

Research Article

Improved Music Recommendation Algorithm for Deep Neural Network Based on Attention Mechanism

Xin He 

School of Fusion Media, Xinyang Agriculture and Forestry University, Henan, Xinyang 464000, China

Correspondence should be addressed to Xin He; 2008310030@xyafu.edu.cn

Received 11 April 2022; Accepted 23 May 2022; Published 24 June 2022

Academic Editor: Wen Zhou

Copyright © 2022 Xin He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increasingly close combination of the Internet and people's production and life, the total amount of global data and information also grows increasingly. In order to save users the time to find their favorite music among many music types, the music recommendation service arises at the historic moment and is widely concerned by scholars. Traditional music recommendation system based on the collaborative filtering algorithm has low recommendation accuracy, poor real-time performance, sparsity, system cold start, and so on. Moreover, the traditional music recommendation algorithm only simply uses user behavior characteristics and does not make good use of user history for listening to audio characteristics. In view of the above question, this section based on the attention mechanism of the deep neural network music recommendation algorithm, through the use of improved MFCC audio data preprocessing, the extracted audio combined with the user's own portrait features, through the AIN RNN network recommended list, by learning user history listening to songs, improves the model-recommended accuracy.

1. Introduction

With the increasingly close combination of the Internet and people's production and life, the total amount of global data and information has also grown increasingly. With the continuous development of digital multimedia technology, the music library has become more and more large, and the music resources are becoming more and more rich. In order to save users the time to find their favorite music in many music, music recommendation services emerge at the historic moment and are widely concerned by scholars. Music, as a kind of multimedia information, not only its own data volume is increasing but also the user demand is increasing, which puts forward higher requirements for the research of the music recommendation algorithm. In the music market, with the continuous development and application of digital multimedia technology, more and more music industries have turned to online music services. People can use Spotify, NetEase cloud music, Shrimp music, and other platforms. It is convenient to obtain music resources through online audition, online download, and other ways [1]. However, as the music library becomes larger and larger and the music

resources increase, it takes a lot of time and energy for users to find their favorite music. In the past, people could only search for music through the keywords such as music name, singer, and classification, and the search results not only take into account the differences of users but also would lead to the long tail phenomenon of music. The emergence and application of the music recommendation system can effectively solve this problem. The recommendation system can predict the users' behavior preferences according to the user behavior information and music data characteristics and actively push the music to the users that meets their tastes [2]. Of course, the growing amount of music data and user demand put forward higher requirements for the research of music recommendation algorithm. On the one hand, traditional recommendation methods such as cold start in collaborative filtering need to be solved, and the original recommendation algorithm needs to be upgraded. On the other hand, with the development of machine learning and deep learning, the continuous emergence of new computing technology helps to fully tap users' potential preferences and improve the performance of the recommendation system.

Based on this background, this paper based on the deep neural network, from the analysis of users, music characteristics, introduce attention mechanism, puts forward a personalized music recommendation algorithm. It not only can bring great convenience to music users but also each music software provider want to achieve goals, so it has important research significance and broad application prospects. The text is divided into 5 chapters. Chapter 1 introduces the research background and research necessity and the main methods and effects of the music recommendation at present; Chapter 3 models the music recommendation algorithm based on deep neural network and introduces the attention mechanism to realize the accurate recommendation of the algorithm. Chapter 4 mainly uses the designed recommendation algorithm to recommend music and evaluates the recommendation effect. Chapter 5 mainly summarizes the work of the full text and puts forward the imagination of the next work.

The long tail theory was put forward by Chris Anderson of the United States. He concluded that the new business model brought by the Internet is the long tail market. According to the August 2 theory, only 20% of products can become popular products, have good marketing, and be known by others. As the Internet brings the characteristics of product diversification, the remaining 80% of products can also be known through the Internet platform. Although they are niche products, the Internet provides a publicity channel for them to have direct contact with the public, so as to increase the market of niche products. A little makes a lot. On the whole, niche products sometimes generate more overall profits than popular products.

2. State of the Art

However, in the era of big data, the recommendation methods for massive music data are facing new challenges and opportunities. On the one hand, the processing of massive data becomes complicated and needs personalized user needs. The original recommendation algorithm needs to be upgraded accordingly; on the other hand, the problems existing in the original recommendation algorithm also need to be solved, such as the cold start problem of collaborative filtering recommendation.

The NetEase cloud divides music recommendation into three parts: private FM, daily song recommendation, and recommended song list. Private FM has low accuracy and high diversity. Diversity and high energy bring freshness to users. If you find a song you have never heard but especially like, it will bring a sense of surprise and mobilize users' positive emotions.

Due to low accuracy, it is likely that the new song will not be liked by users, so set the "delete" and "next" buttons on the playback interface of private FM to facilitate users to switch songs. Daily song recommendation has high accuracy and low diversity. The high accuracy makes the 20 songs recommended every day better meet users' tastes, but there is the problem of simplification of music types. Therefore, playlists are set to provide users with the right to browse and operate and make up for the disappointment of users caused

by simplification of tracks. In the accuracy and diversity of the recommended song list, the recommended song list is different from the other two personalized recommendation functions. The threshold of its accuracy and diversity is not only determined by the algorithm but also determined by its functional form. Firstly, the object-oriented function is divided into two categories: one is the user and the other is the UGC song list. The system labels the song list and the user, respectively, to improve the accuracy. Because the UGC song list is created by many users, the UGC song list has diversity. The combination of the two ensures the coexistence of accuracy and diversity.

With the powerful filtering and push function of the recommendation system, we can well solve the problem of poor search engine effect when users cannot clearly describe the needs. The recommendation system not only provides convenience to consumers but also has an impact on the decision of businesses. By analyzing user data, merchants provide services that are more in line with user preferences. At present, the recommendation system has been widely used in the products of the Internet giants, including music push, information retrieval, social networks, location services, and news push [3]. Personalized music recommendation is a very special field in the recommendation system. Burke proposed in ACM RecSys that there are many kinds of music. The cost of listening to a song is very low. A song only lasts for a few minutes, and the music is mostly used as background music, so users do not need to pay full attention to the music. Compared with books and movies, music has a high reuse rate, and many users will share their own music. To sum up, the music field is very suitable for using a personalized recommendation system for recommendation. In recent years, making music recommendations according to user preferences has become an integral part of online digital music services. For example, the foreign Tidal Review, Apple Music, Pandora, etc. [4], Domestic Douban FM, Kugou Music, Xiami Music, and others integrated into their own recommendation algorithms, greatly improve the engagement of users to the product, expand the product's voice in this field, and help users save a lot of time finding music. Spotify has added a music discovery feature that provides better recommendations, playlists, and music discovery. Like the more foreign well-known music websites, such as Last. The FM recommends the songs of common interest with the higher similarity between the users, which is judged by the user's listening list and collection list [5]. Pandora uses the basic characteristics of music, such as using labels and singers, to calculate the similarity between songs. Baccigalupo Claudio and Plaza Enric believe that the recommended results should be broader, so the more highly recommended songs are integrated into playlists to recommend to users, making users more satisfied with the recommended results [6].

A convolution matrix factorization model proposed by Guan et al. and Van et al. analyzed the hidden features of music and audio signals to solve the cold start problem of the system and achieved good results [6]. Jearanaitanakij has achieved good results by learning from user history preferences to build models for recommendations. However, because the model iterates the parameters in deep network

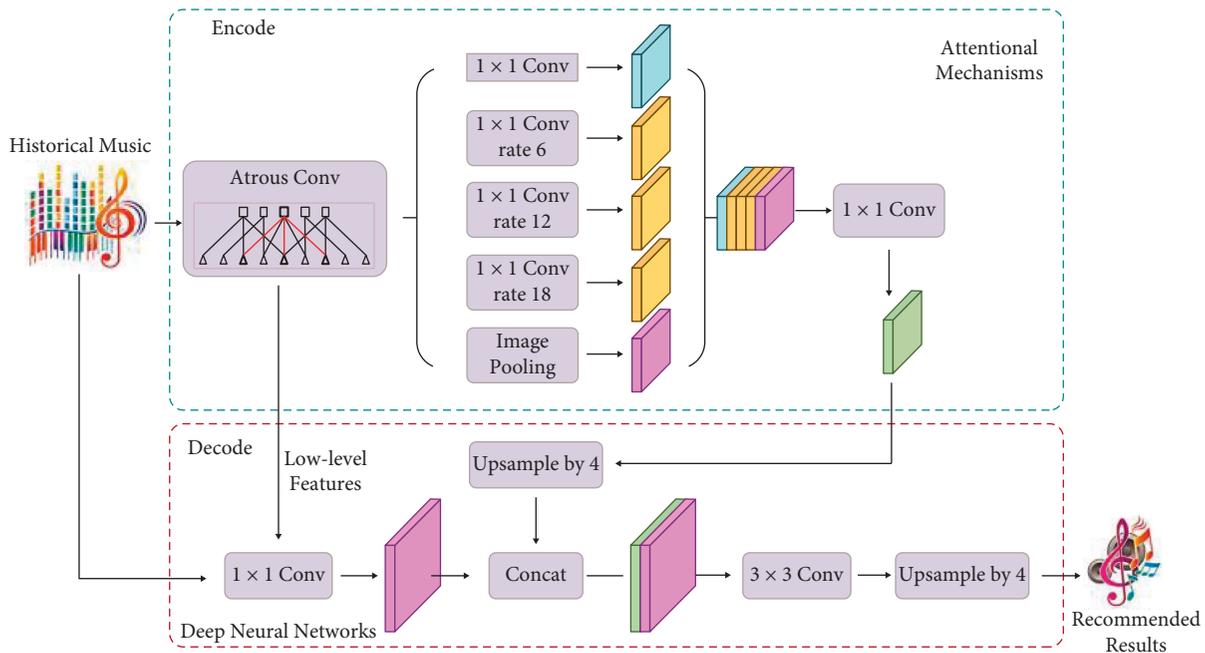


FIGURE 1: An improved framework of deep neural network music recommendation algorithm based on the attention mechanism.

training, the problem of parameter gradient disappearing or explosion occurs [7]. Jia extracted and segmented the long text features according to the language order, and the results obtained from the LSTM recommendation model were 4.1% higher than the RNN model [8]. Kowald et al. use the trust relationship in social networks to propose a method to determine the user influence degree and use the obtained influence degree to define a new collaborative filtering similarity index, giving greater weight to the users with greater influence [9]. Magron and Fevotte proposed a fusion model based on deep learning and collaborative filtering, using improved automatic encoder and convolutional neural network mining to climb the hidden features of music, combining deep learning model and collaborative filtering model, considering the music characteristics and user preferences, supervising the collaborative filtering process, and solving the scoring problem in the sparse matrix of prediction accuracy which is not high [10]. Zhou and Zhou proposed a simple effective convolutional neural network (MCRN) to learn the audio content features of music, specifically converting audio into a “spectrum map” through the Fourier transform. MCRN can effectively extract music content features from the spectrum and experimentally demonstrate that MCRN outperforms other models in music classification and recommendation accuracy [11]. Zhao et al. use matrix decomposition technology to obtain long-term features of users and songs, then natural language processing technology to obtain musical context features, and finally use long and short memory network model to train the real dataset, which gives good experimental features [12].

From the analysis of a large number of research literature and results, we can see that, due to the various features of audio signals and the long time scale, how to reasonably allocate computing power to the important characteristics of

model learning users is also a problem to be solved. In recent years, the attention mechanism has given a solution to the above problems. The attention machine mechanism has evolved from human vision, and it was born to solve the high complexity and long time of the neural network. Niyazov and Mikhailova integrated attention mechanism on the basis of deep network DNN, and experiments showed that it has a certain effect on improving the accuracy of recommendation [13]. Scholar Zhang uses CNN networks based on attention mechanism to predict the next step of user behavior. Zhang et al. proposed a low accuracy and long music classification model by adding the attention mechanism to improve the accuracy of classification and proved that this model greatly improves the accuracy of classification [14]. But as of June 2020, there are very few studies on music recommendation areas introducing attention mechanisms. Aiming at the low accuracy of independent recurrent neural network mentioned above and inspired by the application of attention model, this paper proposes a music recommendation system based on attention mechanism, as shown in Figure 1. By integrating the attention mechanism, more attention can be paid to the audio characteristics that affect users’ personal preferences, so as to better learn from users’ music preferences and improve the accuracy of recommendation.

Move on to the next type of music recommendation—special topics. PGC can be understood as a special project done by professional editors to edit some relevant songs together for “packaging” recommendation. With the advent of the era of Web 2.0, this production right has been delegated to ordinary users. Users can make any theme according to their own interest and music calendar, that is, the so-called UGC, user-generated content (strictly speaking, the selected collection made by users is not comparable to the theme made by PGC in terms of layout beauty and flexibility,

but it is not the core value of this kind of products). Not only the production right but also the popular recommended position has gradually been transferred from PGC to high-quality UGC. Now, this mode of using high-quality UGC as PGC is very mainstream, and so is Internet music services. Typical domestic cases that encourage UGC to do music topics include the selected collection of <https://xiami.com> and the song list of NetEase cloud music. It is worth mentioning that some of the music services still dominated by PGC in China have made good achievements, such as the arrest of more than ten years of operation, and still adhere to high-quality independent music recommendation.

From PGC to UGC, music recommendation sources have been greatly enriched, and under the effective design mechanism, the quality of the displayed music content has only increased. However, this process only enriches the content source and still does not change the mechanism of music recommendation. The way for users to obtain information is still to search or browse popular content (similar to song list). Next, let us take a look at the most mainstream music recommendation mechanism.

By the way, if you think about it carefully, are the essence of the several types of music product forms we have mentioned, from the original album, to the list, to the special topic, and the radio station to be mentioned later, all of which are song lists! The core difference between different types of song lists is only the correlation of songs for different reasons. A concept album may consist of ten songs around a story, a theme may consist of several representative works of a singer, and so on. At present, the most thorough implementation of the concept of song list in domestic music products is NetEase cloud music. By using it, you can concretely understand the view that “everything is song list” I put forward here.

3. Methodology

3.1. Music Recommendation Algorithm Based on the Neural Network. Conventional machine learning techniques are limited in processing natural data in their original form. Model building for machine learning systems or pattern recognition has long been very difficult and takes considerable time and effort to complete. Because the feature extractor needs the assistance of professional knowledge to effectively analyze the original data (such as pixels and audio signals of the image and appropriate feature representation) for the classification subsystem to classify and predict the input original data [13]. Deep learning is one of the most popular research fields today, and its proposal brings new opportunities for machine learning. Unlike other machine learning, deep learning methods have multiple learning methods, which are divided into multiple representation levels, and each learning method can be composed of simple but non-linear modules, each converting the previous level into a higher abstract level representation. With sufficient combinatorial transformations, the learning system can learn very complex functions, just like simulating the perceptual process of human brain neural networks on external stimuli [15].

In order to understand the importance of attention mechanism, we must consider that neural network is

actually a function approximator. Its ability to approximate different types of functions depends on its architecture. The typical implementation form of neural network is chain operation composed of matrix multiplication and nonlinearity on elements, in which the input elements or eigenvectors will only interact with each other through addition.

For the classification task, higher-level representations amplify aspects of distinguishing important features and suppress unrelated changes. For example, the first layer of deep learning learns only partial features, regardless of location; the second layer begins to learn the correlation between local features, and the third layer begins to combine partial features according to the correlation to obtain partial features on a larger scale. Therefore, deep learning has better adaptability than other machine learning. More importantly, these feature extraction models are not manually designed, but are adjusted by neural networks through continuous learning, which are effective for different data sources. In terms of application, deep learning has made great progress in intelligent recognition in the field of speech and image and achieved good results, which has greatly promoted the development of artificial intelligence and the leap of human intelligence interaction technology. Deep learning is unique in that it allows it to be composed of multiple processing layers, each of which can be both a traditional neural network layer or processing algorithms in other fields, so that the computational model can not only extend but also learn the representation of data with multilevel abstraction [16].

As shown in Figure 2, the recommendation system based on deep learning mainly includes three layers. The input layer data can be explicit or implicit feedback data, such as the user’s score, browsing or clicking behavior data, or the user’s portrait and item content data, such as the user’s preference, age, image, and audio content, and the data can also be user-generated auxiliary data such as comments. The model layer can use other deep learning models such as deep belief networks, convolutional neural networks, recurrent neural networks, and autoencoders [17]. The role of the output layer is to generate a recommendation list of items for users by using Softmax classification function, similarity calculation, combined with the hidden representations of users and items learned by the model [18]. In addition to the research content to be covered, deep learning can be integrated into any traditional recommendation system. In content-based recommendation systems, collaborative filtering and hybrid recommendation systems can also be applied to recommendation systems based on social networks and situational awareness.

Suppose the number of input layer nodes is N , the number of hidden layers is M , the output layer nodes are L , and the order input vector is x , $x = x_1, x_2, \dots, x_N$. The output vectors are y , $y = y_1, y_2, \dots, y_L$. The neural network is calculated in the following form:

$$\begin{aligned} h_i &= \omega_{ji} \cdot x_j + b_i, \\ o_j &= f_i(\beta_i \cdot h_i), \\ y_l &= v_{lk} \cdot o_l, \end{aligned} \quad (1)$$

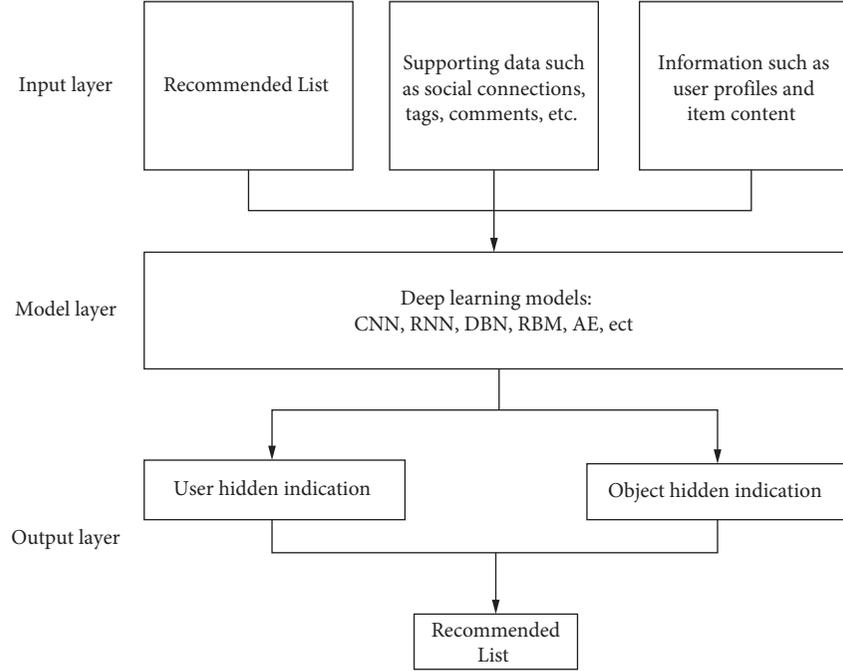


FIGURE 2: A recommended system framework based on deep learning.

where ω_{ji} ($j = 1, 2, \dots, N, i = 1, 2, \dots, M$) represents the weights from the input layer to the hidden layer and b_i represents the bias of the i th neural node. f_i represents the activation function of the first neural node. f_i represents the coefficients of the activation function, and this parameter is introduced to regulate the activation ability of neurons in the hidden layer. v_{ik} represents the weights from the hidden layer to the output layer, o_i and b_i are two intermediate quantities, h_i is the output value of the input layer to the hidden layer input signal after calculation, and o_i is the output value of the neural node after the activation function activation. The sum of error (SSE) is defined as

$$\text{SSE} = \sum_{i=1}^k (y - y')^2. \quad (2)$$

There are three commonly used activation functions. Choosing different activation functions will affect the training and prediction of CNN network models. Compared with sigmoid and tanh, ReLU is more widely used in deep learning, which can not only converge quickly and improve the training efficiency but also effectively alleviate the problem of gradient disappearance. For the above reasons, this paper finally chose the modified linear unit ReLU as the activation function of the training model of the deep neural network.

$$\begin{aligned} \text{sigmoid}(x) &= \frac{1}{1 + e^{-x}}, \\ \text{tanh}(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}}, \\ \text{ReLU}(x) &= \max(0, x). \end{aligned} \quad (3)$$

3.2. Attention Mechanism of Neural Networks. Note that the mechanism calculates a mask for multiplying features. This seemingly innocuous expansion will have a significant impact: suddenly, there is a lot more function space that can be approximated by neural networks, making new use cases possible. Why is that? Although I have no evidence, the intuitive idea is that there is a theory that neural network is a general function approximator, which can approximate any function and achieve any accuracy. The only limitation is the limited number of hidden elements. In any practical setup, this is not the case. We are limited by the number of hidden units that can be used. Consider the following case: we want to approximate the product of neural network inputs. Feedforward neural network can only simulate multiplication by using (many) addition (and nonlinearity), so it needs a lot of neural network foundation. If we introduce multiplicative interaction, it will become simple and compact.

The birth of attention mechanism is to solve the problems of excessive neural network complexity and long time. It is derived from the mechanism of human visual attention. Human vision first browses the global image information, from which the areas need to focus on attention, and second, pay attention to this area, so as to obtain the detailed information of this area, while reducing the reading of useless information and reducing the waste of resources [19]. The core goal of the attention mechanism is to select information that is more critical to the current task objectives from large amounts of information, so as to improve accuracy and remove information redundancy to reduce algorithmic complexity [20].

Figure 3 shows the weights of the Attention mechanism. Understand each element in Source as a set of key value pairs of key and Value. Given a data in the target, by calculating the correlation or similarity of each data with each key, it

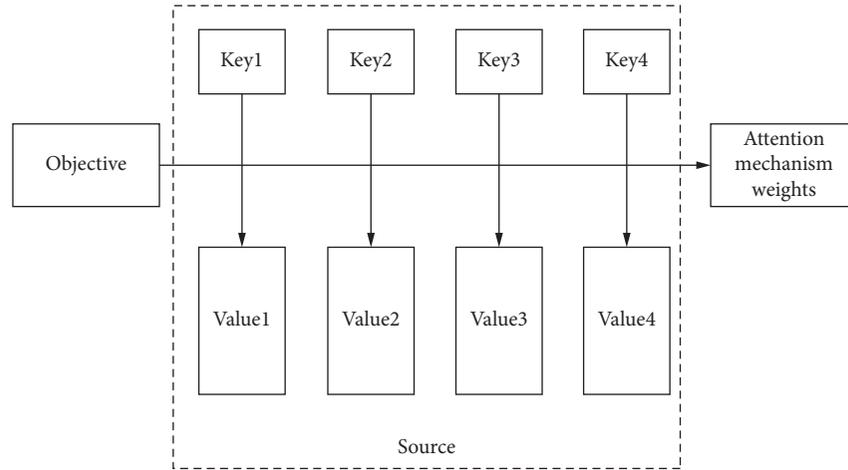


FIGURE 3: Attention mechanism.

obtains the weight coefficient of each key and then weighted sums the Value and finally obtains the weight value of the attention mechanism. The weight represents the importance of the information; the larger the weight, the more the focus on its corresponding Value value, and the Value is its corresponding information. So, essentially, the Attention mechanism is a weighted sum of the Value values of the elements in the Source, while the target and Key are used to calculate the weight coefficient of the corresponding Value. That is, its essential idea can be rewritten into the following formula:

$$\text{Attention}(\text{Target}, \text{Source}) = \sum_{i=1}^{L_x} \text{Similarity}(\text{Target}, \text{Key}_i) * \text{Value}_i, \quad (4)$$

where L_x represents the length of the source. Since the music recommendation system needs to analyze the audio that users listen to historical music and the music audio has many features and long timing, using the attention mechanism to weigh the audio can reduce the complexity of the algorithm and improve the accuracy of the algorithm.

3.3. Improved Music Recommendation Algorithm for the Deep Neural Network Based on Attention Mechanism. Three-dimensional user portraits can bring subversive traditional big data aided decision-making solutions to traditional retail chains or enterprises, rather than making decisions and judgments through various qualitative analysis results by upstream and downstream service providers at all levels. In fact, it is also an impact and break on the value chain of the existing retail chain industry.

Personalized music recommendation algorithm learns users' preferences through users' behavior data and finally predicts users' favorite songs and pushes them to the system users. This purpose can be abstracted as a dichotomous mathematical problem [21]. A hybrid recommendation algorithm based on attention mechanism and independent recurrent neural network by RNN (AINRNN) is improved based on the deep learning neural network.

In the algorithm data preprocessing stage, the training set is divided into two parts: user history listening to audio and user portrait. User history listening audio is provided by the user listening to music. The user portrait is composed of three parts: user age, user listening to song language, and song type [22]. The framework diagram of this algorithm is shown in Figure 1. During the feature extraction phase, first, the user history listening audio is extracted through the improved MFCC (Mel Frequency Cepstral Coefficient), HFC (high frequency content), HPCP (harmonic pitch class profiles), and other audio feature extraction algorithms. At the same time, the user portrait data is feature extracted, subsequently combining the audio extraction features with the user features and training through the AINRNN network. Finally, the recommendation list is obtained through the first layer (softmax). This is the framework of this paper based on an improved personalized hybrid recommendation model of AINRNN. The attention mechanism used represents the number of users listening to songs as the user's preference of song characteristics, so that the model can better learn the personalized weight of users' music preferences.

The user history listening records and user portrait in the training set are taken as input to the model, and the output recommendation list is the favorite song set of the model. AINRNN recommendation algorithm builds a deep INDRNN network by stacking the basic neural structure of INDRNN layer by layer, while introducing attention mechanism and changing the network input processing mode of residual connection to full connection mode, giving a hybrid recommendation algorithm based on attention mechanism and improved RNN. With the help of the ReLU unsaturated activation function, over a time step, the model gradient can be consistently mapping, while being directly propagated to other layers. The structure diagram and flow charts of the algorithm model are shown in Figures 4 and 5.

Figure 5 shows a flowchart of the hybrid recommendation algorithm based on attention mechanism and improved AINRNN. It is divided into data preprocessing and prediction stages. In the data preprocessing stage, the data is extracted from the input list in order, the data is extracted into music audio data and user portrait, the user history

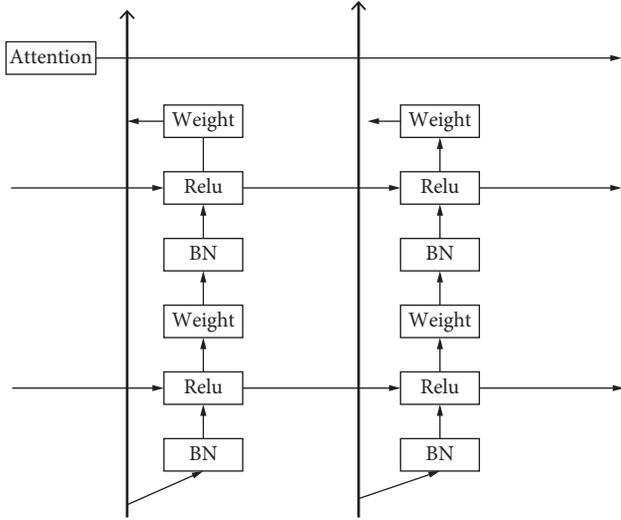


FIGURE 4: Model structure diagram of the AIRNN recommended algorithm.

music audio features are extracted by scattering transformation, and the extracted audio features are taken as input to the above algorithm. In the prediction stage, the above extracted features are used through the attention mechanism and the improved AIRNN hybrid recommendation algorithm, and finally, the list is generated and the data is recorded, while judging whether the input list is completely traversed. If so, display the results to the user from high to low; otherwise, process the next data and repeat the above steps until completion.

4. Result Analysis and Discussion

4.1. Experimental Dataset and Preprocessing. In the field of deep learning vision, relevant enterprises and organizations at home and abroad have opened many excellent benchmark datasets such as MNIST, COCO, CIFAR, and Open Image. Image Net, for different application scenarios, is commonly used for researchers. These publicly available datasets play a crucial role in promoting research and development in related fields. However, unlike the large number of publicly available images or texts, the field of music information retrieval or recommendation has been lacking in large, mature, complete, and easy-to-use benchmark datasets. This has to some extent limited the research and application of models such as deep neural networks that usually require large amounts of data training in this field. Some publicly available music data sets are listed in Table 1.

From the above table, it can be seen that audio files may be subject to strict copyright control by record companies, and only some small data sets will distribute audio files. For large data sets, such as audio sets and acoustic Brainz, audio files are not directly provided, and important information such as user records is not included. Only audio features are included, or audio download links are provided. In all of the above datasets, only the MSD dataset contains both the audio and the required user record information. The music dataset used by the algorithm is Million Song Datasets or

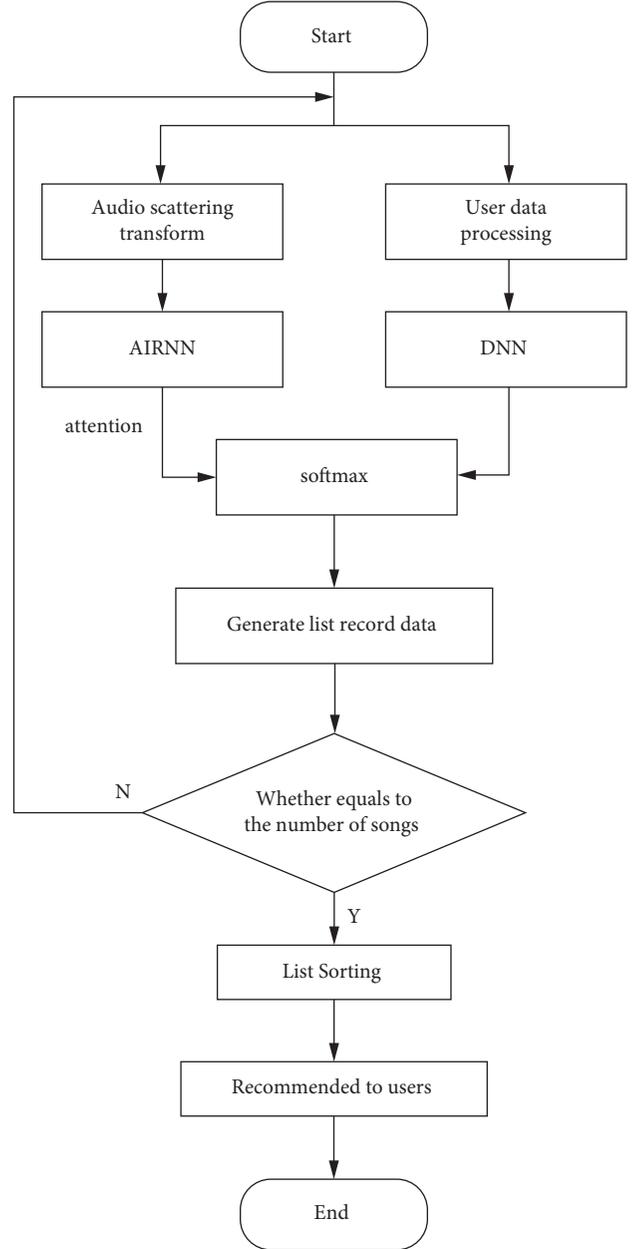


FIGURE 5: Flow chart of the AIRNN algorithm.

TABLE 1: Currently commonly used publicly available music data sets.

Data set	Sample number	Audio frequency	User records
Ballroom	698	Have	Not have
GTZAN	1000	Have	Not have
ISMIR 2004	1458	Have	Not have
MagnaTagAtune	25863	Have	Not have
MSD	1000000	Interlinkage	Have
Audioset	2084320	Interlinkage	Not have
AcousticBrainz	252473	Not have	Not have

MSD, the million music dataset. MSD is the first public dataset in the field of music recommendation. It is a resource integration platform, which brings together multiple authoritative and well-known music community data,

containing information on 1 million songs. The simulation comparison test mainly uses the audio feature set in the core data set of MSD and the data provided by its subset Taste Profile.

Since MSD has information about 1 million songs and a total of 283 GB, not screening will seriously slow down the training speed of the model. First, to ensure that the model calculation was valid, users with less than 100 historical listening tracks were removed, with 359,687 songs and 82,310 users remaining. Then, using the mainstream ten-fold cross-validation method, the filtered user data was evenly divided into 10 copies, one as the test set, and the rest as the training set. During the model training process, the overfitting parameter was set to 0.5, and the learning rate was set to 0.0005. When the convergence range of the simulation accuracy curve is stable, we will stop training and save the model.

4.2. Feature Extraction of the Fused Residual Network and Identification Algorithm of MRF Grayscale Information. We use the designed algorithm to analyze the recommended results of music and test the effect of the algorithm. Due to the diversified evaluation indexes of the recommendation algorithm, the user satisfaction, accuracy, and normalized loss cumulative gain indexes are selected for simulation verification. The accuracy is defined as shown in the following equation. Taking the first k of the recommended list calculated by the model, we can intuitively judge the prediction results.

$$\text{accuracy} = \frac{\sum_k (L(k) \cap T(k))}{\sum_k T(k)}. \quad (5)$$

NDCG represents the ranking position score situation of the high correlation degree of model output and input and is an important indicator to evaluate and measure the performance of music recommendation algorithm. The NDCG definition is as follows:

$$\text{NDCG} = \frac{\sum_{i=1}^p (\text{rel}_i / \log_2(i+1))}{\sum_{i=1}^{|\text{rel}|} (2^{\text{rel}_i-1} / \log_2(i+1))}. \quad (6)$$

Here, rel_i represents the benefit of the i 'th result in the recommendation list, and $|\text{rel}|$ represents the results ranked in the optimal way.

User satisfaction index can make users intuitive to evaluate the performance of the recommendation algorithm. This simulation uses a questionnaire survey to obtain user satisfaction. There are 10 grades on the questionnaire, and 10 is the highest score. To evaluate the importance of the index by using the subjective empowerment method Liszt scale, the score was used as the basis for the weight calculation, while the mean score was used as the original relative influence coefficient. Thirty volunteers were investigated to experience three different music recommendation algorithms based on LSTM, INDRNN, and AINRNN. The anonymous scores of volunteers were investigated by a questionnaire. The weighted sum was used as the satisfaction score. The overall satisfaction

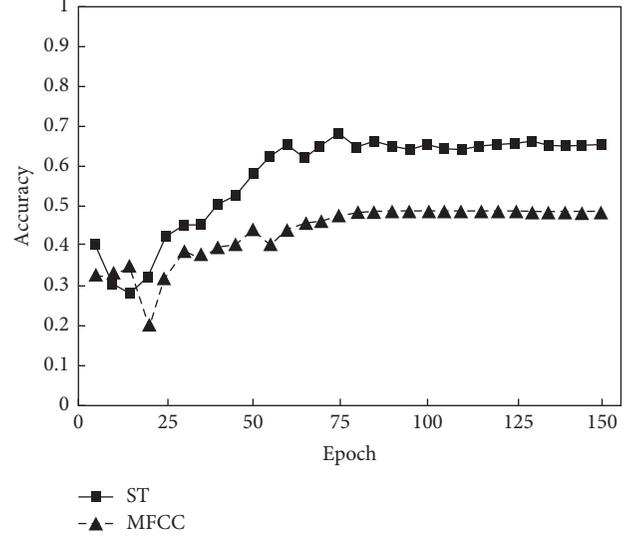


FIGURE 6: Effect of ST and MFCC preprocessing methods on algorithm accuracy.

score of the algorithm was the average value of the satisfaction score of volunteers.

Figure 6 shows the influence of MFCC and ST on multiple algorithms for two different preprocessing methods of audio, and the model is trained with a 5-layer AINRNN network. The results show that when the accuracy stabilizes, i.e., the training reaches 75 epochs, the ST of the simulation group is 19.7% higher than the MFCC accuracy of the simulation group. This also verifies the conclusion that the algorithm gives in the data preprocessing stage and that the scattering transformation can compensate MFCC for losing audio features over 25 ms during feature extraction. At the same time, it also intuitively proves that the audio features of more than 25 ms are potentially related to whether the user likes the song, and the longer the features, the better the prediction effect.

Figure 7 compares the accuracy of the AINRNN algorithm with the accurate INDRNN network algorithm and the classical LSTM network algorithm. As can be seen from Figure 7, the accuracy of the LSTM algorithm is always far from the other two algorithms, with AIN RNN with an accuracy of 67.8% and IND RNN with an accuracy of 61.8%.

Although, AINRNN and IND RNN in the figure above, the NDCG index of AINRNN is 7.3% higher than INDRNN in Figure 8. It shows that AINRNN can learn the potential relationship between audio sequences well and solve the problem of RNN and LSTM gradient explosion to some extent, thus confirming that AINRNN algorithm has a good performance on music recommendation. As k is worth increasing, the accuracy trend of the three algorithms gradually changes from rising to decreasing and finally approaches zero. As can be seen from Figure 7, for the inflection point of this function at k values of 8–10, it is appropriate to set the recommendation list to 10 heads. Compared with the current popular music portals such as QQ and NetEase cloud, the mainstream online music recommends 15 songs every day. Excluding a small number of recommended songs is not determined by the recommendation algorithm, but by the

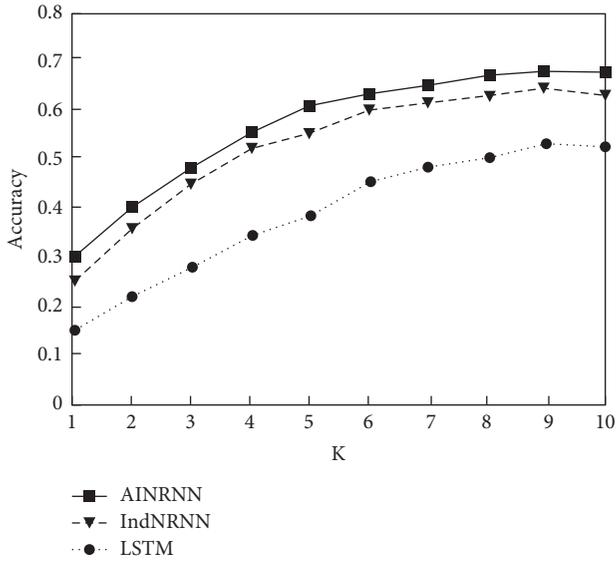


FIGURE 7: Precision of the three algorithms.

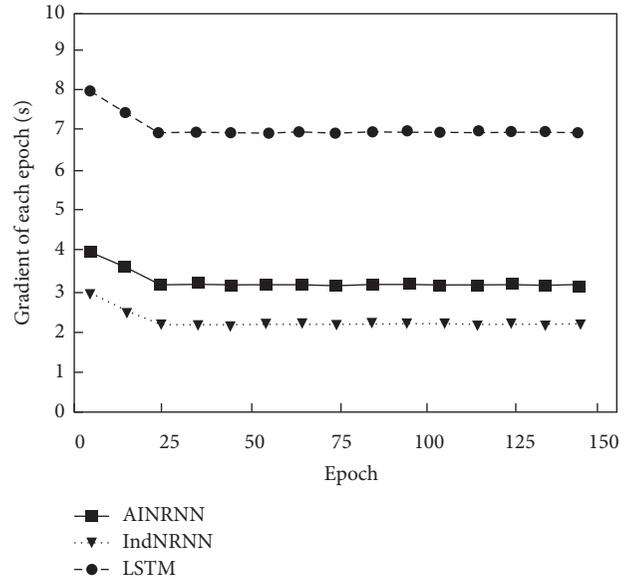


FIGURE 9: Training time of the three algorithms.

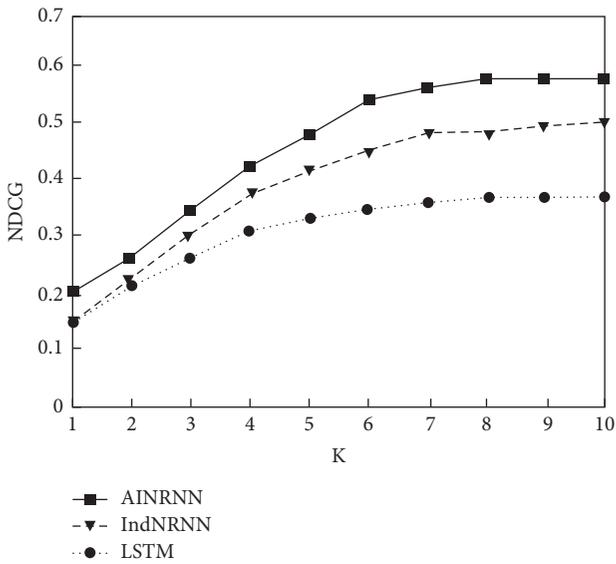


FIGURE 8: The effect of k values on NDCG indicators.

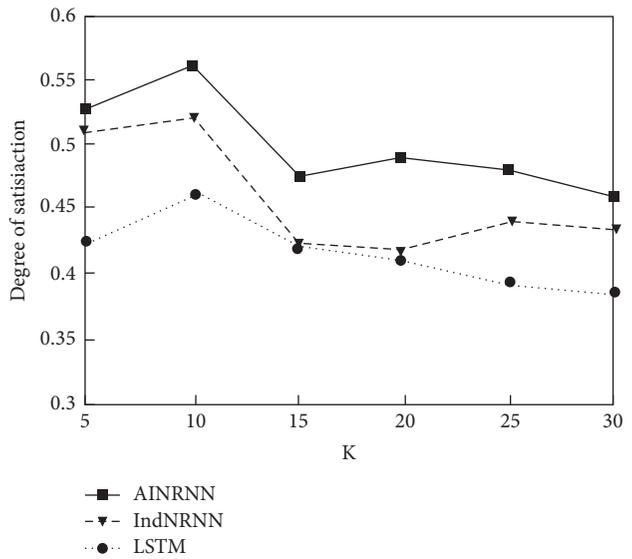


FIGURE 10: User satisfaction.

business promotion model. The number of remaining music recommendations is similar to the simulation results in this section.

Figure 9 represents the training time of the different algorithms in the music recommendation systems. It can be seen that the classical LSTM algorithm not only has a long training time but also has the worst effect. The training time is the shortest, the AIND RNN state algorithm increased by 1.1 s compared with IND RNN, but combined with Figures 8 and 9, AINRNN NDCG index is 58.1%, 7.8 percentage points higher than IND RNN, and the accuracy index is also 6% higher. According to the model of mainstream commercial music business, it is mainly divided into offline calculation and online recommendation, with the frequency of once a day. It is acceptable to sacrifice a small amount of

computing time to improve the accuracy of the algorithm and NDCG index.

Figure 10 shows that the user satisfaction based on the hybrid recommendation algorithm of attention mechanism and improved RNN is also the highest. It can be shown that the above improved algorithm can be better applied in the field of music recommendation.

5. Conclusion

The recommendation algorithm resets the user's recent behavior right high, so the system is easy to recommend a large number of similar types of songs, reducing the user's aesthetic fatigue and excitement. It is suggested to add the analysis of users' use scenes into the algorithm and limit the

number and location of songs of the same type. For long tail unpopular songs, due to the relatively small amount of data, we should pay more attention to effect feedback. Deeply mine data and enrich recommendation elements, such as recommending music, albums, and singers that affect a user's favorite singer according to his/her creation/growth background. With the increasingly close combination of the Internet and people's production and life, the total amount of global data and information has also grown increasingly. In order to save users the time to find their favorite music in many music, the music recommendation service arises at the historic moment and is widely concerned by scholars. Traditional music recommendation system based on the collaborative filtering algorithm has low recommendation accuracy, poor real-time performance, sparsity, system cold start, and so on. This paper focuses on solving the problem of low accuracy of the recommendation system and considering the cold start problem of the system. To solve the problem of low accuracy of music recommendation algorithm, a hybrid recommendation algorithm based on attention mechanism and improved AINRNN is proposed and is implemented by a neural network composed of independent recurrent neural network model and attention mechanism. The scattering transformation can be used to extract the characteristics of the long-time signal, to data preprocess the historical audio heard by the user and to extract the effective features. Then, the model is trained through the independent recurrent neural network with mixed attention mechanism. The attention mechanism can solve the problem of long and difficult deep learning training time, and finally, the recommendation list is obtained through the softmax layer. The results show that the algorithm proposed in this paper can solve the problem of music recommendation well and improve the highest accuracy of the independent recurrent neural algorithm by 8.5% and 20.9% compared with the classic LSTM (long and short-term memory network) music recommendation algorithm. Therefore, the hybrid algorithm can indeed better improve the recommendation accuracy, and we hope to conduct more indepth experiments in future research.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This study was supported by the Role of Creative Dissemination of Red Music Culture of Dabie Mountains in the Construction of Modern National Identity in the New Era; Phase Achievements of Henan Soft Science Research Project (project no. 202400410223); the Role of Creative Inheritance of Red Opera Culture in the Construction of National Identity in the New Era; Henan Education Department

Humanities and Social Sciences Research General Project Stage Achievement (project no. 2020-ZZJH); and Study of Xinyang's Red Music Culture from the Perspective of History; Xinyang Soft Science Research Phase (project no. 20200052).

References

- [1] A. Baijal, V. Agarwal, and D. Hyun, "Analyzing images for music recommendation," in *Proceedings of the 2021 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, Las Vegas, NV, USA, January 2021.
- [2] P. Åman, *Musical serendipity: designing for contextual music recommendation and discovery*, Aalto University, Espoo, 2018.
- [3] G. S. Berns, C. M. Capra, S. Moore, and C. Noussair, "Neural mechanisms of the influence of popularity on adolescent ratings of music," *NeuroImage*, vol. 49, no. 3, pp. 2687–2696, 2018.
- [4] Y. Deldjoo, M. Schedl, and P. Knees, "Content-driven music recommendation: evolution, state of the art, and challenges," 2021, <https://arxiv.org/abs/2107.11803>.
- [5] M. S. Fathollahi and F. Razzazi, "Music similarity measurement and recommendation system using convolutional neural networks," *International Journal of Multimedia Information Retrieval*, vol. 5, pp. 1–11, 2021.
- [6] F. Guan, C. Yu, and S. Yang, "A gan model with self-attention mechanism to generate multi-instruments symbolic music," in *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, July 2019.
- [7] K. Jearanaitanakij, T. Boonloy, A. Thongpuek, R. Chawuthai, and B. Pasaya, "Music recommendation by using favorite level," in *Proceedings of the The TNI Academic Conference 2017*, Bangkok, Thailand, May 2017.
- [8] Y. Yiyi, L. Jia, and L. Depeng, "Effect of perceived in-depth differences on employees' voice behavior: the mechanism of perceived emergent states," vol. 39, pp. 97–112, 2017.
- [9] D. Kowald, M. Schedl, and E. Lex, "The unfairness of popularity bias in music recommendation: a reproducibility study," in *Proceedings of the Advances in Information Retrieval, 42nd European Conference on IR Research, ECIR 2020*, Lisbon, Portugal, April 2020.
- [10] P. Magron and C. Fevotte, "Leveraging the structure of musical preference in content-aware music recommendation," in *Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, June 2021.
- [11] Q. H. Nguyen, T. Do, T. B. Chu et al., "Music genre classification using residual attention network," in *Proceedings of the 2019 International Conference on System Science and Engineering (ICSSE)*, Dong Hoi, Vietnam, July 2019.
- [12] H. F. Zhou and L. Y. Zhou, "Music session recommendation algorithm with Word2Vec," *Communications Technology*, 2019.
- [13] A. Niyazov, E. Mikhailova, and O. Egorova, "Content-based music recommendation system," in *Proceedings of the 2021 29th Conference of Open Innovations Association (FRUCT)*, Tampere, Finland, May 2021.
- [14] H. Zhang, K. Zhang, and N. Bryan-Kinns, "Exploiting the emotional preference of music for music recommendation in daily activities," in *Proceedings of the 2020 13th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, December 2020.
- [15] M. Schedl, "Investigating country-specific music preferences and music recommendation algorithms with the LFM-1b

- dataset,” *International Journal of Multimedia Information Retrieval*, vol. 6, no. 1, pp. 71–84, 2017.
- [16] C.-D. Wang, W.-D. XI, L. Huang, Y.-Y. Zheng, Z.-Y. Hu, and J.-H. Lai, “A BP neural network based recommender framework with attention mechanism,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 99, p. 1, 2020.
- [17] S. Liu, X. Tu, X. A. Cheng, and L. Renfa, “Deep neural networks with attention mechanism for monocular depth estimation on embedded devices,” *Future Generation Computer Systems*, vol. 131, 2022.
- [18] W. Jiao, H. Wang, and G. Zhong, *MusicCNNs: a new benchmark on content-based music recommendation*, Springer, Cham, 2018.
- [19] T. Yoshida and T. Hayashi, “OtoPittan: a music recommendation system for making impressive videos,” in *Proceedings of the IEEE International Symposium on Multimedia*, San Jose, CA, USA, December 2016.
- [20] M. Schedl and C. Bauer, “Online music listening culture of kids and adolescents: listening analysis and music recommendation tailored to the young,” in *Proceedings of the 11th ACM Conference on Recommender Systems (RecSys 2017): International Workshop on Children and Recommender Systems (KidRec 2017)*, Como, Italy, August 2017.
- [21] Y. Tao, Y. Zhang, and K. Bian, “Attentive context-aware music recommendation,” in *Proceedings of the 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*, Hangzhou, China, June 2019.
- [22] L. Zhu, Y. Chen, and W. Jiang, “A hybrid approach for music recommendation,” in *Proceedings of the 2009 International Conference on Computational Intelligence and Natural Computing*, June 2009.