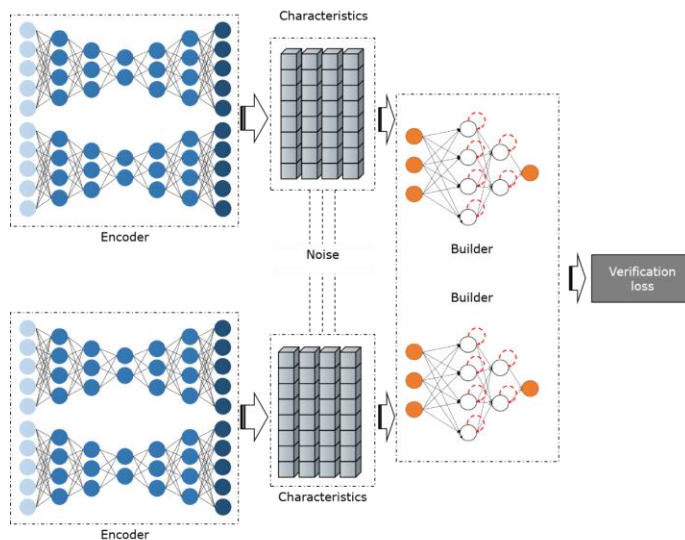


Figure 2: A network model for music feature analysis.

$$f_{x_f} = \text{Max } p_i \quad (4)$$

Here, f_{x_f} denotes the pitch feature vector of the current musical work.

Based on the records of students' music listening habits, we calculate the frequency of each music piece listened to by different students. Subsequently, we determine the playing frequency of the song v that the student u has listened to:

$$f_{u,v} = \frac{p_{u,v}}{p_u} \quad (5)$$

Here, $p_{u,v}$ represents the count of times the student u plays music v , while p_u denotes the total number of instances where the student u listens to music.

When analyzing the emotion of a new text, the first step is to use advanced word segmentation tools to segment the text carefully and mark each word accurately to ensure that each word is accompanied by symbols representing its part of speech. Then, these processed words and phrases are matched with the established phrase model base, which contains rich phrases and their corresponding emotional tendency values. Through this matching process, the emotional inclination of each phrase can be effectively obtained, and then the emotional inclination of the whole text can be calculated. Finally, based on the calculation result of this emotional tendency, the emotional classification result of the text is obtained, which provides strong support for the subsequent text processing and application. The flow chart is shown in Figure 3.

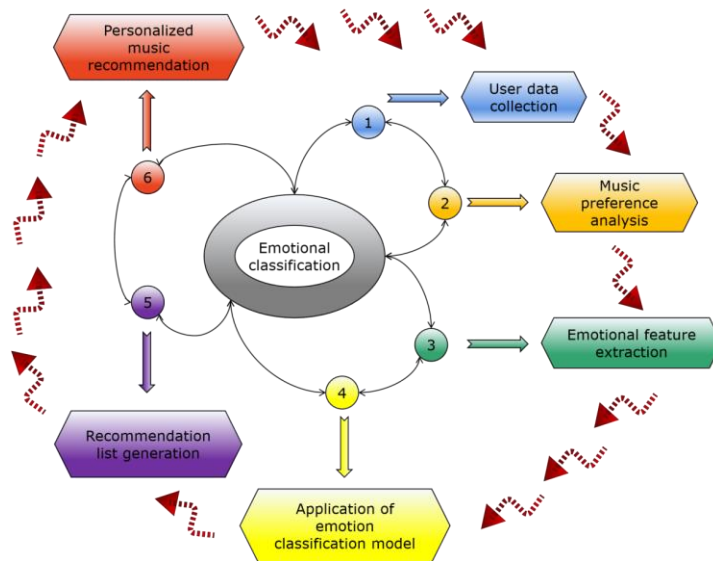


Figure 3: Emotion classification process.

During the data collection stage, user learning history, behaviour data, and pertinent information about music works are gathered. In the preprocessing stage, data is cleaned, duplicates are removed, and formatting is applied to ensure data accuracy and consistency. The feature extraction stage necessitates the use of music analysis and machine learning technologies to extract feature information from music works and construct user portraits. In the model training stage, an appropriate recommendation algorithm model is selected, and training data is used to refine the model parameters. Ultimately, in the recommendation generation stage, we can offer users the most suitable music works and learning resources tailored to their individual needs.

To enhance the accuracy of students' evaluation classification, it is imperative to extract precise emotional features leveraging contextual dependency extensively. The simplified GRU model can be formulated as:

$$h_t = GRU(x_t, h_{t-1}) \quad (6)$$

Upon inputting the context $X = x_1, x_2, \dots, x_n$ of the comment, the hidden layer state \vec{h}_t is outputted to the GRU:

$$\vec{h}_t = GRU(x_t, \vec{h}_{t-1}) \quad (7)$$

By utilizing the LDA (Latent Dirichlet Allocation) theme model, n themes are derived, with m words being closely associated with each theme. Define w_{ij} as the j word about the i theme, which can be represented in matrix form as follows:

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix} \quad (8)$$

Through iterative attention to memory and information extraction, the output vector y_i^t from the final calculation layer forms the ultimate emotional feature y_o . This feature can be viewed as a conditional probability and is subsequently input into a SoftMax classifier for emotional classification purposes:

$$p_c = \frac{\exp(W_s \cdot y_o + b_s)}{\sum_{c=1}^C \exp(W_c \cdot y_o + b_c)} \quad (9)$$

Here, C denotes the number of categories, while p_c representing the probability of predicting a particular category c .

4 EXPERIMENTAL ANALYSIS AND DISCUSSION

4.1 Experimental Design

The core objective of the experiment is to validate the effectiveness, accuracy, and practicability of the algorithm while exploring the impact of various factors on recommendation performance. It aims to determine whether the algorithm can precisely capture users' individualized needs and interest preferences, assess its performance across different user groups, music types, and periods, and investigate the influence of incorporating interest similarity information on user relationship prediction. To achieve this, extensive user behaviour data from a music teaching system are collected, including learning history, browsing records, collection records, etc., along with detailed information on music works such as style, genre, composer, etc., forming the basis of our experimental dataset. In the data preprocessing stage, the data is cleaned, duplicates are removed, and formatting is applied to ensure accuracy and consistency. Additionally, music work features are extracted and transformed into digital feature vectors for subsequent algorithm processing.

4.2 Experimental Results

When the similarity between users decreases, the accuracy of recommendations also shows a significant downward trend. This is because the algorithm relies more on the similarity information between users during recommendations, and when the similarity decreases, the algorithm finds it

difficult to find enough similar users to support accurate recommendations. This experimental result validates our assumption about the importance of user similarity in algorithm design. Figure 4 shows the impact of different users on accuracy.

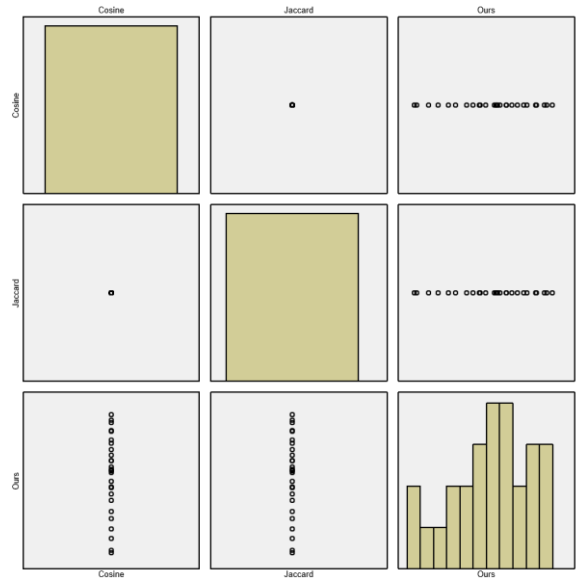


Figure 4: Influence of dissimilar users on accuracy.

Furthermore, the influence of dissimilar users on recommendation diversity is analyzed. Figure 5 shows the experimental results of this relationship. When the similarity between users decreases, the diversity of recommendations shows an upward trend. This is because when it is difficult for the algorithm to find similar users, it will try to recommend more different types of music works to hit the potential interest points of users. This experimental result also verifies the hypothesis of the relationship between recommendation diversity and user similarity when the algorithm is designed.

In music teaching systems, the similarity of interests among users (i.e., learners) is not only the foundation for them to establish deep learning partnerships but also a key factor in improving the quality of personalized recommendations. Interest similarity can be reflected in multiple dimensions, such as preferred music genres, composers, music periods, instrument preferences, and learning progress. When the system can accurately capture and quantify these similarities, it can more effectively promote communication and cooperation among learners while optimizing learning paths and recommending content. As shown in Figure 6, when the music teaching system introduces interest similarity information, the prediction accuracy of user relationships is significantly improved. This result directly reflects the high efficiency of interest similarity in identifying potential learning partners and predicting learning interaction behaviour. Specifically, the system can match learners with similar learning paths and interest preferences based on shared music interests and learning goals, promoting deeper learning communication and cooperation.

Finally, the algorithm proposed in this article underwent comprehensive testing and was compared with other algorithms. Figure 7 displays the results regarding algorithm recommendation accuracy. The proposed algorithm achieved a recommendation accuracy of 96.35%, significantly surpassing the contrast algorithm's accuracy of 81.35%. This remarkable improvement in accuracy verifies the superior performance of this algorithm in personalized recommendations.

After the user reviews the separation results, they can save the results. After clicking save, the system will call the API interface provided by Alibaba Cloud Object Storage OSS to upload and save the music source files and spectrograms, without using the user ID to distinguish between the user's

files. Users can click the upload music file button on the separation page when using their music data for music source separation, provided that they have logged in.

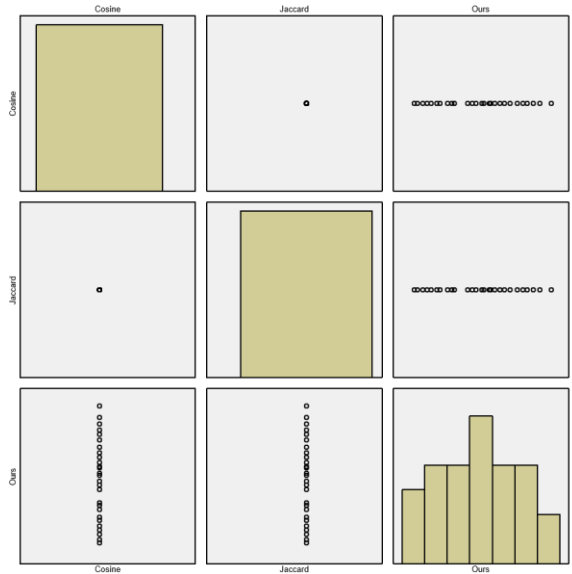


Figure 5: Influence of dissimilar users on diversity.

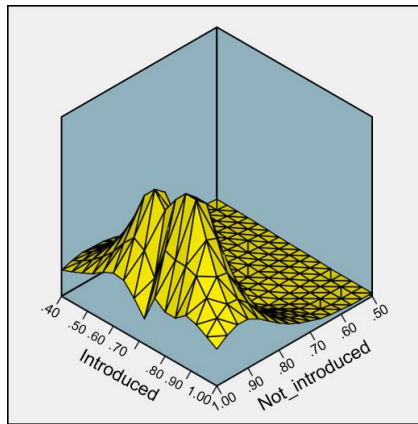


Figure 6: Recommendation accuracy in music teaching system

After the separation is completed, the separated music source files and spectrograms are obtained separately. The backend packages these data and returns them to the frontend browser, which parses and renders the data onto the page. After the upload is completed, the user needs to configure the parameters for separating the music source, including the separated music source, the length of the separated audio file, and the separation algorithm to be used.

Then select your own target music file, confirm it, and upload it. After determining the parameters, the browser submits the separated parameters using a form, and the backend first finds the user-uploaded audio file in the temporary folder according to the settings in the parameters. After receiving an upload request, the time-series browser that performs music source separation will transcode the audio file and use Axios for POST request transmission to the backend server.

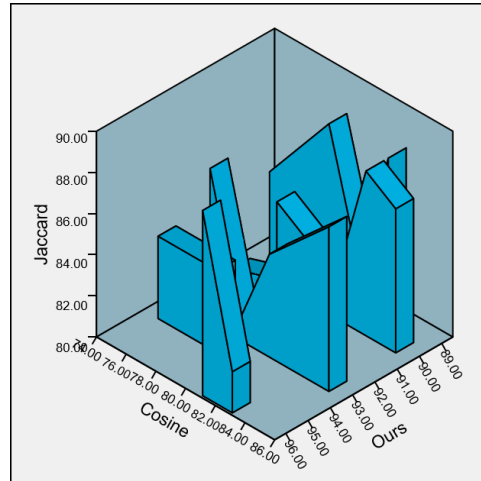


Figure 7: Algorithm recommendation accuracy.

Perform data preprocessing, and after preprocessing is completed, call the corresponding model to separate the audio data. The user's ID is stored in the first level directory of OSS, the upload date is in the second level directory, and the spectrogram file and music source audio file are entered using the suffix.

5 CONCLUSIONS

This article mainly studies the music source feature extraction and separation algorithm of the personalized recommendation algorithm in CAMIS. In response to some challenges encountered in current music source separation tasks, a corresponding neural network model is designed to obtain the music source masking contained in the amplitude spectrum of mixed music signals and achieve music source separation of mixed music. In addition, the algorithm performed well in comprehensive testing, with a recommendation accuracy of 96.35%, far higher than the comparison algorithms. The Transformer network designed based on a self-attention mechanism has achieved excellent performance in multiple different deep learning fields. Through research on traditional and deep neural network-based music source separation algorithms, it was found that data-driven deep neural network-based music source separation algorithms have more advantages than traditional music source separation algorithms. In addition to better separation performance, they also have better generalization. Previous masking-based music source separation algorithms have only focused on processing the amplitude spectrum information of mixed music, and there have been few studies that combine phase spectrum information as well. The phase spectrum of mixed music signals also contains information about the amplitude spectrum, which can also be utilized.

Chun Zhong, <https://orcid.org/0009-0006-2112-4274>

Zhengzhao Qiu, <https://orcid.org/0009-0001-3717-1730>

REFERENCES

- [1] Abdul, M.-N.; Harun, S.-N.; Baharom, M.-K.; Kamaruddin, N.: Preferred learning styles for digital native and digital immigrant visitors in the Malaysian music museum, *Asian Journal of University Education*, 16(3), 2020, 234-246. <https://doi.org/10.24191/ajue.v16i3.11085>

- [2] Bishop, L.; Cancino, C.-C.; Goebel, W.: Moving to communicate, moving to interact: Patterns of body motion in musical duo performance, *Music Perception: An Interdisciplinary Journal*, 37(1), 2019, 1-25. <https://doi.org/10.1525/mp.2019.37.1.1>
- [3] Chen, W.; Yang, T.: A Recommendation system of individualized resource reliability for online teaching system under large-scale user access, *Mobile Networks and Applications*, 28(3), 2023, 983-994. <https://doi.org/10.1007/s11036-023-02194-8>
- [4] Dotov, D.; Bosnyak, D.; Trainor, L.-J.: Collective music listening: movement energy is enhanced by groove and visual social cues, *Quarterly Journal of Experimental Psychology*, 74(6), 2021, 1037-1053. <https://doi.org/10.1177/1747021821991793>
- [5] Georges, P.; Seckin, A.: Music information visualization and classical composers discovery: an application of network graphs, multidimensional scaling, and support vector machines, *Scientometrics*, 127(5), 2022, 2277-2311. <https://doi.org/10.1007/s11192-022-04331-8>
- [6] He, N.; Ferguson, S.: Music emotion recognition based on segment-level two-stage learning, *International Journal of Multimedia Information Retrieval*, 11(3), 2022, 383-394. <https://doi.org/10.1007/s13735-022-00230-z>
- [7] He, X.; Dong, F.: Vocal music teaching method using a fuzzy logic approach for musical performance evaluation, *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 45(6), 2023, 9289-9302. <https://doi.org/10.3233/JIFS-233020>
- [8] Klein, K.; Melnyk, V.; Voelckner, F.: Effects of background music on evaluations of visual images, *Psychology & Marketing*, 38(12), 2021, 2240-2246. <https://doi.org/10.1002/mar.21588>
- [9] Lattner, S.; Nistal, J.: Stochastic restoration of heavily compressed musical audio using generative adversarial networks, *Electronics*, 10(11), 2021, 1349. <https://doi.org/10.3390/electronics10111349>
- [10] Liu, J.: An automatic classification method for multiple music genres by integrating emotions and intelligent algorithms, *Applied Artificial Intelligence*, 37(1), 2023, 2211458. <https://doi.org/10.1080/08839514.2023.2211458>
- [11] Lopes, A.-M.; Tenreiro, M.-J.-A.: On the complexity analysis and visualization of musical information, *Entropy*, 21(7), 2019, 669. <https://doi.org/10.3390/e21070669>
- [12] Melchiorre, A.-B.; Penz, D.; Ganhör, C.; Lesota, O.; Fragoso, V.; Fritzl, F.; Schedl, M.: Emotion-aware music tower blocks (EmoMTB): an intelligent audiovisual interface for music discovery and recommendation, *International Journal of Multimedia Information Retrieval*, 12(1), 2023, 13. <https://doi.org/10.1007/s13735-023-00275-8>
- [13] Pan, F.; Zhang, L.; Ou, Y.; Zhang, X.: The audio-visual integration effect on music emotion: Behavioral and physiological evidence, *PLoS One*, 14(5), 2019, e0217040. <https://doi.org/10.1371/journal.pone.0217040>
- [14] Pei, Z.; Wang, Y.: Analysis of computer-aided teaching management system for music appreciation course based on network resources, *Computer-Aided Design and Applications*, 19(S1), 2021, 1-11. <https://doi.org/10.14733/cadaps.2022.S1.1-11>
- [15] Peng, L.; Li, D.: Personalised recommendation algorithm of music resources based on category similarity, *International Journal of Reasoning-based Intelligent Systems*, 15(3-4), 2023, 323-331. <https://doi.org/10.1504/IJRIS.2023.136369>
- [16] Tian, Y.: Multi-note intelligent fusion method of music based on artificial neural network, *International Journal of Arts and Technology*, 13(1), 2021, 1-17. <https://doi.org/10.1504/IJART.2021.115763>
- [17] Velankar, M.; Kulkarni, P.: Employing cumulative rewards based reinforcement machine learning for individualized music recommendation, *Multimedia Tools and Applications*, 83(16), 2024, 48007-48020. <https://doi.org/10.1007/s11042-023-17448-6>
- [18] Xu, Y.; Su, H.; Ma, G.; Liu, X.: A novel dual-modal emotion recognition algorithm with fusing hybrid features of audio signal and speech context, *Complex & Intelligent Systems*, 9(1), 2023, 951-963. <https://doi.org/10.1007/s40747-022-00841-3>