Hindawi
Mobile Information Systems
Volume 2022, Article ID 3387598, 14 pages
https://doi.org/10.1155/2022/3387598

Research Article

Music Recommendation System and Recommendation Model Based on Convolutional Neural Network

Yezi Zhang

The Catholic University of Korea Graduate School, Seoul 14662, Republic of Korea

Correspondence should be addressed to Yezi Zhang; 181406018@njnu.edu.cn

Received 19 January 2022; Revised 4 March 2022; Accepted 20 April 2022; Published 12 May 2022

Academic Editor: Chia-Huei Wu

In today's era of big data with excess information, music is common and everyday, which shows the huge amount of music data. How to obtain one's favorite music from the massive music database has become a problem, and the emergence of music recommendation systems is also inevitable. In this paper, we take digital piano music as the research object, form comprehensive features using spectrum and notes, design classification methods using convolutional neural networks, and further process the classification results and design recommendation algorithms. The basic method of music recommendation of this algorithm is to determine the structure of the network model, determine the corresponding training model, and improve the parameters on the basis of the typical source network model used in the system experiment. Historical behavior chooses to collect information. Then, it reads the audio data on the system and retrieves it from Mel, which reveals the identity of the music. The classification proposal achieves its goal by denying the similarity between customer preferences and the potential of two musical characteristics. Two recommended methods based on convolutional neural networks are tested in this article. On the whole, the accuracy of the user's comprehensive feature, recommendation method is higher than the recommendation accuracy rate of the multicategory user. In the comparison experiment of the single-category and multicategory recommendation methods, the average accuracy rate of single-category user feature recommendation is 50.35%; and the recommendation accuracy rate of multicategory user features is higher than the recommendation accuracy rate of single-category user features. The experimental results show that the two recommendation methods can achieve better recommendation results.

1. Introduction

1.1. Background. While multimedia technology is developing rapidly following the pace of the times, we have also entered the era of big data with massive amounts of data. Music is one of the contents of multimedia information. While there is a huge user demand, its own data volume is also very considerable and continues to grow [1]. Nowadays, with abundant music resources, how to efficiently obtain the songs of interest in the vast and complicated sea of music, a targeted solution of a personalized music recommendation system is proposed. Music recommendation algorithms predict and push user behavioral preferences based on user behavioral information and music characteristics. The development of music recommendation algorithm has also shifted from the technical route recommended by users' personal preferences to the mutual recommendation among users. In recent years, the focus of attention has been on tapping potential preferences. In general, the technology of music recommendation algorithms is getting better and better.

1.2. Significance. Amounts of data, different groups of people have different understandings and feelings of music, and a single attribute recommendation algorithm cannot meet the diverse and personalized needs of users. The music recommendation system that emerged from the massive music database is at a time when the data sea is becoming more and more complicated and full of personalized needs. The original recommendation algorithm needs to solve the original problems while upgrading itself. The implementation

and development of the same big data has also brought new computer technology, which can provide help in designing and improving the accuracy of the algorithm.

1.3. Related Work. Combining previous research on neural network recommender systems based on deep learning by domestic and foreign scientists with research results in these fields, there is still plenty of room for expansion of music recommendation systems. In it, Liu et al. reported new behavior and comments based on a standard agreement (NERAR) for approval that combines features and recommendations. A large number of experiments show that the established model is better than the most advanced recommendation model in terms of recommendation accuracy [2]. Anantha and Battula use a pretrained model to extract knowledge from the data using the concept of transfer learning. The established model uses the knowledge of the pretrained model to extract patterns between users and items. In order to achieve this goal, the article introduces a method to generate recommendations in two stages. In the classification stage, the classification of product images and its experimental analysis is discussed next. The ranking stage and its experimental analysis of sorting product images to users are discussed. The result analysis of the discussion has achieved gratifying results [3]. In order to overcome the problem of making recommendation difficult as the number of products and users increase, Son and Shim proposed a matrix factorization algorithm that uses text data related to products. Among matrix factorization algorithms, a word-level convolutional neural network method is used to extract word-level features to effectively reflect text data. There is a problem with word-level convolutional neural networks; that is, there are a lot of parameters to learn. Therefore, in their research, they proposed a matrix factorization method that uses character-level convolutional neural networks to extract character-level features from text data. In addition, in order to verify the performance of the proposed matrix factorization method, experiments were carried out using real data [4]. To solve the problems of the personalized recommendation system, the main purpose of this model by Zhang et al. is to study social connections and trust relationships among users and recommend places of interest to users through social impact and geographic information between users and tourist attractions. The results of the data prove the feasibility and effectiveness of the model algorithm actually proposed by them, providing better prediction accuracy compared to other recommended algorithms [5]. Based on the theory of social communication, Bi et al. proposed an effective social reference network model. It combines the attention mechanism and bidirectional LSTM in the same frame and uses multilayer perceptrons [6]. In their research, Lee et al. proposed an automatic melody extraction algorithm using deep learning. In this algorithm, the feature image generated by the band energy is extracted from the chord audio file [7]. In order to solve the problem of finding one's interest in a wide range of products, Yan Y et al. proposed a model algorithm that uses a binary convolutional neural network and radical basis function (RBF). They preprocess the input data to 0 or 1 based on the convolutional neural network, which can fully save the data storage and space, and the side help improves the efficiency of the recommendation. They also use RBF to establish a kinship network and make recommendations based on useful information screened out by similar users in the kinship network [8]. However, the above-mentioned studies need to learn a large number of parameters, which require a large amount of labeled training data and a large amount of professional knowledge to ensure correctness, ignoring issues such as the importance of the strength of the relationship between users of different levels in the network.

1.4. Innovation. For modern people who always carry smart phones in their daily lives, listening to music is a very personal and contextual behavior. The rapid development of mobile devices and cloud-based music services enables consumers to enjoy music, anytime and anywhere [9, 10]. Therefore, there is an increasing demand for researching smart technologies to promote context-aware music recommendation. At this stage, the commonly used music system recommendation algorithms are divided into two methods: content-based and context-based. The theme of this paper is to take digital music as the research object, design and classify the music recommendation system based on convolutional neural network [11, 12], and further process the recommendation algorithm model in combination with personal preference information.

2. Music Recommendation Technology Based on Curly Neural Network

In order to adapt to people's increasingly personalized needs and the rapidly increasing amount of information, music recommendation technology proposes personalized recommendations. Music recommendation based on user tastes is one of the hottest applications in the field of personalized recommendation at present, and personalized recommendation is indispensable on all major music websites at home and abroad, which shows the importance of recommendation systems. For the classification of recommendation algorithms, extending the development on the original basis, it is mainly divided into content recommendation algorithm and context recommendation algorithm [13] and deep learning recommendation algorithm [14] according to different objects.

2.1. Content-Based Recommendation Algorithm System. The core point of content-based recommendation algorithm is to recommend items with a high degree of similarity based on historical selection behaviors that have been liked [15]. In digital music, it refers to extracting audio data as a music carrier, extracting part of the underlying sound characteristics through audio signal analysis to calculate similarity, and then constructing a data structure describing their preferences through audio similarity, and then using the similarity of user models to judge the similarity of two music to achieve the purpose of recommendation [16, 17]. The process of this recommendation algorithm has three steps.

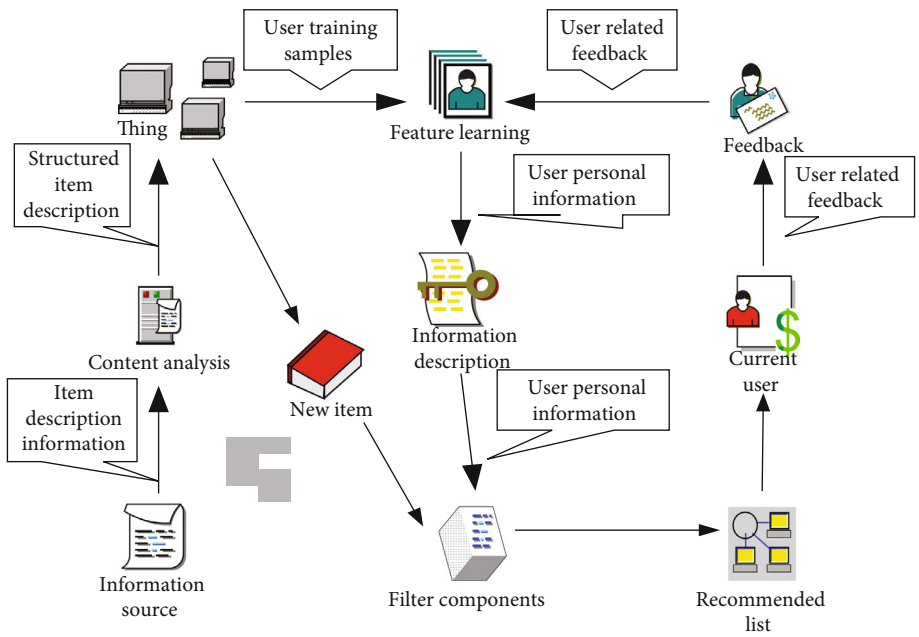

Figure 1: Process diagram of content-based algorithm.

As shown in Figure 1, the three parts are as follows: first, extract some features for products that users have liked in the past. Second, based on the extracted features of disliked in the past, a comprehensive analysis of the likes is made. Third, recommend the most similar products for users by comparing the preference characteristics and the characteristics of favorite items [18].

In digital music, it is the study of audio data, and audio data needs to be processed, such as Fourier transform, variable, and windowing using Mel spectrogram [19]. Mel spectrogram is abbreviated as MFCC, and the MFCC coefficient is determined by the mask phenomenon of different wavelengths and frequencies that affect the sensitivity of the human ear. Receive a simple signal by filtering the water from low frequency to high frequency. Features fit the body and face of the human ear, with good recognition [20]. The speech extraction process of MFCC is shown in Figure 2.

Using the Mel scale [21] to illustrate the characteristics of human physiological auditory frequency, the extracted parameters and the frequency formula are

$$\mathrm{Mel}(f) = 2595 \times \lg\left(1 + \frac{f}{700}\right). \tag{1}$$

Calculate the MFCC coefficient, calculate the logarithm of the spectral energy according to the MEL scale, and perform the discrete cosine transform. Use g to represent the number of triangular filters, Xk to x output filters, and n to order fi to MFCC coefficients; the calculation formula and process are as follows:

$$fi = \int_{g}^{2} \sum_{x}^{g} \log(Kx) \cos\left[i\left(x - \frac{1}{2}\right)^{\pi}_{g,i=1,2,\cdots,n}\right. . \tag{2}$$

To get the feature coefficients of higher priority, further processing parameters are needed, and Gaussian mixture model (GMM) is used to describe the spatial distribution of the linear prediction coefficient cepstrum [20], and the formula expression of the style feature vector is obtained:

$$p = \sum_{g=1}^{f} \log \sum_{y=1}^{m} MyKy(Xg). \tag{3}$$

In the formula, p is the likelihood ratio, Xg is the style feature vector, f is the total number of frames contained in the music segment, $ky(Xg)$ is the probability function of the Gaussian component, and my is the weight of the y-th component.

There are two advantages to extracting vector values from audio data. (1) It avoids directly processing the original data, improves efficiency, and reduces the amount of complex data calculations. (2) The data features obtained are more accurate and obvious, and subsequent recommendations are more convenient [22].

The content-based recommendation system can establish a profile that avoids complicated calculations and can describe music features. It has similar functions to feature vectors. Compared with the computing features of audio, content-based recommendation algorithms have greater advantages in text information and more mature [23].

2.2. Context-Based Recommendation Algorithm. Now is the society of the information age. Big data permeates all aspects of life everywhere. Users are also acquiring data while creating data. The closer the connection between them, the more relevant information generated over time. Context-based

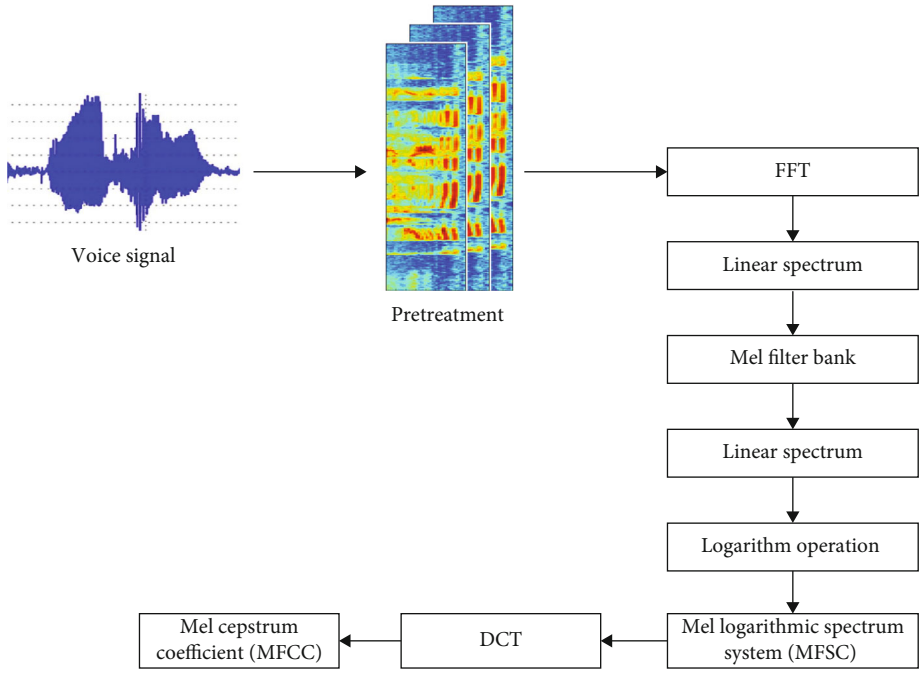

Figure 2: MFCC feature parameter extraction process.

recommendation algorithm is to mine the user's behavior data to calculate its relevance for recommendation [24].

Collaborative filtering recommendation algorithm is a relatively mature recommendation system at present. Once it is launched, it will play a role in various fields of the Internet. It is different from content-based algorithm recommendation which requires calculation of a large number of complex audio data features. The algorithm process can be briefly summarized into three steps: first, collecting users' preferred behaviors and historical choice information and calculating the similarity degree between individual users' information; then, evaluating and analyzing users' behaviors and choices based on their similarity degree and deriving users' potential choice behaviors relative to other users; and finally, recommending users based on their potential choice behaviors. For users with the same interest, the collaborative filtering recommendation algorithm will use the average score of other users as the score of the unrated user. The disadvantage of this method is that it cannot take into account the differences between users. After further improvement, we must pay attention to the differences in preferences between different users.

The collaborative filtering recommendation algorithm is simply things and things and people and people. As shown in Figures 3 and 4, the calculation is based on the user and the item, respectively. These are all subcategories of general memory-based algorithms. Memory-based algorithms use known product rating information, the so-called "memory" to guess user preferences or recommendations. Assume a collection of user $A = \{A1, A2, \cdots, \mathrm{An}\}$. Then, $B = \{B1, B2, \cdots, Bm\}$ is the product set. Let RAB be the user's rating of the product. Given that Aa has a higher degree of similarity to A, the inferred form of R is

$$RAB = \frac{1}{n} \sum_{A \in Aa} RAB, \tag{4}$$

$$RAB = k \sum_{A \in Aa} \mathrm{sim}(A, Aa) \cdot RAaB, \tag{5}$$

$$RAB = \overline{RA} + k \sum_{A \in Aa} \mathrm{sim}(A, Aa) \cdot \left(RAB - \overline{RA}\right). \tag{6}$$

In the formula, k is a standardized factor, usually

$$k = \frac{1}{\sum_{A \in Aa} |\mathrm{sim}(A, Aa)|}, \quad sim(i, u), \tag{7}$$

Among them, the i and u in $\mathrm{sim}(i, u)$ represent the similarity between users, and the average score represented by $\overline{RA}$ can be defined as

$$\overline{RA} = \left(\frac{1}{|BA|}\right) \sum_{A \in Aa} RAB. \tag{8}$$

In

$$BA = \{b \in B\} | RAB \neq 0\}. \tag{9}$$

In formula (4), the weighted average is commonly used. Since formula (5) uses a standardization factor k, formula (5) can calculate the similarity measure between users of different recommendation systems. Formula (6) only pays attention to the deviation of the mean value of the user's

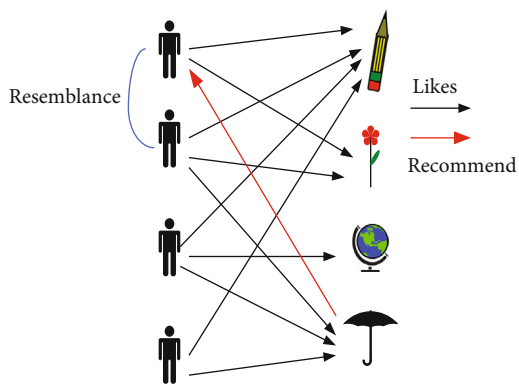

Figure 3: Based on user similarity.

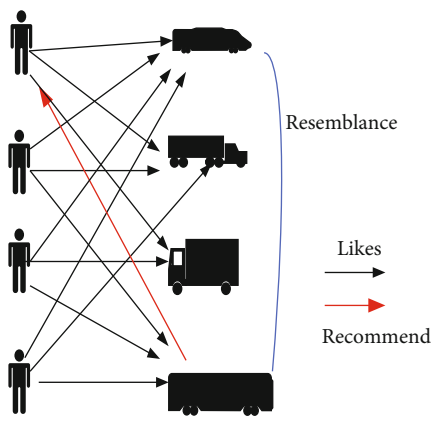

Figure 4: Based on item similarity.

preference, without the shortcomings of different evaluation standards. In this way, the accuracy of formula (6) is higher.

The main research object of the context-based recommendation algorithm is the context, that is, all the information associated with music, and generally recommends music based on the user's preference behavior and historical choices. Some scholars believe that social information based on the application content model, such as the similarity of art genres, emotional semantic space, and other contextual information, can describe the characteristics of music from the perspective of users. However, the public opinion information of popular music can cause chain problems of recommendation reactions.

The combination of contextual information and collaborative filtering can expand the connection dimensions of the data collection and form and achieve more accurate personalized recommendations. Also, the concept of disproportionate sharing of consortia is always present in the data model. Not only will users be associated with data information, but users and data will also be associated with contextual information separately, so that contextual information can also be used as the basis of collaborative filtering to evaluate the degree of similarity between users. Use this indirect link to recommend, even if there is no link between the user and the data; it can be evaluated through the contextual relationship between the user and the user.

2.3. Recommendation Algorithm Based on Neural Network Deep Learning. In the era of information big data facing a large amount of data information, with the development and application of artificial intelligence, it provides convenience for people in various situations in daily life, and the personalized intelligent recommendation system will provide effective services in life. However, the construction of machine learning systems and model models is very difficult. Researchers have begun to look for other feature extraction methods to improve recommendation efficiency.

The deep learning proposed in recent years has brought new possibilities to machine learning. Unlike traditional machine learning, deep learning technology has been successful in the processing of computer vision and voice images. At this time, it has great performance in representation learning. The potential of deep learning, through sufficient combination changes, can simulate a perception process of the human brain and nervous system network the received external stimuli. Therefore, deep learning can be obtained through learning and adjustment through neural networks on the feature extraction model, rather than artificial design. Deep learning, a branch of machine learning, is an algorithm that attempts to perform high-level abstraction of data using multiple processing layers that contain complex structures or consist of multiple nonlinear transformations. The uniqueness of deep learning is that it allows multiple processing layers to be composed while the processing layer can be a traditional neural network or algorithms in other fields, as shown in Figure 5. Such a computing model can not only be extended but also learnt.

Currently, as one of the representative algorithms of deep learning, convolutional neural networks have achieved the best current results in computer vision, classification, and other fields. In a typical deep learning system, each neuron uses weights to "save" the acquired features, and in order better fit external stimuli, nonlinear structure models are added to the original "memory unit" linear structure model. For the linear activation function, the sigmoid function is often used as a typical activation function of neural networks, which can compress continuous values to (0,1).

$$\lambda(z) = \frac{1}{1+e^{-z}}. \tag{10}$$

Among them, the reciprocal z can be expressed by itself:

$$\lambda'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \lambda'(z)\left(1-\lambda'(z)\right). \tag{11}$$

The z in the formula refers to the linear model of the neural unit:

$$Z = \sum_i wixi + b. \tag{12}$$

In the past, some scientists estimated that only 1-4% of the active neurons in the human brain accounted for the lack of functional neurons. From the perspective of extracting sparse features better and faster, the selective response of

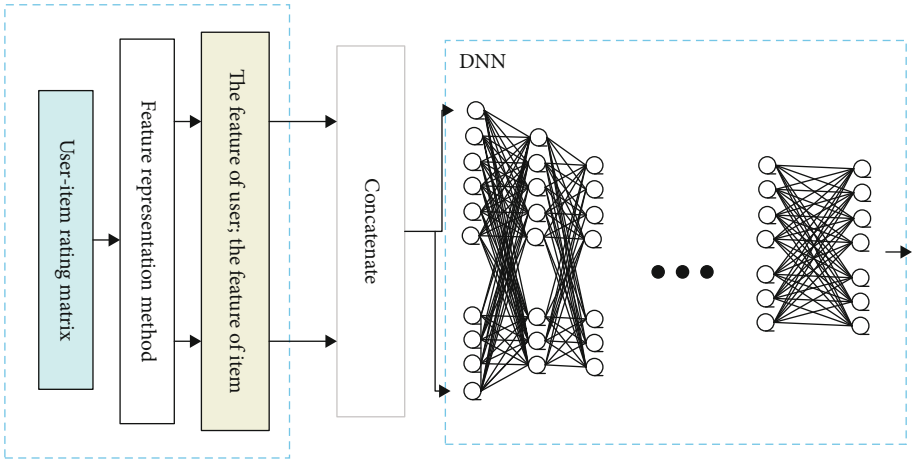

Figure 5: Deep learning neural network model.

neurons to a part of the input signal will deliberately ignore a large amount of complex information, thereby improving the learning accuracy. Sparsity has the ability to enhance the generalization ability of the network space and prevent overallocation, but excessive input will reduce the effective capacity of the neural network model, resulting in too few effective features for clear identification and classification.

At present, deep neural networks often use gradient descent methods for network optimization. Now, hot deep learning includes gradient descent algorithms. The implementation steps of gradient descent algorithms are as follows: (1) determine the learning rate; (2) given initial values; (3)) determine the downward direction and the reserved rate and update it; and (4) when the downward height is less than the defined value or the iteration condition is reached, the descent will stop. The gradient descent method formula is as follows:

$$\theta A(x) = A0 + A1x, \tag{13}$$

$$J(A0, A1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\theta A\left(x^{(i)}\right) - y^{(i)}\right)^2, \tag{14}$$

$$Aj \coloneqq Aj - \alpha_{\partial Aj}^{\partial} j(A0, A1), \tag{15}$$

$$tmp0 = A0 - \alpha_{\partial A0}^{\partial} j(A0, A1) \coloneqq tmmp0j = 0, \tag{16}$$

$$tmp1 = A1 - \alpha_{\partial A1}^{\partial} j(A1)A1 \coloneqq tmmp1j = 1. \tag{17}$$

Equation (17) is the weight update process, and equation (16) is the bias update process. $\theta A(x)$ represents a simple linear regression model. $j(A0, A1)$ is the loss function, used to evaluate the fit between the output and the expectation, m represents the training number set, and α is the learning rate. Convolutional neural networks will use parameter sharing, local cognition, pooling, and other methods to reduce the number of calculations. The convolution kernel and the image are only partially connected, rather than fully connected. Because the statistical properties of some parts of the image are the same as others, the learned features of one part can also be used on another part. Therefore, the convolution kernel can share the same set of parameters in the process of convolving the input. After convolution, the feature will be further compressed through the pooling operation. Commonly used pooling operations include maximum pooling and average pooling. Convolutional neural network combines the advantages of image processing and deep learning. It is a typical deep neural network, which can improve the accuracy of feature recognition and significant neural calculations. This network is mostly used in image recognition and recommendation systems and is particularly suitable for processing image recognition and classification.

2.4. Comparison of Recommendation Methods. This paper lists the content-based recommendation algorithm, the context-based collaborative filtering algorithm, and the neural network-based recommendation algorithm, and the three algorithms are compared here.

The deep learning-based data processing is the same as the content-based one in that features are extracted from audio metadata. This approach avoids the cold start problem but increases the data processing burden. Therefore, context-based recommendation algorithms such as collaborative filtering are more popular in the industry, as they are easy to process and deploy, do not rely on metadata, and keep up with users' popular preferences. In addition, the cold-start problem is a persistent problem in context-based recommendation algorithms, where new users cannot be recommended if they do not generate information or have too little information.

Extracting music features from audio signals can essentially represent the genre to which the music belongs and is more relevant to human intuition about music. Content-based and deep-learning-based recommendation algorithms have this advantage. However, due to the scalability of deep learning models, deep learning-based recommendation algorithms can be adapted to a wider range of platform environments and are highly scalable.

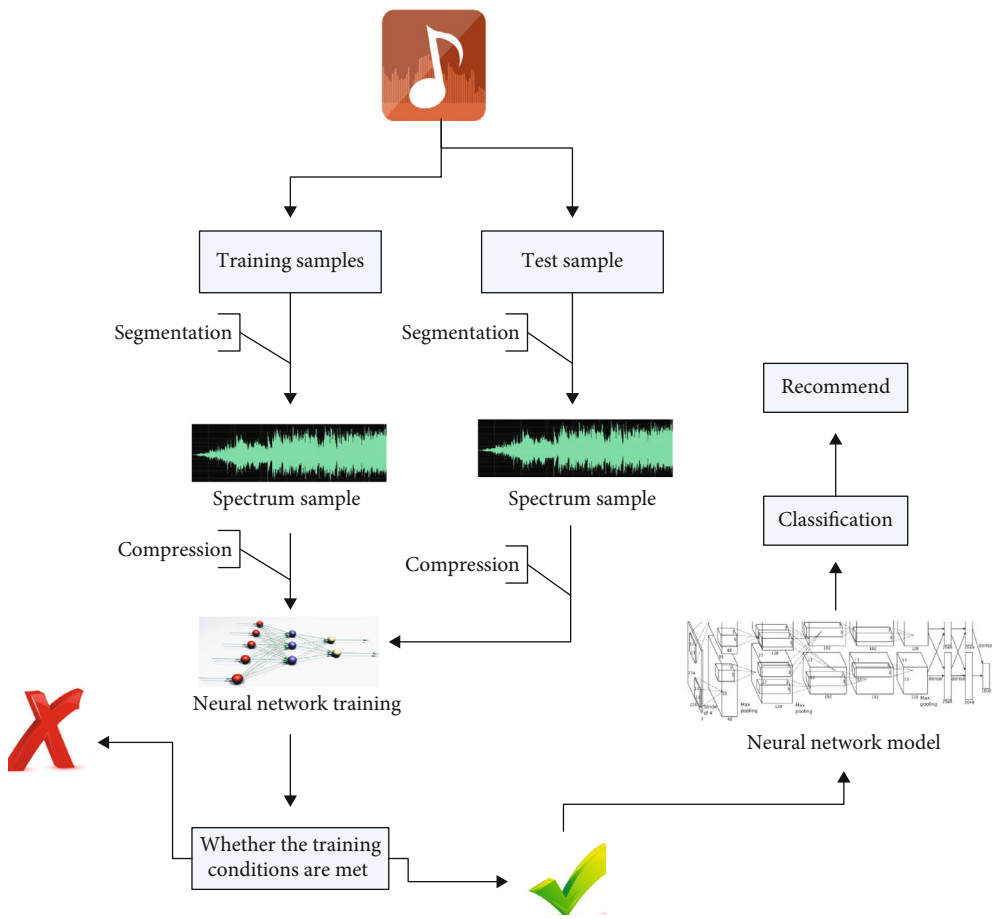

Figure 6: CNN training process flowchart.

3. Music Recommendation Model Based on Convolutional Neural Network

3.1. CNN Training Process and Experimental Environment. As shown in Figure 6, the general main process is divided into three parts: (1) the music data file is randomly divided into training samples and test samples, and the audio samples are segmented to generate spectrum data. (2) Compress the frequency spectrum data generated by the training sample processing and transfer it to neural network training. (3) Finally, the neural network model is used for classification and recommendation.

Due to various restrictions, the use of TensorFlow to train the model, the training state information comes from distributed machine learning, and execution on TensorFlow can simplify the creation, configuration, training, evaluation, and piloting of machine learning models. TensorFlow is easy to use. The training samples come from NetEase Cloud Music and Global Music Network. It was divided into four standard categories: classical, pop, rock, and pure music. 100 songs were selected for each, and then, random samples were selected. The music format was in mp3 format. The pixels of the divided spectrum segment are 128×128. Take 240 seconds of audio as a demonstration, broken down into about 93 fragments. The audio data was segmented, and more than 8,000 image samples were obtained from the fragments. 40% of the images are used as training examples, 30% are used as images, and 30% are used as measurement models.

The sample image is represented in grayscale, the x-axis represents time, the y-axis represents frequency, the grayscale represents the amplitude of frequency 29, and the free region (i.e.,, the larger the gray value) has higher amplitude.

Figure 7 shows four typical spectrum images. Each atlas represents a piece of music. From the atlas, you can see that the overall range of the "light music" gray scale is weak and the music is calm. The "classic" frequencies are high and the backs of the gray scale more intense, indicating relatively high notes and strong rhythms. The change in the width of the "rock" gray scale is particularly noticeable and atherosclerosis is bright. The frequency of "bursts" is usually low, the weaknesses are equal, and the sense of unity is strong.

After many adjustments in experiments, based on the classic convolutional neural network model, the final model is shown in Figure 8.

The model has 4 convolution+pooling layers, the neural network has 1024 neurons, and it uses fully connected training.

This experiment refers to other excellent experimental schemes of deep neural networks, namely, the back propagation algorithm of neural networks. BP multilayer perception algorithm is explored in relation to focusing on the study

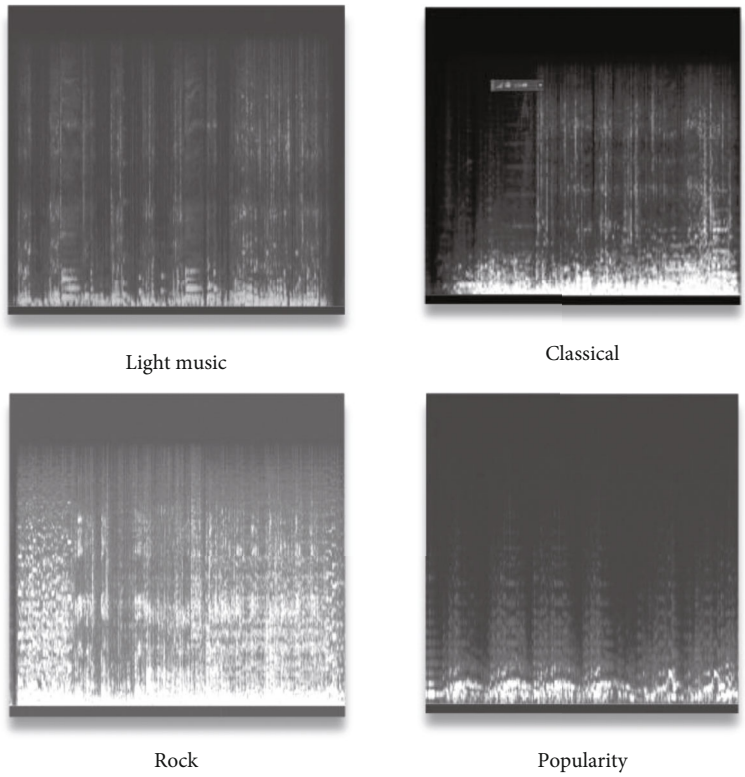

Figure 7: Spectral samples of four types of music.

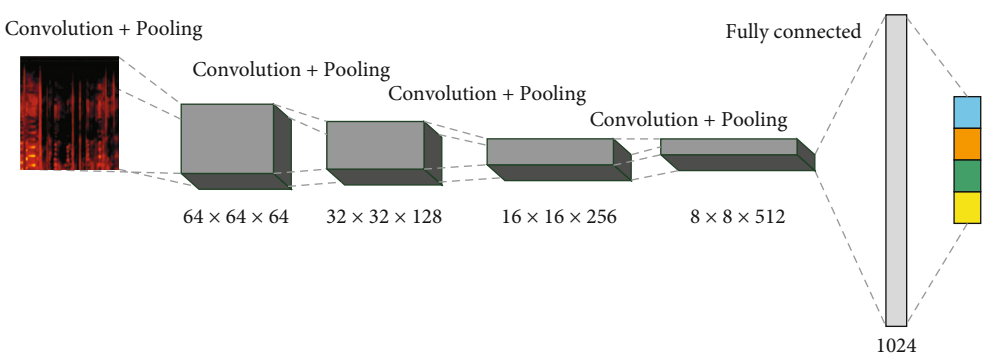

Figure 8: Convolutional neural network training model.

and comparison of two activation functions, linear unit and nonlinear unit. The controller compares the two gradient descent methods, Adam and RMSProp.

3.2. Comparison and Analysis of Activation Functions. After comparing multiple sets of activation functions, this article focuses on the comparison of ELU and ReLU activation functions. First, it is the most widely used ReLU, whose activation equation is

$$i(f) = \max\,(0, wf + a). \tag{18}$$

The ReLU activation functions because the derivative is a constant, so the calculation is very small, so the calculation is very efficient. Also because the derivative is constant, the error can move smoothly in the entire range, and there is no problem of gradient disappearance and gradient explosion. Also, the output portion of the neuron is set to 0. This improves network differences and at the same time reduces parameter dependencies. However, if the learning rate of the ReLU activation function is not set reasonably, and there happens to be a large gradient at this time, the neuron will be "paralyzed" at this time. The real-time performance is that the gradient of the neuron is displayed as 0 and no longer changes, so that the neural network no longer has the learning ability. In general, "paralysis" is only an individual phenomenon, not the whole, and has no effect on the entire network.

When the ELU activation function inputs x in the interval, when the derivative is equal to 1, the equation is as

$$i(f) = \begin{cases} f, & \text{if } f > 0, \\ \alpha(\exp(f) - 1, & \text{if } f \leq 0. \end{cases} \tag{19}$$

Compared with the ReLU activation function, the ELU activation function can take a negative value, which can make the average unit activation close to zero, reducing the amount of calculation while reducing the deviation value. The ELU mitigates the gradient dispersion problem to a certain extent and has a soft saturation property for small values of the input, which improves the robustness to noise.

3.3. Comparative Analysis of Optimized Control Modules. The optimal solution obtained by batch gradient descent (BGD) is the global optimal solution, but the process is particularly time-consuming because all the data in the training set are called for each training iteration, and the sample training speed is severely reduced if the number of samples is large.

The stochastic gradient descent (SGD) iterative training method draws only one batch of samples to update the parameters. Considering a single sample, the loss function corresponds to the granularity of each sample in the training set.

Compared with the batch descent method, the stochastic gradient descent method has a faster descent speed and has a good learning convergence effect under a large number of sample data. However, the gradient change in the stochastic gradient descent method only comes from the random part and cannot be considered as a whole. This also means that there is a gradient error, which will affect the irregular changes in the next iteration of the gradient descent. The iteration process is as follows:

$$J(A0, A1) = \frac{1}{2}\left(HA\left(x^{(i)} - y^{(i)}\right)^2,\right. \tag{20}$$

$$Aj := Aj - \alpha^{\partial}_{\partial Aj} j(A0, A1). \tag{21}$$

Based on the stochastic cathode gradient method, the RMSProp algorithm describes the concept of energy, the cumulative gradient is represented by w, and the changes are as follows:

$$w = pw + (1 - p)\widehat{g} \cdot \widehat{g}, \tag{22}$$

$$\nabla\theta = -\frac{\alpha}{\partial + \sqrt{w}} \cdot \widehat{g}, \tag{23}$$

$$\theta i = \theta i + \nabla\theta, \tag{24}$$

where p represents the attenuation rate, represents the$\widehat{g}$ of the gradient value, andδ represents the stability of the value. Plus the RMSProp algorithm can realize the automatic change of the learning rate. If the value of the gradient descent this time is relatively large, the decomposition of the learning rate needs to be accelerated; similarly, when the value of the gradient descent is small, the decomposition of the learning rate needs to be slowed down.

The other Adam, which also uses an adaptive learning rate method, is an optimizer based on the RMSProp algorithm. The average of the historical gradient ui is

$$mi = \beta 1 mi - 1 + (1 - \beta 1)\widehat{g}, \tag{25}$$

$$ui = \beta 2 ui - 1 + (1 - \beta 2)\widehat{g} \cdot \widehat{g}. \tag{26}$$

In this, mi represents the weight of the mean slope value and ui represents the weight equal to it. $\beta 1$ and $\beta 2$ are inefficient, but if the measure is still close to 1, mi and ui will be close to 0. The equation is

$$mi = \frac{mi}{1 - \beta 1}, \tag{27}$$

$$ui = \frac{ui}{1 - \beta 2}, \tag{28}$$

$$\nabla\theta = -\frac{\alpha}{\partial + \sqrt{ui}} \cdot mi. \tag{29}$$

Adam can further modify the RMSProp result, which is equivalent to adding another pulse compensation. Whether it is descent speed or accuracy, Adam's perfect controller is excellent.

4. Combining User Preference Characteristic System Experimental Model and Its Analysis Results

4.1. User Preference Feature Calculation. By combining the relationship between music and distribution and the relationship between music and users, you can obtain the relationship between users and the characteristics of distribution. Suppose that the relationship between a certain user and multiple pieces of music is $Su, a = [a1, a2, a3 \cdots am]$, and the optimization result of music classification is $Ca, b = [[b11, b12, b13 \cdots b1n], [b21, b22, b23 \cdots b2n] \cdots [bm1, bm2, bm3 \cdots bmn]]$, where Sj represents the degree of preference for a certain piece of music, $[bi1, bi2, bi3 \cdots bin]$ is the classification feature vector of a certain music, and the others are 0. When only three ratio values are not 0, the relationship of user preferences is

$$qu, c = Ca, b^t \times Su, a^t, \tag{30}$$

$$qu, c = \left[\sum_{i=1}^{m} Sjbi1, \sum_{i=1}^{m} a1bi2 \cdots \sum_{i=1}^{m} aibin\right]. \tag{31}$$

qu, c indicates the user's preference that describes the user's preference for music from the perspective of music characteristics. It has multiple uses in recommendation. (1) If it is a content-based recommendation method, the user's preference feature can be used as a typical feature. We can find which music is similar to the user's preference. (2) It can be used as a measure of the user. The similarity standard

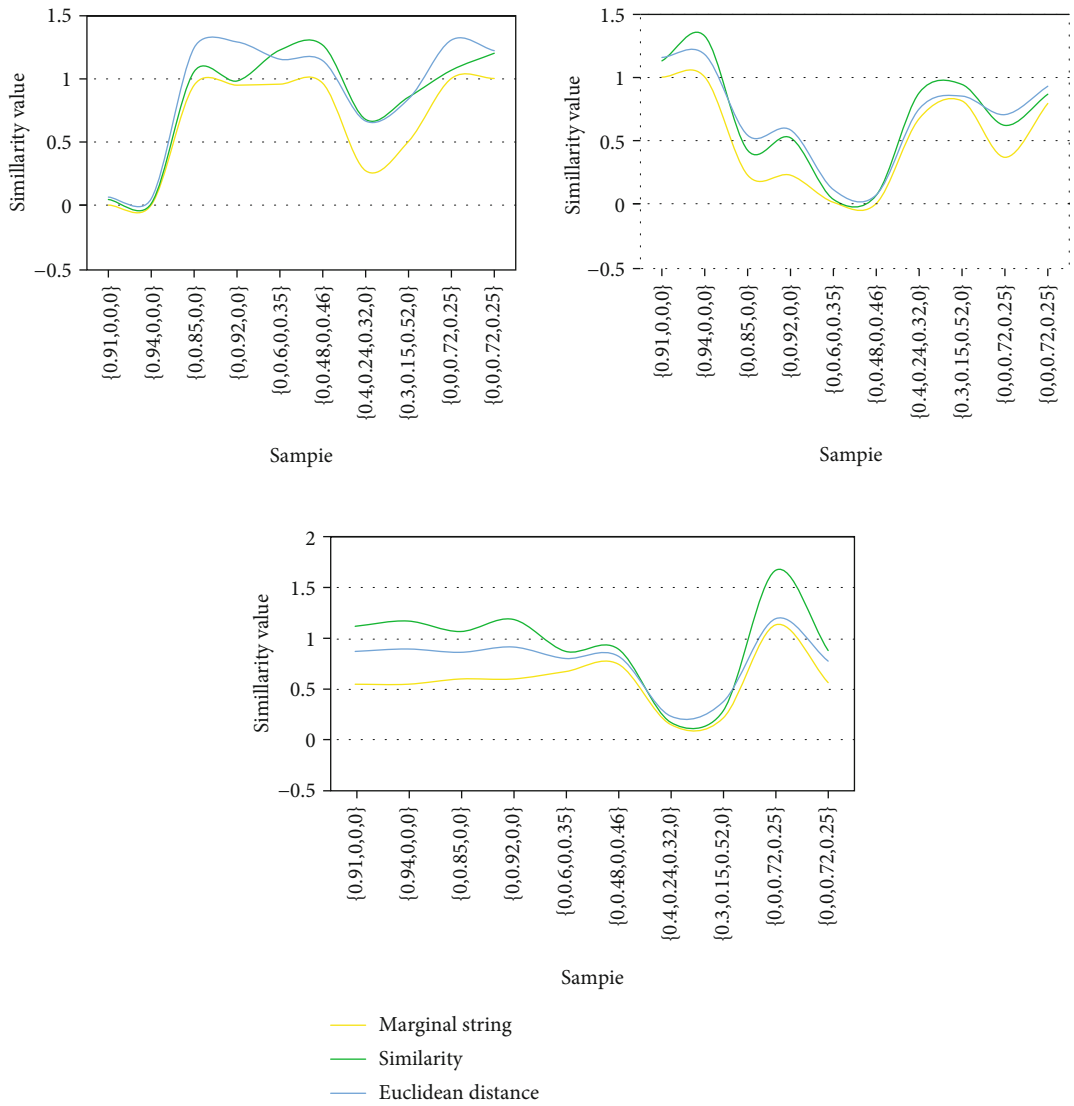

Figure 9: Comparison of three similarities of users.

Table 1: Classification state transition probability.

	Popularity	Rock	Classical	Light music
Popularity	0.768	0.024	0	0.25
Rock	0.05	0.833	0.068	0.05
Classical	0	0.034	0.833	0.134
Light music	0.406	0	0	0.594

can also improve the accuracy of the collaborative filtering recommendation algorithm. (3) After a large amount of user preference data statistics, the common preference characteristics of users can be found.

4.2. Similarity Analysis Combined with User Preference Features. When using the angular cosine method in this experiment, we found some problems that are not suitable for this research. But when the sample characteristics $A1 = \{0, 0, 0, 0.8\}$ and $A2 = \{0, 0, 0, 0.9\}$ and the user feature $P = \{0, 0, 0, 1\}$, the cosine calculation results are all 1. This is because cosine only calculates the included angle, and there is no comparison of vector lengths. In order to correct this problem, a vector ratio is added to the cosine:

$$ZA, P = 1 - \frac{\vec{A} \cdot \vec{P}}{\|A\| \cdot \|P\|} + \left|1 - \frac{\|A\|}{\|P\|}\right|. \tag{32}$$

Among them, A is the distribution feature vector of some music,A is the relationship vector between the user and the distribution, and it is the similarity between$A$ and P.

The three values in Figure 9 above are the cosine angle, ZA, P, and Euclidean distance in sequence. The smaller the value, the higher the similarity.

This experiment introduces Euclidean distance as a reference, and it is a random sample of 10 categorical variables, and the cosine and ZA, P similarity are calculated separately. In Figure 9, (1) user sample $y1 = \{0.94, 0.05, 0, 0\}$, (2) user

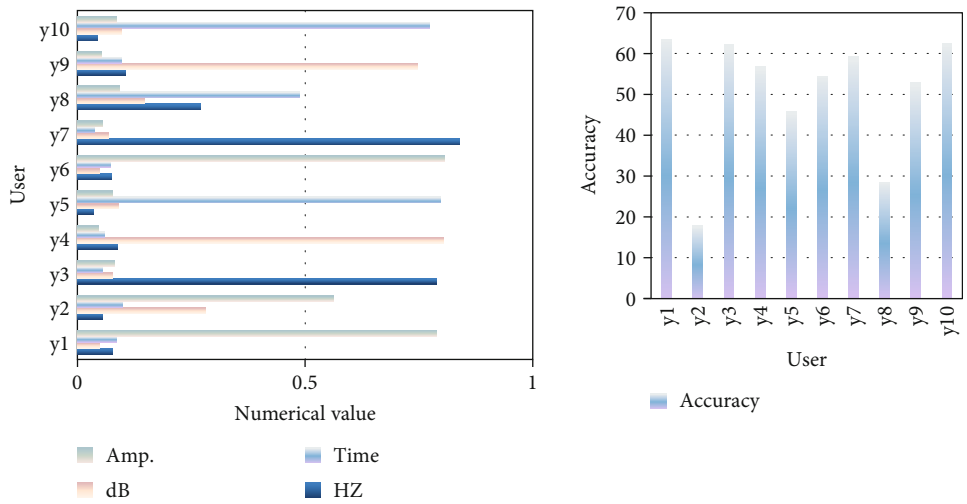

Figure 10: The recommendation result of comprehensive evaluation of user characteristics and the accuracy of recommendation.

Table 2: User characteristic multicategory evaluation recommendation table.

User	User characteristics	Main category	Accuracy
Y1	{0.074,0.039,0.085,0.726}	P4,p10,p8	33.9%
Y2	{0.098,0.72,0.091,0.05}	P2,p5,p8	45.6%
Y3	{0.719,0.093,0.043,0.072}	P7,p1,p5	41.2%
Y4	{0.07,0.041,0.078,0.75}	P4,p7,p9	40.7%
Y5	{0.28,0.131,0.462,0.076}	P6,p3,p8	35.3%
Y6	{0.082,0.739,0.052,0.053}	P2,p5,p8	51.4%
Y7	{0.032,0.08,0.742,0.08}	P3,p10,p8	48.3%
Y8	{0.046,0.278,0.095,0.528}	P9,p4,p10	49.5%
Y9	{0.778,0.063,0.031,0.054}	P1,p5,p7	43.6%

sample and $y2 = \{0, 0.55, 0, 0.5\}$, (3) user sample $y3 = \{0.35, 0.32, 0.35, 0\}$. The results showed that ZA, P can reflect the difference of classification characteristics; on the other hand, the difference of energy body value is close to the value of current Euclidean distance. But adding the same ZA, P to the modulus ratio reduces the similarity.

4.3. Comprehensive Evaluation and Multicategory Evaluation of User Characteristics. If the qu, c user only likes a piece of music, the user only has a one-dimensional vector image and can instantly trust the similarity between the vector image and file function. If the user likes some songs, the qu, c user represents the number of feature vectors. You can get the $\overline{qu, c}$ by calculating the average overlay effect on feature vectors. This may indicate that the user likes the full feature. This is a guide. It also supports the analysis of user characteristics and the calculation of similarity. The formula is

$$\overline{quc} = quc/m. \tag{33}$$

Suppose the classification features of the music you listen to are {0.8,0,0,0} and {0,0.9,0,0}; then, $\overline{qu, c} = \{0.4, 0.45, 0, 0\}$. User preferences are all single categories, and $\overline{qu, c}$ as the basis is indeed multicategory. A single pair of user preferences is not reasonable but should be recommended in categories. Based on the selection of music data, only two categories are considered in this paper, i.e., single category and double category, plus "other," for a total of 11 category distributions. If more music data were available, the category distribution could be larger. However, the research in this paper considers the short-term behavior of users, and users do not switch between multiple music genres in a short period of time. Therefore, the data of multiple categories are less.

The data sample of this experiment is 100 pieces of music. After the trained convolutional neural network model predicts the classification, 1000 classification features are obtained, and the classification is optimized. In this paper, the user's music list is obtained based on the hidden Markov's music classification model, such as formula (32), 10 songs for each user, as a reference for the results.

$$Xt + 1 = Xt \times Phmm. \tag{34}$$

As shown in Table 1, if the education distribution status is irreversible, it is the previous music distribution, the next one obtained after guessing, the music file distribution is compared to the music file distribution function, and the similarity is that I have. It has the following name.

In the recommendation method experiment combining the comprehensive evaluation of user characteristics, the article first calculates the mean value $\overline{qu, c}$ of the user's preference characteristics, as shown in Figure 10, to calculate the degree of ZA, P similarity. Randomly select 10 of which are less than 0.1 for recommendation. $\overline{qu, c}$ is the average characteristic of users. The recommended results of the user's comprehensive evaluation are as follows.

The y2 and y8 in Figure 10 belong to users with multiple categories, and the average hit rate is 50.35% through calculation. Experiments show that the recommendation accuracy rate of a single category in the average feature preference set

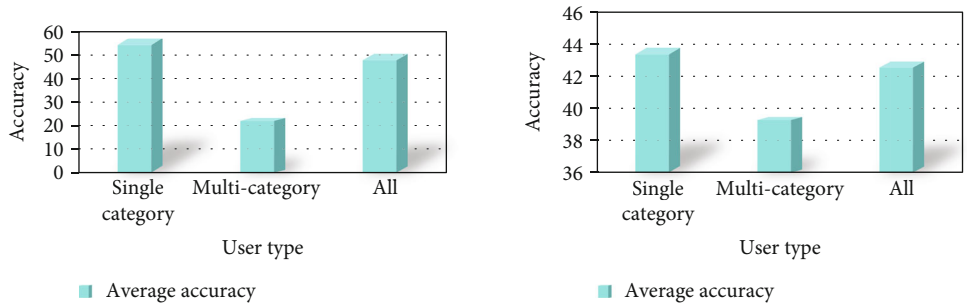

Figure 11: The average accuracy of the two recommended methods.

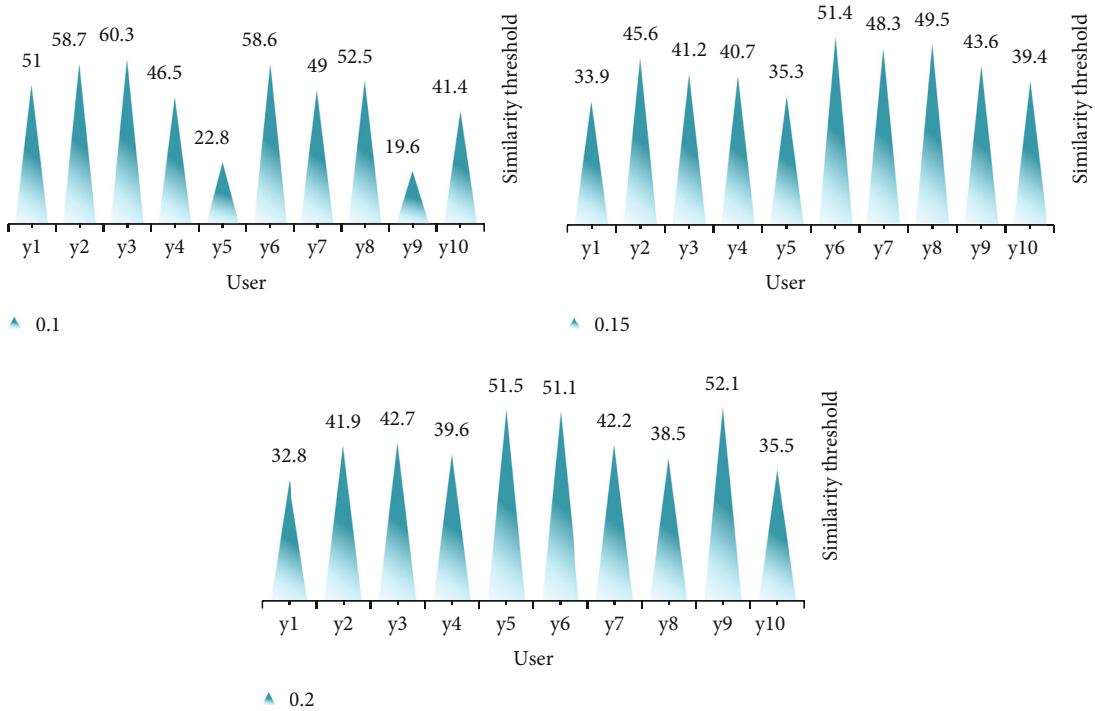

Figure 12: Recommended hit rates for different similarity thresholds.

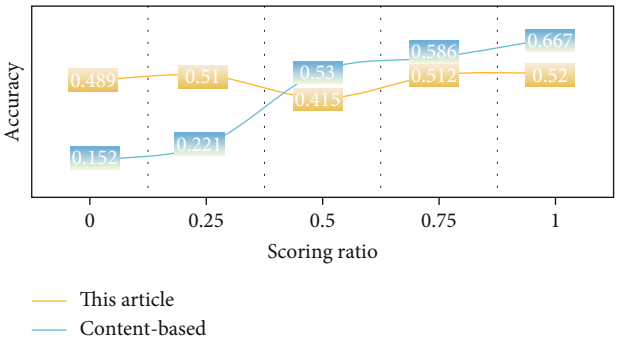

Figure 13: Comparison of different scoring ratios.

is higher than that of users with multiple categories, which may be the selected sample less, so the probability of randomness should be smaller. Of course, it is also possible that the user prefers a single category. It can be seen that the characteristics of preference are not very accurate.

The multicategory evaluation results of user characteristics are shown in Table 2.

The average hit rate calculated from Table 2 is 42.89%, and users y5 and y9 in the table are multicategory users. Because the results of multicategory similarity calculations are too large, this paper sets the similarity threshold to 0.15. The empirical results show that the recommendation rate of multicategory recommendation is generally lower than that of average feature, but average feature is better for multicategory users, because multicategory classification avoids a single expression of average feature.

The comparison of the recommendation results of the two recommendation methods for different user types is shown in Figure 11: (1) is a general evaluation of user characteristics and (2) is a comprehensive evaluation of user characteristics.

From the comparison in Figure 11, it turns out that the method of measuring user characteristics is generally better than the consent of different types of measurements. For an individual user, the process of generally evaluating user characteristics is superior to the process of approving multiple groups of user characteristics. The result is good. However, for multiuser groups, measurements of different types of user attributes are better than the recommended measurements.

In this experiment, it is found that different similarity thresholds have a significant impact on the multicategory evaluation recommendation method of user characteristics. When the maximum proportion of a piece of music exceeds the threshold, the music is considered as a single classification and the classification is based on the result of the voting method; if the maximum proportion is less than the threshold, the music is considered to contain multiple classification features. In this way, the predicted classification results of the CNN are divided by the threshold, and two results are obtained more accurately: one is a single classification, and the other is a multiple classification. Regardless of the classification result, it can be expressed as the music classification feature vector. The recommendation results under different thresholds are listed in Figure 12.

The user sample in Figure 12 uses the user characteristics of Table 2. From the graph results, when the similarity threshold is 0.1 or 0.15, the probability of multicategory users y5 and y9 is lower, but the similarity threshold is at 0.2; the recommendation accuracy rate of multicategory users y5 and y9 is much higher than the probability of the comprehensive recommendation evaluation method, indicating that its accuracy rate has improved. But for a single category, if the similarity threshold is too large, the accuracy will decrease. Therefore, the intermediate value 0.15 is the equilibrium result of the experiment.

In general, it turns out that measuring user characteristics is better than approving different types of measurements. For a single user, this process usually measures user characteristics better than the multiuser recognition process. The result is good. However, for many user groups, measurements for different behavioral uses are better than the recommended measurements. In addition to music preferences, users' behavior preferences should also involve users' personal information, search records, favorites lists, etc. Compared with the traditional content-based music recommendation method, this paper uses the convolutional neural network classification model to obtain the feature vector of music, which simplifies the processing of massive audio data and the complex audio feature extraction process. The user characteristics are also considered in the recommendation process. difference. However, the method of musical note recognition is currently more suitable for pure music recommendation.

The content-based convolutional neural network recommendation algorithm also uses audio feature training, but different from this method is that based on the content, a large number of user scores are required to predict the popularity of music, and then, recommendations are based on the popularity. This article compares the recommended accuracy rates under different scoring ratios as shown in Figure 13.

It can be seen from the results that the recommendation algorithm based on the content-based convolutional neural network that relies on user evaluation data has a low recommendation accuracy rate during cold start, but the recommendation accuracy rate begins to increase when user evaluation data is added. Since the method in this paper does not rely on user scoring, the accuracy of recommendation does not change much. Therefore, this method has certain use value in the absence of user scoring data for recommendation.

5. Conclusions

During the analysis, this task finds data and classification errors in the content delivery and advises the model to optimize them. The results can be used as a vector representation of the distribution of music features and can be combined with music user settings to calculate the user experience. This paper presents a comprehensive evaluation of user characteristics and suggestions for measuring different types of user characteristics based on the process of calculating differences in characteristics and compares the two methods. Experimental results show that a good rationale for a user-suggested group is the extensive evaluation of user features, and the logic for various user feature metrics is appropriate for approving multiple user groups. On the whole, the recommendation result of comprehensive evaluation of user characteristics is better than recommendation results of multicategory evaluation of user characteristics. This paper proposes two recommendation methods based on user preference characteristics. The experimental results show that the two recommendation methods can achieve better recommendation results. In the recommendation method of comprehensive evaluation of user characteristics, the recommendation accuracy rate of single-category user characteristics is higher than that of multicategory user characteristics, with an average accuracy rate of 50.35%, while in the recommendation method of multicategory evaluation of user characteristics, the accuracy of feature recommendation of multicategory users is higher than that of single-category user features, with an average accuracy of 42.89%. Comparing the results of the content-based convolutional neural network recommendation algorithm shows that although the recommendation accuracy rate of this method is not high, it has certain advantages in cold start and can also have better recommendation results.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

References

[1] Y. T. Chen, C. H. Chen, S. Wu, and C. C. Lo, "A two-step approach for classifying music genre on the strength of AHP weighted musical features," *Mathematics*, vol. 7, no. 1, p. 19, 2019.

[2] Y. Y. Liu, B. Yang, H. B. Pei, and J. Huang, "Neural explainable recommender model based on attributes and reviews," *Journal of Computer Science and Technology*, vol. 35, no. 6, pp. 1446–1460, 2020.

[3] N. L. Anantha and B. P. Battula, "Deep convolutional neural networks for product recommendation," *Ingenierie des Systemes d'Information*, vol. 23, no. 6, pp. 161–172, 2018.

[4] D. Son and K. Shim, "Improving on matrix factorization for recommendation systems by using a character-level convolutional neural network," *KIISE Transactions on Computing Practices*, vol. 24, no. 2, pp. 93–98, 2018.

[5] Z. Zhang, H. Pan, and G. Xu, "Research on personalized tourism attractions recommendation model based on user social influence," *Advances in Modelling and Analysis B*, vol. 59, no. 1, pp. 62–75, 2016.

[6] Z. Bi, L. Jing, M. Shan, S. Dou, and S. Wang, "Hierarchical social recommendation model based on a graph neural network," *Wireless Communications and Mobile Computing*, vol. 2021, 10 pages, 2021.

[7] J. Lee, K. Yoon, D. Jang, S. J. Jang, S. Shin, and J. H. Kim, "Music recommendation system based on genre distance and user preference classification," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 5, pp. 1285–1292, 2018.

[8] Y. Yan, G. Yu, and X. Yan, "Online doctor recommendation with convolutional neural network and sparse inputs," *Computational Intelligence and Neuroscience*, vol. 2020, 10 pages, 2020.

[9] Q. Wang and Z. Mu, "Application of music in relief of driving fatigue based on EEG signals," *EURASIP Journal on Advances in Signal Processing*, vol. 2021, 89 pages, 2021.

[10] C. L. Liu and Y. C. Chen, "Background music recommendation based on latent factors and moods," *Knowledge-Based Systems*, vol. 159, no. NOV.1, pp. 158–170, 2018.

[11] X. Du, X. He, F. Yuan, J. Tang, Z. Qin, and T. S. Chua, "Modeling embedding dimension correlations via convolutional neural collaborative filtering," *ACM Transactions on Information Systems*, vol. 37, no. 4, pp. 1–22, 2019.

[12] H. Wei and N. Kehtarnavaz, "Determining number of speakers from single microphone speech signals by multi-label convolutional neural network," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, Washington, DC, USA, Oct. 2018.

[13] A. Aurisano, A. Radovic, D. Rocco et al., "A convolutional neural network neutrino event classifier," *Journal of Instrumentation*, vol. 11, no. 9, p. P09001, 2016.

[14] Satellite, Venture, Business, Laboratory, Muroran, & Institute, "Music artist recommendation system based on user's tastes," *Joho Chishiki Gakkaishi*, vol. 15, no. 3, pp. 59–70, 2016.

[15] Z. Cheng and J. Shen, "On effective location-aware music recommendation," *ACM Transactions on Information Systems*, vol. 34, no. 2, pp. 1–32, 2016.

[16] H. G. Kim, G. Y. Kim, and J. Y. Kim, "Music recommendation system using human activity recognition from accelerometer data," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 3, pp. 349–358, 2019.

[17] C. H. Chen, "An arrival time prediction method for bus system," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4231-4232, 2018.

[18] 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.

[19] R. Wang, X. Ma, C. Jiang, Y. Ye, and Y. Zhang, "Heterogeneous information network-based music recommendation system in mobile networks," *Computer Communications*, vol. 150, pp. 429–437, 2020.

[20] Q. Lu and F. Guo, "A novel e-commerce customer continuous purchase recommendation model research based on colony clustering," *International Journal of Wireless and Mobile Computing*, vol. 11, no. 4, pp. 309–317, 2016.

[21] H. Pan and Z. Zhang, "Research on context-awareness mobile tourism e-commerce personalized recommendation model," *Journal of Signal Processing Systems*, vol. 93, no. 2-3, pp. 147–154, 2021.

[22] A. Schindler and A. Rauber, "Harnessing music-related visual stereotypes for music information retrieval," *ACM Transactions on Intelligent Systems& Technology*, vol. 8, no. 2, pp. 1–21, 2016.

[23] T. M. Phuong, T. C. Thanh, and N. X. Bach, "Neural session-aware recommendation," *IEEE Access*, vol. 7, pp. 86884–86896, 2019.

[24] X. Li, H. Liu, W. Wang, Y. Zheng, H. Lv, and Z. Lv, "Big data analysis of the internet of things in the digital twins of smart city based on deep learning," *Future Generation Computer Systems*, vol. 128, pp. 167–177, 2022.