

Research Article

Personalized Song Recommendation System Based on Vocal Characteristics

Jingzhou Yang 

Zhengzhou University of Light Industry, Zhengzhou 450002, China

Correspondence should be addressed to Jingzhou Yang; yuchuzhisheng@zzuli.edu.cn

Received 14 January 2022; Revised 18 February 2022; Accepted 22 February 2022; Published 16 March 2022

Academic Editor: Naeem Jan

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

To find favorite songs among massive songs has become a difficult problem. The song recommendation algorithm makes personalized recommendations by analyzing user's historical behavior, which can reduce user's information fatigue and improve the user experience. This paper studies a personalized song recommendation algorithm based on vocal features. The specific work includes three parts. Firstly, the spectrum feature extraction and observe feature extraction of songs. The spectrum includes three types of features: time domain, frequency domain, and amplitude, which implicitly describe the rhythm, notes, and high-pitched or soothing properties of songs. Furthermore, automatic note recognition methods are explored as explicit classification features. The characteristic of this work is to use the comprehensive features of spectrum and musical notes as the classification basis. Secondly, based on song of convolutional neural network (CNN) classification, it sets different types of song classification. For the training of CNN, ELU, and ReLU, RMSProp and Adam were explored, and their performance and characteristics during training were compared. The classification methods were compared under the two configurations based on the spectrum as the classification basis and the comprehensive characteristic frequency of the spectrum and the note as the basis. Thirdly, a personalized song recommendation method was based on CNN classification. Also, the reasons why classification of CNN is not suitable for direct song recommendation are analyzed, and then, a recommendation method based on song fragment classification is proposed. A threshold model that can distinguish between pseudodiscrete and true-discrete is proposed to improve the accuracy of song classification.

1. Introduction

The amount of information available to us is expanding at a breakneck pace. A person's ability to accept and process the diversity of information available on the Internet is severely limited. Access to knowledge is a basic human right, and the problem of information oversaturation must be addressed if we are to continue to thrive and advance as a species. Prior to the introduction of recommender systems, search engines were a primary method for addressing the problem of information overload. However, a search engine is rendered useless if the user's information requirements are unclear or if the user is unable to explain his or her wants effectively. Locating personal demand data in the network quickly and accurately has become a challenging problem that must be addressed in the real world. As time has progressed,

personalized recommendation systems have evolved as a viable option for Internet-based intelligent products. The goal of a personalized recommendation system is to learn about a user's preferences, interests, and other relevant data in order to tailor services to meet their needs [1–5].

Spiritual fulfillment becomes increasingly important as one's wealth increases. In entertainment ventures, songs have always been an integral component of the experience. In the past, individuals relied primarily on cassettes and vinyl for their entertainment needs. However, with the shift to digital song consumption, the problem of information oversaturation has emerged. More than 60 billion songs exist in the globe, and they are growing at a rate of around two songs each second. Customers will not be able to accept the big song data if a business lists all the songs on a music website. As a result, people will experience information

overload because this is much above their capacity to consume and is therefore inescapable. An overwhelming volume of unrelated data will get in the way of people's attempts to find their favorite songs. Using this method will take a lot of time for users, as well as causing them to suffer in anguish and agony on a regular basis. This problem can be alleviated by using on-site searches on many music sites. As a result, the website's overhead will increase, and the user's pleasure will be negatively affected as a result of this passive and passive solution. It is critical that the music library's personalized recommendation system be implemented so that users may find great songs faster and more correctly [6–10].

In the context of big data, standard recommendation algorithms, calculation methods, storage methods, and so on cannot handle the exponential development in data requirements and individualized user needs. As a result of the challenges presented by big data, the recommendation system's performance will be substantially hampered. As a result, changing the way songs are recommended has become an unavoidable tendency. The goal is to completely exploit the value of data, enhance the quality of recommendation results and the efficiency of calculation, and alleviate the problem of information overload even more. Recommendation systems face new challenges in the big data environment. This mainly includes the following: (1) more data will need to be analyzed in the recommendation process. The multidimensional big data of users brings high-dimensional sparsity problems, and at the same time, there will be more redundancy and noise in the data. (2) Data collection needs to cover explicit data and implicit feedback data, and the combination of the two data can better improve the performance of the recommender system. (3) Faster and more new data is generated, which has higher requirements on the data processing capability of the recommendation system. (4) Massive data puts forward new requirements for data storage. Most of the traditional databases are relational, which cannot meet the requirements of existing recommendation systems and have high risks [11–15].

Scholars at home as well as abroad have paid close attention to song recommendation and have made many research results, and many successful products have appeared in the industry. There are famous music websites such as Pandora abroad, NetEase Cloud, and Douban Radio in China, but these products are also seeking changes in the big data environment. Therefore, based on practical application, this paper studies the technical difficulties and problems faced in specific applications. The song personalized recommendation system can provide users with satisfactory personalized services, meet the various needs of different users, and can effectively adjust and respond according to the user's personal situation. Song recommendation is a very special field in the recommendation system, and the research value and practical significance of this paper are very high.

The arrangement of the paper is as follows: Section 2 describes the related work of the research. Vocal features extraction on songs is completely described in Section 3. This

section has some subsection to thoroughly explain the features. Song's classification based on CNN model is completely explained in Section 4 that is easy to understand. Section 5 talks about song recommendation combined with short-term user behavior. Section 6 conversed about experiments that made the research strong and discusses the results. Section 7 concludes the paper.

2. Related Work

T-Methods like amatrix decomposition and user feedback are laboriously incorporated into the collaborative filtering process. In most cases, the way in which people interact with music takes on a variety of shapes and forms, such as through their search habits, listening logs, tags, and so on. The history of songs is key implicit feedback because the system automatically collects them. Collaborative filtering usually decomposes explicit information such as user ratings for items or rating matrices obtained by text processing of evaluation contents for recommendation [16]. However, the collaborative filtering method can also use the processed implicit feedback to make recommendations, so that the recommendation results have high accuracy and bring high satisfaction to users. For the purpose of discovering possible connections between listeners and musical elements like bands and instrument genres, [17] recommended the implementation of collaborative filtering. According to [18], one must meet a certain standard. Users' listening frequency is converted into a score according to an established threshold, and the approach recommends songs to users based on their similarity.

The content-based recommendation method has been used in the fields of information retrieval and information filtering and is widely used in text classification and other fields. The processing and use of musical features are now replaced by tags. Tags are used in various fields [19], like tag-based literature retrieval [20], tag-based image processing [21], and tag-based music information retrieval [22]. Tags reflect essential characteristics for music to a large extent. At home and abroad, many researchers have integrated tags into music recommendation. Reference [23] downloads tags from social media sites and uses an LDA model to predict the next song. The topics in the LDA are composed of different tags, and the corresponding tags are screened out according to the predicted next topic, and song recommendations are made to the target users. Reference [24] proposes automatic audio classification based on the user's labeling behavior for music. Reference [25] augments the entire model with the labels of singers, uses a community detection algorithm to divide all music into different preference communities, and then constructs a recommendation list based on the labels of each subcommunity. Reference [26] describes the user's preference according to the label and calculates the similarity between users on this basis.

The hybrid recommendation method can make up for the above two methods, and the fusion of multiple methods can make up for the shortcomings of a single method, and the performance of the recommendation has been significantly improved. Reference [27] infers the user's song genre

preference according to the list name of the user's listening songs and generates a music recommendation list. Reference [28] proposed a car music recommendation. Reference [29] and others proposed similar methods for different environments, using natural environments such as weather, temperature, and lighting, and some labeled music matching recommendations. The abovementioned music content characteristics and context can be integrated into the recommendation model, which greatly improves the performance of the recommendation algorithm. According to the calculated music preferences of each user, there are great differences in the diversity and satisfaction of recommendations made to different users. Current research on users' music preferences has also achieved great success. These user music preference vectors include novelty [30], diversity [31], and mainstream [32]. Reference [33] studied different features and proposed a simple linear combination of multiple features to make recommendations.

3. Vocal Feature Extraction of Songs

There are many voice audio elements that may be extracted from a song's content by analyzing its content. Further analysis based on these variables can yield high-level description forms such as rhythm, melody, emotion, and musical notes. This paper analyzes the time-domain features of songs, obtains frequency domain features and amplitude features through fast Fourier transform and generates frequency spectrums based on these three-dimensional features. This paper also improves the traditional method of extracting notes by autocorrelation function, which improves the accuracy of note recognition, and combines note features into the spectrum to obtain note spectrum samples.

3.1. Spectral Representation and Extraction Methods of Songs.

The frequency domain is a description of the frequency characteristics of the signal. The frequency domain coordinate system takes the frequency value as the abscissa and the ordinate as the statistical quantity of each frequency value. Frequency domain analysis is an important part of signal analysis, mainly used in electronics, acoustic signals, and other fields. Digital audio information, which is generated by sampling the analog audio data, is now stored in songs on the network. In terms of digital audio quality, the higher the sampling frequency, the more digital audio data there is to work with. Digital audio can also be divided into tonal components of different frequencies because it is a form of sound wave signal. This shows that the frequency domain contains numerous sinusoidal functions, each with different amplitude and phase, indicating that the sound is full with information.

The direct expansion of the energy amplitude over time is the trend in the digital audio information encoded in the song. Audio signals can be represented in the time domain, which is the most obvious way to understand the signal. Both time and frequency domain analyses take into account signal properties. In the time domain, digital audio has a lot of data. The more data collected at a given time with a greater sample frequency necessitates more calculations. Frequency domain

analysis has a lesser data volume than time domain analysis, but it can better capture some important properties. This has led to a progressive shift in signal analysis toward frequency domain methods. If you want to characterize an audio signal, you will need to know what is called a frequency domain feature. To represent music, frequency domain features are straightforward to apply and limit the quantity of data that must be processed, making it more efficient.

Fourier transform is one of the important basic algorithms of frequency domain analysis. This algorithm uses sine and cosine functions to fit time domain signals, and the characteristic information carried by each sine and cosine function is frequency domain characteristics. At first, the computational load of Fourier transform was still very large, and many subsequent algorithms have been greatly improved on this basis, such as fast Fourier transform, which greatly saves calculation time while maintaining frequency domain accuracy.

Spectrum is a data representation of frequency domain features obtained by Fourier transform of audio time domain features, and the data comes from the periodic frequency of the fitted sine and cosine functions. The abscissa of the spectrum represents the frequency scale, and the ordinate represents the number of frequencies corresponding to the scale value within the statistical time range. To a certain extent, this form of expression can express the changing characteristics of musical tones from the side, and the amount of data can be effectively reduced by compressing the frequency range.

3.2. Note Feature Extraction and Improved Algorithm.

The musical tone signal of a song is composed of the fundamental tone and the overtone, and the fundamental tone determines its pitch. Therefore, the detection of the fundamental tone period is the key to the identification of the song notes. The pitch detection method based on the autocorrelation function is a classic time domain detection algorithm. The algorithm is simple, but there are pitch frequency octave or half frequency errors. The short-term autocorrelation function is

$$R_i(k) = \sum_{m=1}^{N-m} y_i(m)y_i'(m+k). \quad (1)$$

On this basis, it is a classic improved algorithm to perform three-level center clipping operation before calculating the autocorrelation function:

$$y_i'(x) = \begin{cases} 1, & |z_i(x)| > C_L, \\ 0, & |z_i(x)| < C_L, \\ 1, & \text{others.} \end{cases} \quad (2)$$

Ideally, after trilevel clipping and autocorrelation function calculations for a data frame of about 70% of the notes, the first maximum peak point is exactly the peak point corresponding to the pitch period. In a few cases, the signal is affected by the formant, and there will be interference of frequency doubling waves, resulting in the shift of the maximum peak point.

In response to this problem, this paper proposes to use the data frame translation to estimate the peak ratio and use this to select the correct peak point. Shifting the data frame refers to increasing the upper and lower bounds of the selected signal interval by 64; that is, the shifted signal interval is

$$s(i) = w[S(i) + 64, S(i) + 64 + \text{len}]. \quad (3)$$

3.3. Spectrum and Note Spectrum Generation. In this paper, the SoX tool is used to generate multidimensional spectrum. SoX, the full name of Sound exchange, is a well-known foreign open-source audio processing software and the most famous open-source sound file format conversion tool. It is widely used in the field of acoustic processing and has been widely ported to multiple operating system platforms, with strong compatibility. Sox can process sound files in a variety of formats and can also perform common audio signal processing such as sound filtering and sampling frequency conversion. This tool is highly integrated and uses the latest processing algorithms, making it ideal for audio signal researchers.

In this experiment, the music spectrum is generated using the SoX command line tool. This tool can automatically complete the cutting and spectral drawing of large amounts of audio data. Spectrograms are presented as portable network graphics files and display time on the x -axis, frequency on the y -axis, and audio signal amplitude on the z -axis. The magnitude of the z -axis is represented by the color of the pixels in the XY plane. If the audio signal contains multiple channels, the channels are represented from top to bottom, starting with channel 1. In this way, a two-dimensional grayscale image can fully represent the multidimensional features of audio data.

Note spectral samples are superimposed note features in the upper half of spectral samples divided into $128 * 128$. The specific operation is to take the pitch frequency of the first 128 notes of each song as the note feature and draw from top to bottom and left to right. The pitch frequency of the notes is vertically compressed to 128 levels, and the gray scale is also 256 levels.

4. Song Classification Based on CNNs

CNN is a kind of artificial neural network, which is often used in the field of image recognition. In this paper, the spectrum of each classified song is used as the input image of the CNN, and the classification of songs is indirectly realized through image recognition.

4.1. CNN Training Model Design. LeNet-5 is a relatively classic CNN model, which is used in a variety of image recognition scenarios. The training model in this paper is based on LeNet-5 and refers to other excellent CNN models and has been adjusted many times during the experiment. The final model is shown in Figure 1. The model has 4 convolution and 4 pooling layers, and the neural network has 1024 neurons, which are trained with full connections.

The parameter initialization operation adopts random initialization, the weights are randomly initialized, the biases are all initialized to 0, and the learning rate adopts the empirical value of 0.001.

4.2. Training and Classification. After the training model is designed, the training scheme needs to be determined. This scheme refers to other excellent deep neural network experimental schemes. In the experiment, we focus on two kinds of activation functions and compare them, ELU and ReLU, and the controller compares Adam and RMSProp gradient descent methods.

ReLU is a very efficient and widely used activation function:

$$\text{ReLU}(x) = \max(0, x). \quad (4)$$

ReLU needs to set a reasonable learning rate. If the learning rate is set unreasonably, it will cause the gradient to fluctuate violently. If a large gradient is generated during this period, the neurons are likely to be paralyzed due to overstimulation and lose the ability to perceive small gradients. The specific performance is that the gradient of the neuron does not change and remains at 0. In the general training process, the paralysis of neurons is only an individual phenomenon, which has certain contingency and does not affect the overall learning ability of the network. Of course, if the learning rate is set unreasonably, this will cause most of the neurons in the network to be paralyzed, and eventually the neural network will no longer have the ability to learn.

ELU alleviates the gradient dispersion problem to some extent by taking the input x itself in the positive interval:

$$\text{ELU}(x) = \begin{cases} x, & x > 0, \\ \alpha(\exp(x) - 1), & \text{others.} \end{cases} \quad (5)$$

Compared with ReLU, ELU can take negative values. This allows the activation mean of the unit to be closer to 0, which reduces the amount of calculation while reducing the variation of the bias value. When the input takes a small value, it has the characteristics of soft saturation, which improves the robustness to noise.

5. Song Recommendation Combined with Short-Term User Behavior

Although the accuracy of the CNN classification model is high, there are some errors that can be optimized. On the basis of statistics and analysis, this paper further processes the classification results, optimizes the classification results, and obtains the classification characteristics of songs. According to the classification features and user behavior data, this paper calculates the user's behavior preference characteristics and proposes two recommendation methods on this basis.

5.1. Recommended System Overview. The CNN classification model is used to classify the music spectrum and obtain the basic feature vector. By optimizing the feature vector of the

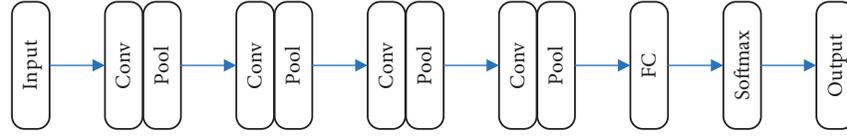


FIGURE 1: CNN model.

song, a relatively accurate song feature library can be constructed. The feature library can be used not only to measure the similarity between songs, but also to calculate user preference features. If the song that the user prefers already exists in the song library, the feature library can be directly searched. If it does not exist, you need to generate spectral samples, use the CNN classification model to predict and classify, and get feature vectors. Based on classification features, this paper designs a recommender system as shown in Figure 2.

The CNN classification model is a trained CNN, which can predict and classify the spectrum of the song and obtain the classification feature vector. User feature calculation is based on the relationship between steel songs and classification features, combined with the relationship between songs and users, to obtain the relationship between users and classification features.

5.2. Classification Optimization. Although the classification accuracy of CNN is high, the classification error cannot be ignored. After the statistical analysis of the classification error, it is found that the pseudodiscrete belongs to a single classification, and the smaller classification proportion can be ignored. The true discrete belongs to a variety of classifications, and a larger proportion of several classifications can be retained. Therefore, distinguishing between false discrete and true discrete can solve the first classification error.

Variance can generally be used to evaluate the degree of dispersion of a set of data, but it is not suitable for the data in this experiment. The reason is that when the number of classifications is larger, the variance will be smaller, so that the degree of dispersion cannot be accurately judged. Therefore, based on the voting method, this paper proposes a threshold formula to evaluate the discrete degree of music classification features, that is, whether music is pseudodiscrete or true discrete:

$$T = \min\left(\frac{a+1}{N}, 0.9\right). \quad (6)$$

When the maximum proportion of a song exceeds the threshold, the song can be considered a single classification, and the classification is subject to the result of the voting method. If the maximum ratio is less than the threshold, the song is considered to contain multiple categorical features. In this way, the predicted classification results of CNN are divided into two more accurate results after threshold division. One belongs to a single category and one belongs to multiple categories. Regardless of the classification result, it can be expressed as the classification feature vector of the song.

5.3. User Preference Feature Calculation. According to the relationship between the song and the classification feature, combined with the relationship between the song and the user, the relationship between the user and the classification feature, that is, the user preference feature, can be obtained. Suppose the relationship between a user and multiple songs is $L_{u,s} = [s_1, \dots, s_m]$, and the classification optimization result of each song is $R_{s,c} = [c_{11}, \dots, c_{1n}], [[c_{21}, \dots, c_{2n}], [c_{m1}, \dots, c_{mm}]]$. Among them, s_i represents the user's preference for a certain song, and $[c_{i1}, \dots, c_{in}]$ represents the classification feature vector of a certain song, of which only three proportions are nonzero, and the rest are zero. Then the user preference feature relationship is

$$P_{u,c} = L_{u,s} \times R_{s,c}, \quad (7)$$

where $P_{u,c}$ represent the user's preference feature, which is to describe the user's preference for songs from the perspective of song features, and it has multiple uses in recommendation. First, if the recommendation is based on content, $P_{u,c}$ can be used as typical features of user preferences, and by calculating the similarity between $P_{u,c}$ and song feature vectors, we can know which songs are similar to user preferences. Second, $P_{u,c}$ can also be used to measure user similarity, thereby improving the accuracy of collaborative filtering recommendations. Third, combined with a large number of users $P_{u,c}$ data, the common preference characteristics of users can be obtained.

5.4. Recommendation Combined with User Preference Features. The angle cosine can be used to calculate the similarity between multidimensional feature vectors, but the cosine formula only calculates the angle of the vector and does not compare the length of the vector. In order to obtain a more accurate similarity, this paper increases the vector modulus ratio on the basis of the cosine formula:

$$S_{A,B} = \left| 1 - \frac{\vec{A} \cdot \vec{B}}{AB} \right| + \left| 1 - \frac{A}{B} \right|. \quad (8)$$

If there is only one song preferred by the user, then user $P_{u,c}$ has only one dimension of feature vector, and the similarity between the feature vector and the feature library can be directly calculated. If there are several preferred songs, then user $P_{u,c}$ represents the superposition effect of multiple eigenvectors, and all eigenvectors of $P_{u,c}$ can be averaged to obtain $\overline{P_{u,c}}$. $\overline{P_{u,c}}$ can represent the user's preference characteristics as a whole, which is equivalent to a comprehensive evaluation of user characteristics.

The preferences of users are all a single category, while the recommendation basis $\overline{P_{u,c}}$ is multicategory. Therefore, it

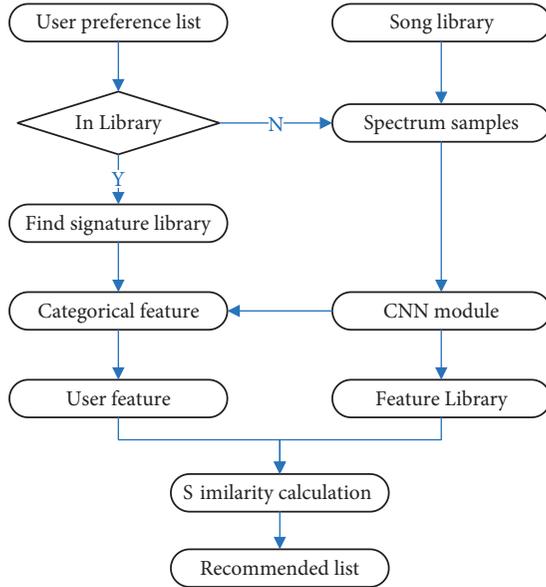


FIGURE 2: Recommended system.

is unreasonable to simply average user preferences, but it should be averaged by category and recommended by category, and the number recommended by a certain category is consistent with the number of that category.

After division, $P_{u,c}$ are calculated according to these 11 category distributions. The recommendation algorithm for multicategory evaluation of user characteristics is to classify user preference characteristics according to these 11 categories, calculate the average characteristics, and recommend them, respectively.

6. Experiment and Discussion

In this section, evaluation of songs is made by some experiments. Take some samples from two different websites of music and dataset information is given in Table 1. Then, evaluate the song recommendation. This recommendation result at the end shows that comprehensive evaluation is better than multicategory evaluation. The complete details of experiments and discussion are as follows.

6.1. Evaluation on Song Classification. The training samples come from two music websites and are divided into four standard classes: Blues, Classical, Jazz, and Pop. The two datasets are named MA and MB, respectively. The data distribution of each dataset is shown in Table 1. The frequency levels are compressed into 128 levels, corresponding to the ordinate of the spectral image. The frequency data per 1 second is mapped to 50 pixels, corresponding to the abscissa of the spectral image. The pixels of the segmented spectral segment are 128×128 , representing the audio signal of 2.56 s. 60% of the image samples are used as training samples and 40% are used as test samples.

First, the training process of the network is evaluated to determine whether the network can converge. The training loss variation of the network is shown in Figure 3.

TABLE 1: The dataset information.

Dataset	Blues	Classical	Jazz	Pop
MA	105	108	113	112
MB	109	106	116	114

Obviously, at the beginning, as the number of training increases, the network loss decreases significantly. But when the epoch is equal to 30, the loss of the network no longer decreases as the training iteration progresses. This shows that the network has reached the convergence state at this time, which verifies the feasibility and correctness of the network designed in this paper.

In this work, we use two different activation functions (ReLU and ELU) to process network song features. In order to verify which activation function is more suitable for the song classification task, this work conducts experiments for different activation functions. The results are shown in Figure 4; the evaluation metrics are accuracy and AUC.

It is obvious that the network is more capable of classifying songs when using the ELU activation function. On the MA dataset, 0.4% accuracy improvement and 0.6% AUC improvement can be obtained. On the MB dataset, 0.6% accuracy improvement and 0.7% AUC improvement can be obtained.

Finally, this work compares the song classification performance when using spectrum and note spectrum, and the activation function used is ELU. The experimental results are shown in Figure 5.

It is obvious that the network is more capable of classifying songs when using the note spectrum. On the MA dataset, 1.1% accuracy improvement and 0.9% AUC improvement can be obtained. On the MB dataset, 1.6% accuracy improvement and 1.6% AUC improvement can be obtained.

6.2. Evaluation on Song Recommendation. This article recommends a single user to recommend songs with high similarity to the user. The data sample of this experiment is 1000 songs, 200 songs per user. After the trained CNN predicts the classification, 1000 classification features are obtained, and the classification is optimized. The threshold model is 0.5, and songs whose maximum ratio is greater than the threshold are considered to have multiclassification features. In order to be closer to the user's behavior and habits, this paper uses the hidden Markov-based song classification model to calculate the user's music list, which is used as the reference standard for the recommendation results.

In the recommendation algorithm experiment combined with the comprehensive evaluation of user features, this paper first calculates the user's average preference feature $\overline{P_{u,c}}$ and calculates its similarity with the song classification feature, and the similarity threshold is set to 0.1. Randomly select 10 songs with $S_{A,B}$ less than 0.1 for recommendation. The preference list of each user is a reference sample generated by HMM, and $\overline{P_{u,c}}$ is the average feature of each user. The recommended results of the comprehensive evaluation

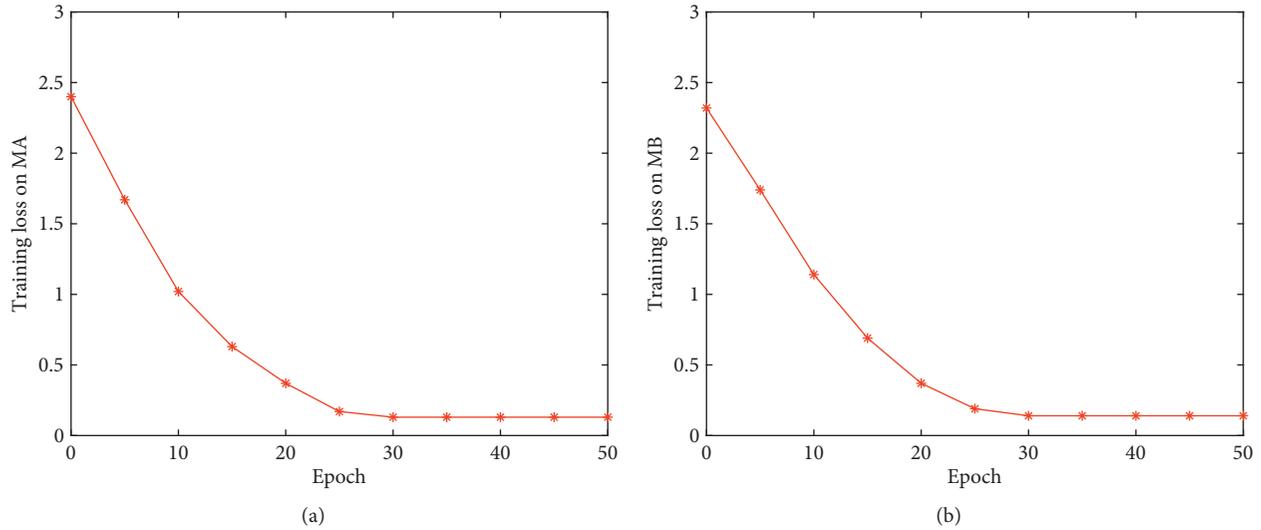


FIGURE 3: Evaluation on training loss.

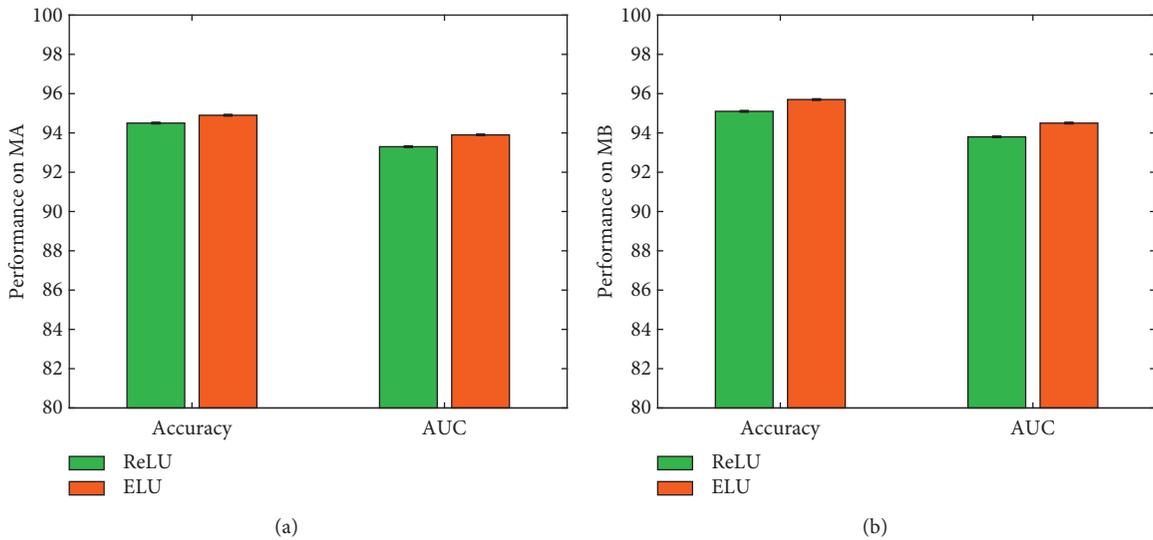


FIGURE 4: Comparison of ReLU and ELU.

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

Experiments show that the recommendation accuracy of users with a single category in the average preference feature set is higher than that of users belonging to multiple categories. This may be due to the fact that there are fewer multicategory songs, resulting in a small probability of being randomly obtained. It may also be that the user prefers a variety of single categories, and the calculated features are relatively discrete, resulting in the recommendation algorithm predicting that the user’s preference features are multicategory. It can be seen that the average preference feature is not very accurate to represent the user’s overall preference feature. The multicategory recommendation is overall worse than the average feature recommendation, on

average, but for multi-category users, the multicategory recommendation works better. This is because the multiclass classification process avoids a single representation of the average feature.

The comparison of the recommendation results of the two recommendation methods for different user types is shown in Table 3.

The results show that the recommendation method based on comprehensive evaluation of user characteristics is generally better than the recommendation method based on multicategory evaluation of user characteristics. For single category users, the recommendation method based on comprehensive evaluation of user characteristics is also better than the recommendation method based on multicategory evaluation of user characteristics. For multicategory users, the recommendation method based on

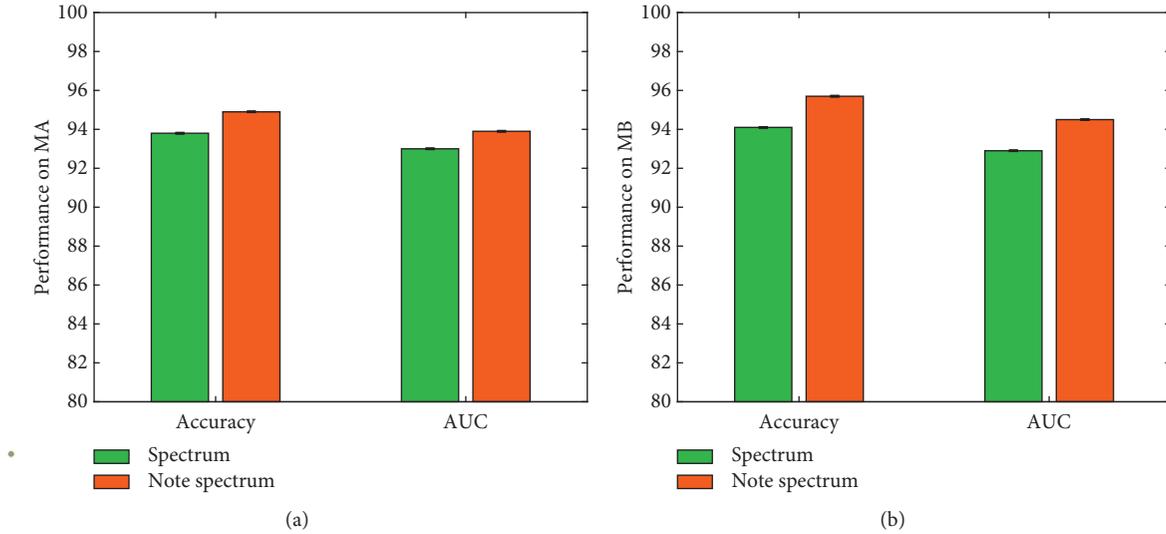


FIGURE 5: Comparison of spectrum and note spectrum.

TABLE 2: Comprehensive evaluation and recommendation of user characteristics.

User	Comprehensive evaluation Accuracy (%)	Multicategory evaluation Accuracy (%)
U1	64.1	61.2
U2	45.9	41.3
U3	54.3	53.3
U4	49.7	47.8
U5	60.3	58.8
Average	54.9	52.5

TABLE 3: Comparison of recommended methods.

Method	User type	Average accuracy (%)
Comprehensive evaluation	Single category	58.3
	Multicategory	25.7
	All	54.9
Multicategory evaluation	Single category	51.7
	Multicategory	41.3
	All	52.5

multicategory evaluation of user features is better than the recommendation method based on comprehensive user feature evaluation.

7. Conclusion

In this paper, the CNN is trained by combining the spectral features and note features of the song, and the CNN classification model is obtained. The final classification of songs is achieved through further analysis of the CNN classification results. Two different strategies for personalized recommendation based on user preference characteristics are explored. The main work is summarized as follows: (1) song spectrum generation and note recognition improvement. In this paper, the SoX tool is used to generate the spectrum of the song, which expresses the characteristics of

the song from three dimensions. At the same time, this paper also improves the traditional autocorrelation note recognition algorithm, improves the correct rate of note recognition, and combines the note with the spectrum to generate the note spectrum. (2) CNN-based classifier design and training: the neural network parameters adjustment and activation function selection are compared and analyzed. The experimental results show that, for the same algorithm model, this paper tests two activation functions, ELU and ReLU, and it is found that the activation function of CNN can often obtain better results by using ELU. (3) Comprehensive analysis of classification results and design of song recommendation algorithm: this paper conducts statistics and analysis on the classification results of CNN and finds that there are data errors and classification errors in the classification results, which are not suitable for direct recommendation. Then a threshold model is proposed for further classification, and the optimized classification result is used as the classification feature of songs. After CNN classification and optimization, the relationship between songs and classifications is obtained. Combined with the behavioral preference relationship between users and music, the prediction of user's preferred music classification is realized. Categorical features, as a representation of user preferences, can be used not only to measure the similarity between users, but also to calculate the similarity between user preferences and music. This paper proposes two recommendation methods based on user preference characteristics.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that he has no conflicts of interest.

References

- [1] S. Sharma, V. Rana, and V. Kumar, "Deep learning based semantic personalized recommendation system," *International Journal of Information Management Data Insights*, vol. 1, no. 2, Article ID 100028, 2021.
- [2] M.-C. Chiu, J.-H. Huang, S. Gupta, and G. Akman, "Developing a personalized recommendation system in a smart product service system based on unsupervised learning model," *Computers in Industry*, vol. 128, Article ID 103421, 2021.
- [3] W. Du, Y. Wang, W. Xu, and J. Ma, "A personalized recommendation system for high-quality patent trading by leveraging hybrid patent analysis," *Scientometrics*, vol. 126, no. 12, pp. 9369–9391, 2021.
- [4] Y. Liu, "Construction of personalized recommendation system of university library based on SOM neural network," in *Proceedings of the International Conference on Information Systems and Computer Aided Education*, pp. 2199–2203, NY, USA, September 2021.
- [5] P. Nitu, J. Coelho, and P. Madiraju, "Improvising personalized travel recommendation system with recency effects," *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 139–154, 2021.
- [6] X. Wen, "Using deep learning approach and IoT architecture to build the intelligent music recommendation system," *Soft Computing*, vol. 25, no. 4, pp. 3087–3096, 2021.
- [7] 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.
- [8] D. Paul and S. Kundu, "A survey of music recommendation systems with a proposed music recommendation system," in *Emerging Technology in Modelling and Graphics*, pp. 279–285, Springer, Singapore, 2020.
- [9] R. Liu and X. Hu, "A multimodal music recommendation system with listeners' personality and physiological signals," in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, pp. 357–360, Virtual Event, China, August 2020.
- [10] S. Visnu Dharsini and B. Balaji, "Music recommendation system based on facial emotion recognition," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 4, pp. 1662–1665, 2020.
- [11] M. Sheikh Fathollahi and F. Razzazi, "Music similarity measurement and recommendation system using convolutional neural networks," *International Journal of Multimedia Information Retrieval*, vol. 10, no. 1, pp. 43–53, 2021.
- [12] Y. Afoudi, M. Lazaar, and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," *Simulation Modelling Practice and Theory*, vol. 113, Article ID 102375, 2021.
- [13] M. V. Manoj Kumar, B. S. Prashanth, S. Sarah, M. Kashif, and H. R. Sneha, "Expression X: emotion based music recommendation system," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 9, pp. 3958–3963, 2020.
- [14] T. Oswari, T. Yusnitasari, and R. D. Kusumawati, "Design and test music recommendation system for online music websites using collaborative filtering approach," *International Journal of Digital Signals and Smart Systems*, vol. 4, no. 1-3, pp. 64–79, 2020.
- [15] S. Chen, S. Sou, and H. Hsieh, "HPCF: hybrid music group recommendation system based on item popularity and collaborative filtering," in *Proceedings of the International Computer Symposium (ICS)*, pp. 43–49, Tainan, Taiwan, December 2020.
- [16] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [17] A. Nabizadeh and Y. Yu, "Predicting user preference based on matrix factorization by exploiting music attributes," in *Proceedings of the International Conference on Computer Science & Software Engineering*, pp. 61–66, Porto Portugal, July 2016.
- [18] J.-H. Su, W.-Y. Chang, and V. S. Tseng, "Personalized music recommendation by mining social media tags," *Procedia Computer Science*, vol. 22, pp. 303–312, 2013.
- [19] S. Lee, M. Masoud, and J. Balaji, "A survey of tag-based information retrieval," *International Journal of Multimedia Information Retrieval*, vol. 1, pp. 1–15, 2016.
- [20] S. Li, J. Li, and R. Pan, "Tag-weighted topic model for mining semi-structured documents," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 2855–2861, Beijing, China, August 2013.
- [21] Y. Sheng Wang, "Image tag recommendation algorithm using tensor factorization," *Journal of Multimedia*, vol. 9, no. 3, 2014.
- [22] A. Nanopoulos and I. Karydis, "Know Thy Neighbor: combining audio features and social tags for effective music similarity," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 165–168, Prague, Czech Republic, May 2011.
- [23] N. Hariri, B. Mobasher, and R. Burke, "using social tags to infer context in hybrid music recommendation," in *Proceedings of the Twelfth International Workshop on Web Information and Data Management*, pp. 41–48, Maui, HI, USA, November 2012.
- [24] J. Y. Liu and Y. H. Yang, "Inferring personal traits from music listening history," in *Proceedings of the International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies*, pp. 31–36, Nara, Japan, November 2012.
- [25] D. Dolgikh, "Graph-based music recommendation approach using social network analysis and community detection method," in *Proceedings of the International Conference on Computer Systems and Technologies*, pp. 221–227, Dublin, Ireland, June 2015.
- [26] J. Shen, H. H. Pang, and M. Wang, "Modeling concept dynamics for large scale music search," *Research Collection School of Information Systems*, vol. 29, no. 4, pp. 455–464, 2012.
- [27] N. Hariri, B. Mobasher, and R. Burke, "Context-aware music recommendation based on latent topic sequential patterns," in *Proceedings of the ACM conference on Recommender systems*, pp. 131–138, Dublin, Ireland, September 2012.
- [28] X. Wang, D. Rosenblum, and Y. Wang, "Context-aware mobile music recommendation for daily activities," in *Proceedings of the ACM international conference on Multimedia*, pp. 99–108, Nara, Japan, October 2012.
- [29] K. Kapoor and V. Komal, "I like to explore sometimes," in *Proceedings of the ACM Conference on Recommender Systems*, pp. 19–26, Vienna, Austria, September 2015.
- [30] V. Saúl and P. Castells, "Rank and relevance in novelty and diversity metrics for recommender systems," in *Proceedings of the ACM Conference On Recommender Systems*, pp. 109–116, Chicago IL, USA, October 2011.
- [31] M. Zhang and N. Hurley, "Avoiding monotony: improving the diversity of recommendation lists," in *Proceedings of the*

ACM Conference On Recommender Systems, pp. 123–130, Lausanne Switzerland, October 2008.

- [32] Y. Zhang, D. Quercia, and T. Jambor, “Auralist: introducing serendipity into music recommendation,” in *Proceedings of the ACM international conference on Web search and data mining*, pp. 13–22, Seattle, WA, USA, February 2012.
- [33] M. Schedl and D. Hauger, “Tailoring music recommendations to users by considering diversity, mainstreaminess, and novelty,” in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 947–950, Santiago, Chile, August 2015.