Chapter 16
Music Personalized Recommendation System Based on Deep Learning

Chaozhi Cheng

Abstract In the Internet age, these have brought great convenience for users to obtain personal preference information. Among them, personalized recommendation algorithm is a very key problem. The music personalized recommendation system studied in this paper combines DL (Deep learning) technology, uses CNN (Convolutional Neural Network) to predict the hidden features of music, and obtains the low-dimensional vector representation of music features. Combined with the hidden representation of user preferences, it learns the hidden layer output of auxiliary information through AE (Autoencoder) and then integrates it into the traditional personalized recommendation algorithm. Finally, a reasonable personalized recommendation is generated for relevant users. According to the experimental results, with the increase in the number of prediction scores, the value of MAE (Mean Absolute Error) decreases continuously. The MAE of this model is lower than other models, and it has a better recommendation effect. The reliability of the personalized music recommendation system established in this paper is verified.

16.1 Introduction

With the rapid rise of network technology and electronic information technology, technologies such as big data, cloud computing, robotics, artificial intelligence and DL (Deep learning) have also developed rapidly, which provide huge computing resources for the progress and development of the whole information age [1]. In the Internet age, these have brought great convenience for users to obtain personal preference information. In the recommendation system, the design of personalized recommendation algorithm is particularly important.

The music industry has gradually turned to online music. Faced with such a huge network user group, intelligent music recommendation has become an inevitable trend in the development of online music platforms. With the popularity of DL

C. Cheng (✉)
College of Music and Dance, Huaihua University, Huaihua, China
e-mail: ccz175@163.com

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*, Smart Innovation, Systems and Technologies 385,
https://doi.org/10.1007/978-981-97-0109-4_16

method in recent years, more and more personalized recommendation systems have begun to use DL method [2, 3]. Some studies have adopted LSTM (Long Short-Term Memory) network to recommend HashTag in Weibo [4]; Researchers put forward a way to extract features from audio signals by CNN (Convolutional Neural Network). Their main method is to get the feature vectors of music by matrix decomposition [5]. Similar music has similar feature vectors, so they can recommend similar music to users. Literature [6] proposes a hybrid recommendation system based on DL, which integrates external information about users and items into deep neural network by using embedding technology, thus alleviating the cold start problem.

Music is an ancient art that can bring happiness to human beings. However, in order to accurately find songs that meet users' requirements in a large number of music works, we must choose songs that meet users' preferences according to their preferences. On this basis, a new method of music personalized recommendation based on DL is proposed. The recommendation system can predict users' usage habits based on their usage habits and the characteristics of music data, and actively recommend music suitable for users [7, 8]. On this basis, this project plans to adopt the classic personalized recommendation method based on DL, and realize the effective mapping of users and songs through AE (Autoencoder) and CNN (Convolutional Neural Network). On this basis, according to different user preferences and different song content, the similarity analysis is carried out to get songs suitable for different songs.

16.2 Research Method

16.2.1 Recommend Algorithm Classification

Content-based Recommendation Algorithm

Based on the content-based recommendation method, the characteristics of users and objects are extracted, the user's preference model is constructed, and the matching degree between the characteristics of information to be selected and the user's preference model is calculated to produce recommendation results. The traditional content-based recommendation algorithm first obtains the items that have interacted with users, and then obtains the user's preference model through similar user active behaviors such as user's praise or rating of the items. Content-based recommendation mechanism uses items that have interacted with users, and obtain users' interest in different items through similar user behaviors such as praise, click, or rating. And collect similar items with high user interest, and then rearrange the recommendation degrees of these collected items and put them in the list to be recommended.

Collaborative Filtering Recommendation Method

Collaborative filtering recommendation is one of the most influential and widely used algorithms in industry. It is mainly divided into two categories: one is collaborative filtering based on users, and the other is collaborative filtering based on items. There are similarities and great differences between them. The main idea of collaborative filtering based on items is to find items with high similarity and positive feedback in the user's historical behavior records by analyzing the user's historical records, and recommending them to the target users after sorting. Its underlying assumption is that if almost all users who like one item like another item, the two items have the greatest similarity; If a user's interest list includes two specific items, it has a high probability of belonging to a specific field, but if two items are included in most users' interest lists, it has a high probability of belonging to the same field.

Matrix Decomposition Recommendation Algorithm

In the scoring prediction scenario of recommendation system, blank records are filled by matrix decomposition, which alleviates the sparsity of data to some extent. In the matrix decomposition algorithm, the scoring matrix is decomposed into the form of multiplication of user feature matrix and item feature matrix, and the potential features of users and items are mined to predict the scoring of unrated items and reduce missing values. In many cases, the form of simple matrix decomposition and multiplication can not restore the scoring matrix well, and there are certain errors, which can be continuously reduced by combining gradient descent. When the user-item scoring matrix is very sparse, the gradient descent method is used to continuously reduce the error loss value, and at the same time, it is easy to produce the phenomenon of over-fitting, that is, the existing data is well fitted but the prediction effect of the missing value is not good.

16.2.2 Overall Design of Recommendation System

In the era of network information overload, how to maximize the information in the network has become the focus of attention. According to the keywords input by users, on this basis, a new recommendation method with wide application prospect is proposed. Although the information transmission mode between search engine and recommendation system is different, there are some differences between them. In the music industry, the earliest cooperative screening algorithms are based on clear feedback, that is, according to users' comments on a song or a singer. Aiming at the current recommendation system and different application scenarios, a common classification algorithm is given. When the information of a person or an object is used, it is considered as a content-based information recommendation. If the recommendation system uses interactive data generated by users and items, it is considered as a recommendation system based on collaborative filtering.

For music recommendation system, collaborative filtering algorithm has more practical experience. Because most users don't have the habit of evaluating music, users' rating data of music is sparse. The application of DL method in recommendation system is a new trend, which provides a new idea under the complicated situation of music data processing [9]. Although most music systems or data sets have corresponding music tags, the problem of tag missing is still serious. The problem of missing music labels is quite common. When making music recommendation, it will face the challenge of completing or correcting labels. It is necessary to extract music features by using neural network model and label and classify music. For example, the "list" label under the "theme" standard has nothing to do with the characteristics of music itself, and this part of the label needs to be removed when the classification model is actually operated.

At present, CNN has been widely used in the fields of image and text recommendation. However, the key point of this topic is that, under the framework of CNN, it is fundamentally a composite recommendation model that integrates music content and users' historical behavior. The core idea is as follows: firstly, based on CNN, the hidden characteristics of songs are predicted and the low-dimensional vector expression of songs is obtained; Secondly, on this basis, based on AE learning, the implicit characteristics of songs are combined with the implicit characteristics of songs, and they are combined with classical song recommendation algorithms. In the training process, the tightly coupled model learns the parameters of AE and traditional personalized recommendation algorithm at the same time, and the two methods influence each other, so that the traditional personalized recommendation algorithm can provide prior knowledge when the hidden layer of AE learning features is output, and finally generate reasonable personalized recommendation for relevant users.

On the Internet, faced with a huge group of users, intelligent recommendation of music becomes more important, so that users can better find music that suits their preferences, and music services on the Internet become more attractive, thus increasing the profits of enterprises [10]. Through the research of this project, we can enrich the information of music recommendation, solve the problems of "cold start", "sparseness" and "scalability" in traditional music recommendation to some extent, and better meet the needs of users for personalized music. On this basis, this paper proposes a personalized music recommendation system based on DL (Fig. 16.1).

The system mainly includes user modeling module, music feature extraction module, and personalized recommendation algorithm module. The music feature extraction module is mainly used to preprocess audio content and extract spectral features, so as to prepare for training CNN to obtain the regression model of music potential feature prediction.

The realization of the whole system function mainly includes the following steps:

(1) Collect the historical behavior data of users in the system, and construct a hidden semantic model that can reflect the relationship between users and music according to a unified quantitative standard;
(2) Pre-processing the original music resource file, and extracting the spectrogram which can represent the music audio characteristics;
(3) Construct a convolutional neural network model, take the extracted spectrum as the input of the network model, and finally get the CNN regression model;

(4) When there are new music resources in the system, the audio feature map is also obtained first, and then the potential features of music are predicted by using the obtained CNN regression model. Combined with the user's preference model, the user's interest in music is calculated and sorted, and finally, the first new music resources of TopN are recommended to relevant users.

All data services of the system are exposed through the interface, and authorized calls can use the data services of the system service layer at will. The data service of the service layer interacts with the data layer through the persistence layer. The offline computing engine interacts with the data layer through the persistence layer and stores the offline computing results in the database. Recommend other music lovers who may be interested to users. In QQ music, song lists are created by users, and a user can create multiple song lists. Click on the user's details and the song list entries created by the user, as well as similar user recommendations based on the current viewing user.

16.2.3 Design of Personalized Music Recommendation Algorithm

Traditional recommendation algorithms have great limitations in the process of music recommendation. With the continuous development of DL, many music recommendation methods combining traditional recommendation algorithms with DL have emerged in the field of music recommendation. Most of them pay attention to the historical behavior data of users, but ignore the potential information. Because of the repetitive characteristics of historical behavior data, the recommended music is similar, and there are problems such as sparse data and cold start.

At present, mature commercial solutions are divided into two categories: search engine and recommendation system. Search engine is designed for users who have certain retrieval goals and needs. It can retrieve the information they need through keywords typed by users. On this basis, a new recommendation method with wide application prospect is proposed. There are two common characteristics in the applicable fields of recommendation system: first, there is a large amount of data in this field, and users do not have enough energy and time to contact and understand all items; Second, users have no clear demand for items in this field. Most of the fields that meet this feature are concentrated in the field of pan-entertainment, such as music, books, news, e-commerce, and so on. In the development of artificial intelligence, personalized recommendation algorithm combining DL and reinforcement learning emerged [11]. In the long river of recommendation system development, the traditional personalized recommendation algorithm played a key role in connecting the past with the future, and then the emerging personalized recommendation algorithms all improved and innovated on its basis.

At present, personalized recommendation technology has penetrated into every corner of the network, such as QQ music, which is closely related to users' browsing records and personal information. Personalized recommendation system is based on a variety of information, analyzing users' preferences, thus judging the products that users may be interested in, and then displaying these products within the acceptable range of users, waiting for users to choose, which can be considered as a subjective behavior of personalized recommendation system. The conversion rate of push is also lower than that of personalized recommendation, and it will also cause users' disgust because of the deviation between the pushed content and users' preferences or the frequent push.

In this paper, we will introduce a personalized music recommendation algorithm based on CNN. CNN belongs to a feedforward neural network, which is different from multi-layer receptors in its connection mode. Among them, the connection mode of multi-layer perceptual network is a complete connection, and it is composed of convolution layer, pooling layer, and complete connection layer. CNN convolves multi-dimensional features, which can effectively extract more features, especially for images, voices, and other multi-dimensional features.

In the framework of CNN, there is a pool behind the convolution layer. The role of the pool is to compress the input data to a smaller scale, and then repeat the process, and then extract more data, and finally obtain a complete data. Based on the network structure of Le Net5, this paper studies the basic principles of CNN architecture as summarized above by drawing lessons from several other good convolutional neural networks. On this basis, a personalized music recommendation model based on 7-level network is proposed (Fig. 16.2).

The neural network consists of four convolution-pooling structures, in which two convolution-pooling structures cross each other to form two perfectly connected structures and a prediction output layer. In the input layer, each pixel size is $256 \times 256 \times 1$, which is the Mel spectrum extracted from the sound. This method uses the method of maximum pool and sets the size of the pool window as 2×2.

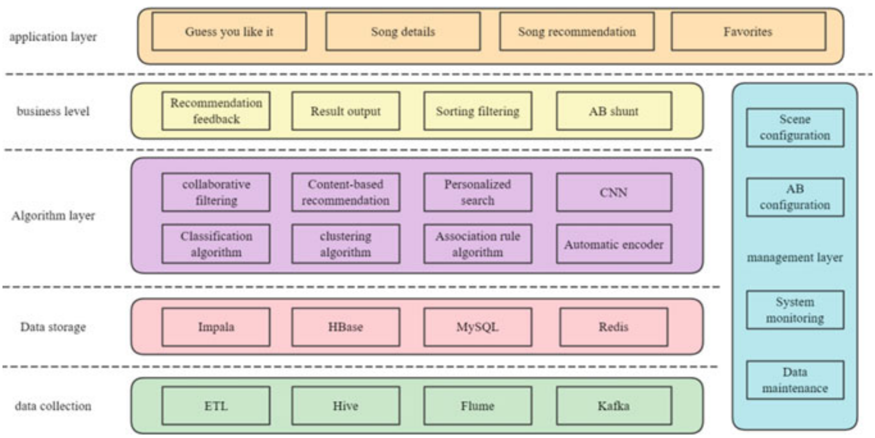

Fig. 16.1 Overall design block diagram of music personalized recommendation system

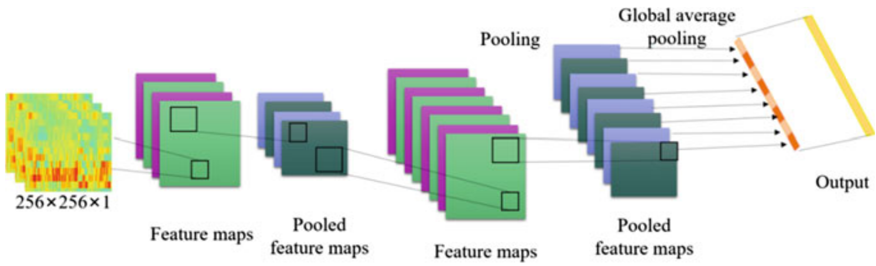

Fig. 16.2 CNN network model structure for personalized music recommendation

For a piece of music, descriptive information (music name, introduction, lyrics) and audio itself can represent its unique attributes. Among them, the feature of audio is the most unique and effective information representation of a piece of music. Through this feature, this paper can distinguish different music to the greatest extent. Converting sound signal into image representation can better use CNN to extract features [12].

Map Hertz frequency f to Mel frequency $mel(f)$, as shown in Formula (16.1):

$$mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \tag{16.1}$$

According to the Mel frequency obtained by Mel mapping formula, the perception of Mel frequency by human hearing is linear. When the Mel frequency of audio is doubled, the tone that human ears can perceive is also doubled. Mel spectrum uses human auditory perception characteristics to generate spectrum, and combining this characteristic with neural network can greatly improve the effect of automatic music classification.

AE is an unsupervised machine learning method based on coding and decoding. This method uses coding and decoding to reconstruct the input information and learn an implicit hierarchical expression from it. Acoustic emission consists of the following two steps:

Encoding process from input layer to hidden layer:

$$h = f(x) = \sigma\left(W'x + b\right) \tag{16.2}$$

Decoding process from hidden layer to output layer;

$$\tilde{x} = g(h) = \sigma\left(W^T h + b\right) \tag{16.3}$$

where W, b is the weight and bias term of the feature, $\sigma(\cdot)$ is the activation function, and AE aims to reconstruct the data in the input layer at the output layer, so its loss function combines different data forms.

Generally, the activation functions of artificial neural networks mainly include Sigmoid, Tanh, ReLU ELU, etc. The CNN model proposed in this paper chooses ReLU as its training activation function and Softmax as its output layer activation function.

ReLU, which is called rectifier linear element, is the mainstream activation function in the current network model. As shown in formula (16.4).

$$f(x) = relu(x) = \begin{cases} x, & x > 0 \\ 0, & other \end{cases} \tag{16.4}$$

ReLU has no negative value, it is hard saturated at $x < 0$, and its derivative is 1 at $x > 0$.

Softmax is often used for multi-classification and is used as the activation function of the prediction output layer in this model, as shown in Formula (16.5).

$$f(x)_j = \mathrm{Softmax}(x)_j = \frac{e^{z_j}}{\sum_{i=1}^{k} e^{z_j}} \tag{16.5}$$

In order to ensure that the CNN network model will neither fail to fit nor overfit, taking the gradient descent algorithm as an example, the parameter momentum update in the network can be expressed as follows:

$$\theta \leftarrow mu * \theta - \eta \frac{\partial L(\theta)}{\partial \theta} \tag{16.6}$$

In the formula, mu represents momentum, and the value of momentum coefficient is 0 ~ 1, and the commonly used values are 0.5, 0.9, 0.95, and 0.99. By adopting this parameter updating method, we can not only speed up the learning rate and increase the stability, but also have some ability to get rid of local optimization.

16.3 Experimental Analysis

The experimental data is a self-built QQ music data set, which contains user data, music data, singer data, song list data, and user music playing history data from QQ music. The physical environment of the experiment consists of Ubuntu as the operating system, Intel(R) Xeon(R) E5-2678 v3 as the processor, with a main frequency of 2.5GHZ and a turbo frequency of 3.1 GHz, and NVIDIA GeForce GTX 1080Ti as the graphics card. Adam is used to optimize the embedding dimension adjustment layer and recommendation model, and the learning rate is set to 0.0001 and 0.003 respectively. The dimension size of the fixed embedding vector is 8. The batch data size of the whole model during training is 500.

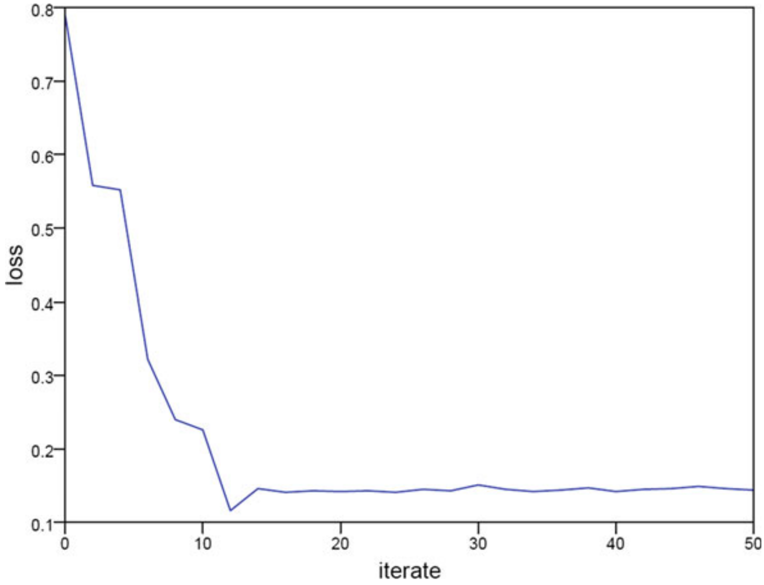

Fig. 16.3 Loss curve of training model

In this experiment, MSE (mean-squared error) is used as the loss function, and the CNN network model is trained by using the training set data. The training result is shown in Fig. 16.3.

We can see that with the increase of iterations, the loss error of the network model will decrease rapidly at first, and then it will decrease slowly. When the age reaches 12, the error will decrease to 0.116, the function will tend to converge, and the training process of the model will basically meet the expected requirements.

For the music recommendation model proposed in this paper, the scores of users' songs are predicted respectively, and the predicted scores of 10, 20, 30, 40, and 50 pieces of music are randomly selected to calculate MAE (Mean Absolute Error). The experimental results are shown in Fig. 16.4.

According to the experimental results, the MAE of this model is lower than the other two models, and it has a better recommendation effect. With the increase in the number of forecast scores, the value of MAE decreases continuously. The results show that the CNN combined with AE model proposed in this paper has the best model efficiency under MAE index, and also has the best anti-sparsity ability. The cold start problem of new items is solved, and the risk of over-fitting and falling into local minimum in the process of scoring matrix is reduced.

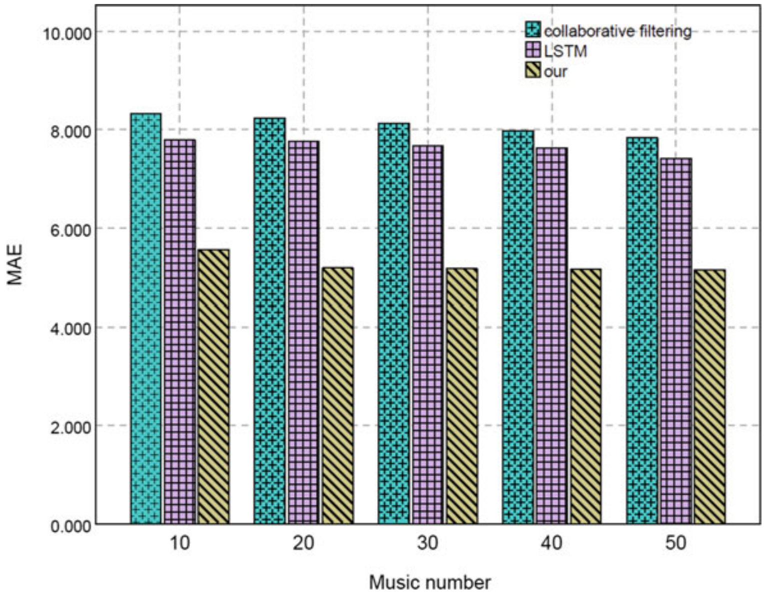

Fig. 16.4 The recommended model compares the experimental results

16.4 Conclusion

Personalized recommendation system plays a very important role in e-commerce system, audio and video playing platform, and news information software. How to predict users' interests based on their historical behavior data and other available information, so as to provide them with more accurate recommendations of products they are interested in, and thus improve their user experience, is the goal pursued by many scholars and application developers. The music industry has gradually turned to online music. Faced with this huge network user group, music intelligent recommendation has become a hot spot of online music service. On this basis, using AE, CNN, and other machine learning methods, a personalized music recommendation system based on DL is constructed to effectively mine information such as users and music. According to the experimental results, with the increase in the number of prediction scores, the value of MAE decreases continuously. The MAE of this model is lower than other models, and it has a better recommendation effect.

References

1. Zhang, J., Yang, Y., Tian. Q., Zhuo, L., Liu. X.: Personalized social image recommendation method based on user-image-tag model. IEEE Trans. Multimed. **19**(11) (2017)

2. Zhou, L., Wang, C.. Research on Recommendation of Personalized Exercises in English Learning Based on Data Mining. Scientific programming, Vol. 2021(14). pp.5042286.1–5042286.9(2021).
3. Wang, C.Y., Wang, Y.C., Chou, S.: A context and emotion aware system for personalized music recommendation. J. Internet Technol. **19**(3), 765–779 (2018)
4. Zhang, L., Tian, Z.: Research on the recommendation of aerobics music adaptation based on computer aided design software. J. Intell. Fuzzy Syst. **2**, 1–12 (2021)
5. Shi, J.: Music recommendation algorithm based on multidimensional time-series model analysis. Complexity **2021**(1), 1–11 (2021)
6. Li, M., Bao, X., Chang, L., et al.: Modeling personalized representation for within-basket recommendation based on deep learning. Expert. Syst. Appl. **192**(8), 116383 (2021)
7. Quintanilla, E., Rawat, Y. S., Sakryukin, A., et al.: Adversarial learning for personalized tag recommendation. IEEE Trans. Multimed. **23**, 1083–1094 (2020)
8. Yang, X., Zhou, Z., Xiao, Y.: Research on students' adaptive learning system based on deep learning model. Sci. Program. **2021**(Pt.13), 6593438.1–6593438.13 (2021)
9. Wang, M.: Applying Internet information technology combined with deep learning to tourism collaborative recommendation system. PLoS ONE **15**(12), 6 (2020)
10. Wang, X. Y., Chen, Q. Y.: Research on personalized recommendation system of libraries based on deep learning. Mod. Sci. Instrum. (6), 4.4–47 (2018)
11. Li, Q. N., Li, Ti. H.: Research progress of context-aware recommendation system based on deep learning. Comput. Syst. Appl. **29**(2), 8 (2020)
12. Shi, Y.Q., Liu, S.Y., Mao, C.: Research on object detection and content recommendation system in short video based on deep learning. Comput. Mod. **11**, 1 (2018)