Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5936156, 8 pages
https://doi.org/10.1155/2022/5936156

Research Article

Automatic Recommendation of Online Music Tracks Based on Deep Learning

Hong Gao

School of Teachers' Training, Eastern Liaoning University, Dandong 118003, China

Correspondence should be addressed to Hong Gao; gaohong@elnu.edu.cn

Received 11 April 2022; Revised 28 April 2022; Accepted 30 April 2022; Published 7 June 2022

Academic Editor: Zaoli Yang

It is one of the main goals of personalized music recommendation system that how to accurately recommend the songs in line with users' interests in the huge music library. In view of the above problems, this study proposes a personalized music recommendation method based on convolutional neural network. First, this study defines a training set containing potential musical characteristics and, combined with the depth of the belief network, design a music information prediction model and the research in the music-type classification method with different dimensions. Based on selecting four different kinds of music information better describing the underlying characteristics of 40D feature vector to every song music composition, the music feature set is constructed. Then, the CNN (convolutional neural network), which is widely used in audio field, is used as the music information prediction model, and its structural parameters are redesigned to complete the multidimensional music information prediction, which solves the cold start problem to some extent.

1. Introduction

By the end of 2020, the number of Internet surfers in China was about 1 billion, and the Internet penetration rate was close to 71%, among the huge digital music market [1]. Despite the economic and trade frictions between China and the US, the online music industry is still growing at a high rate of over 8%. This shows that the development potential of digital music is huge [2].

After many years of development and improvement, recommendation technology has been widely used in many fields, such as short video field, news field, stock field, and e-commerce field. With the rapid development of the Internet, the problem of data overload is becoming increasingly prominent. In the music field, as the song library becomes larger and richer, it is difficult for users to quickly find the songs they are interested in [3]. Most people will use the search function of software for some singers they know in the past or categories of songs they like. However, the search results do not consider that users are different individuals and have different preferences for songs, leading to low user satisfaction.

People generally judge the merits of different feature extraction methods by the experimental results of music classification task [4]. In music classification task, labels are generally the genres to which music belongs, such as pop, rock, electronic, rap, and light music. Therefore, it is not reasonable to classify music only according to the type of music. A good music recommendation requires comprehensive recommendation based on the type of music and users' emotions and hobbies. People in different scenarios and different moods will choose different music in terms of features, the simplest way of feature extractio\n from music metadata access to get music genre, year album information, etc. In the music information retrieval field, it is more about extracting audio features. In the past, people often used related techniques in the field of music recommendation to extract acoustic features, as a feature representation of music and also tried to use machine learning methods to help extract audio features. In general, the effectiveness of the feature extraction method can be evaluated by comparing the experimental results in these application fields [5]. Feature extraction can be done automatically by a computer or manually by

an expert in the music field. However, the ability to pick up good music and audio features, whether automatic or manual, is crucial for content-based recommendation systems. The quality of feature extraction can not only better predict users' music taste but also provide a better training dataset in the follow-up recommendation model [6, 7].

Faced with huge amounts of digital music database overload phenomenon, the traditional music retrieval mode gradually exposed its existing problems. On the other hand, there is no retrieval method to meet the needs of users [8]. Of course, the increasing amount of music data and user demand put forward higher requirements on the research of music recommendation algorithm. The problems existing in traditional recommendation methods, such as collaborative filtering and cold start, need to be solved urgently, and the original recommendation algorithm needs to be upgraded [9]. This study, based on CNN(convolutional neural network), starts from analyzing the characteristic data of music and designs a personalized music recommendation system, which can bring not only great convenience to music users but also what the music platforms want to achieve [10, 11].

To sum up, it is very important to have a personalized song recommendation system for users. Based on users' historical listening behavior and characteristics of songs, the recommendation system can predict users' preferences for various songs and personalized push songs to users. At the same time, this system can also manage the personal listening page of songs, such as collection of favorite songs, view lyrics, song performers, and other functions [12].

2. Related Work

2.1. Research Status of Recommendation System. With the progress of the Internet, people now mainly obtain information through online surfing, such as listening to books, movies, and music, as one of the main hobbies of People's daily life [13]. With the online music platforms developing rapidly, the competition of online music platforms has shifted from the depth and scale of music library to the recommendation and discovery of music. Users' demand for music is no longer simple search or download. In China, the development of music recommendation system is relatively late, and the ability of personalized music recommendation is weak. However, there are some excellent Internet music products [14]. For example, on rainy days, sunny days, and dark days, the user's preference for music will change periodically. Therefore, the music recommendation system needs to integrate these factors and carry out iterative adjustment to realize personalized recommendation for users' pain points. Of course, in recent years, with the rise of e-commerce giants such as Alibaba and Jingdong, the relevant theories and technologies of the recommendation system have been attached importance to by Chinese people, which has set off a boom of research and application in the field of recommendation system and achieve achievement [15]. In this study, the scalability and future development trend of sparsity algorithm for evaluation data of recommendation system are discussed. In order to overcome the disadvantages of single recommendation technique, many researchers put forward a combination of multiple recommendation techniques. At present and in the future, the research content of recommendation system mainly includes user information acquisition and modeling, recommendation algorithm research, evaluation of recommendation system, and application and social influence research of the recommendation system. Nowadays, "Internet +" and the era of big data bring new opportunities and challenges to the research of recommendation system algorithm. On the one hand, the development of data mining, machine learning, artificial intelligence, and other technologies provide new feasible methods for the research of recommendation system. Besides, the massive information and the huge demand of Internet users put forward higher requirements for real-time performance, algorithm scalability, and human-computer interaction of recommendation system [16, 17]. In short, the recommendation system has made great progress in theoretical and applied research, but it has not yet reached a very mature stage. As a frontier exploration field, there are still many problems worthy of in-depth analysis and further discussion.

Many large companies began to use machine learning or deep learning-related technologies to build recommendation systems. Traditional collaborative filtering and all kinds of rule-based recommendation system, which is the current one of the world's largest video sharing sites, to build a massive recommendation system and put a lot of research and development personnel research related technologies, continuously put forward new recommender system technologies, from the decomposition of matrix to the depth of the application of neural network. The deep network model enables the recommendation structure to not only meet users' preferences but also generate some different and diversified contents [18, 19]. The online test results of this model in Google AppsStore show that it has effectively increased the downloads and purchases of applications Google and have also opened the relevant codes for scholars to refer to and learn from. Many well-known enterprises in China have invested in the research of the recommendation system and carried out various competitions to encourage people to participate in and put forward better recommendation strategies [20].

2.2. Research Status of Music Recommendation System. The methods of the recommendation system mainly include collaborative filtering method, content-based recommendation method, knowledge-based recommendation method, and mixed recommendation method. Music is an important application field of the recommendation system [21]. Compared with watching movies, reading books, watching TV, and other related entertainment activities, listening to music is the entertainment activity that people take more frequently. In

general, recommendation methods in various fields can be used for reference. However, although collaborative filtering has been effective in other fields, it is rarely been used seen in music recommendation alone. Most of the research experiments using collaborative filtering for music recommendation are used as a proof of the universality of the recommendation method proposed. The music industry often does not have enough rating data, and collaborative filtering can cause cold starts [22, 23]. In recent years, some scholars use collaborative filtering combined with other information or mixed with other methods to build a music recommendation system, which has attracted more attention. The basic idea of the content-based recommendation method is to collect information describing the content of the item and recommend the item to users who like similar items. This method is very dependent on the characteristics of the extracted item. Based on knowledge, the recommended method is suitable for application in items often used by a one-time purchase or field, so it is not suitable for applications in the field of music [24]. The method is based on the weighted preferences of the users and items, and through a series of technical regulations to recommend knowledge acquisition, we obtained the user interaction information. The mixed recommendation method refers to the recommendation method that combines the above methods. It often makes up for the shortcomings of another method through one method, such as combining the content-based recommendation method with the collaborative filtering method. The two methods are mixed to compensate for the cold start problem in collaborative filtering. Although the mixed recommendation method may be better than the traditional model at present, the research of the hybrid recommendation method is still in its infancy [25, 26].

Deep learning in recent years has developed rapidly in terms of image recognition and speech processing; a neural network model can be based on the input automatic learning characteristics, manual processing, and hierarchical network and can show the characteristics and found the distributed characteristics of the presentation of data [27]. At the same time, current studies have proved that it can be used in retrieval and recommendation tasks. The dynamic attention mechanism depth model is used to deal with the nonexplicit selection criteria of articles in the process of news release, learn the dynamic criteria for editors to choose article styles, and find out whether editors like articles in the article pool for binary recommendation. Recently, many companies have applied the method of deep learning to further improve the recommendation quality of their corresponding services [28]. All of these models have been tested online and show significant improvements over traditional models [29]. Based on the above research development and status quo at home and abroad, this study proposes a music recommendation method based on convolutional neural network to automatically mark user preferences.

Based on the above discussions, the main contributions of this study are shown as below:

(1) Firstly, the feature extraction method is used to extract deep features of music for subsequent model training and testing

(2) Use the convolutional neural network model to recommend and predict music tracks

(3) Determine the hyperparameters of the model by cross validation

3. Automatic Recommendation of Music by CNN

3.1. Principles of Deep Learning. Based on the study of human brain cortex, the brain as the external information processing using the hierarchical mechanism, when the brain receives information from sensory organs, the information will be passed on to the layers of progressive neurons, and each layer represents the feature extraction of things, through the layers of transfer form to the cognition of things. To a large extent, the construction of deep neural network is based on this cognitive process, which abstracts information several times. In deep learning, neurons in the human brain are replaced by processors. Each processor receives features extracted from the upper layer and carries out further feature extraction to the next layer so as to establish a connection between features at the bottom and things at the top [30].

Specific parameters and neural network training process [31] are also an important part of deep learning, and the output can meet people's needs through parameter adjustment [32, 33] (1) Supervised learning: supervised learning is usually applied in the field of classification, general process can be simply summarized as training samples of the labels and then achieving the purpose of classification after judging the result. Commonly used supervised learning methods are KNN, SVM, and so on. (2) Unsupervised learning: unsupervised learning is also used more in the study of learning style, different from supervised learning and unsupervised learning for training [34]. Training sample is not marked, through the learning of the training sample, training focused on potential structural knowledge, so the unsupervised learning samples of ambiguity is higher; researchers have designed many deep learning models, such as convolutional neural network, short and short memory network, and self-encoder [35, 36].

3.2. The Proposed Recommendation Method. In recent years, the convolutional neural network model is often used to solve complex image recognition problems. On the basis of traditional full-connection layer neural network, convolutional neural network adds convolution layer and pooling layer, which is shown in Figure 1.

The function of the convolution layer lies in the extraction of image features. The essence of the convolution kernel is a filter matrix, which can produce many different effects on the original image. The calculation process of convolution is shown below:

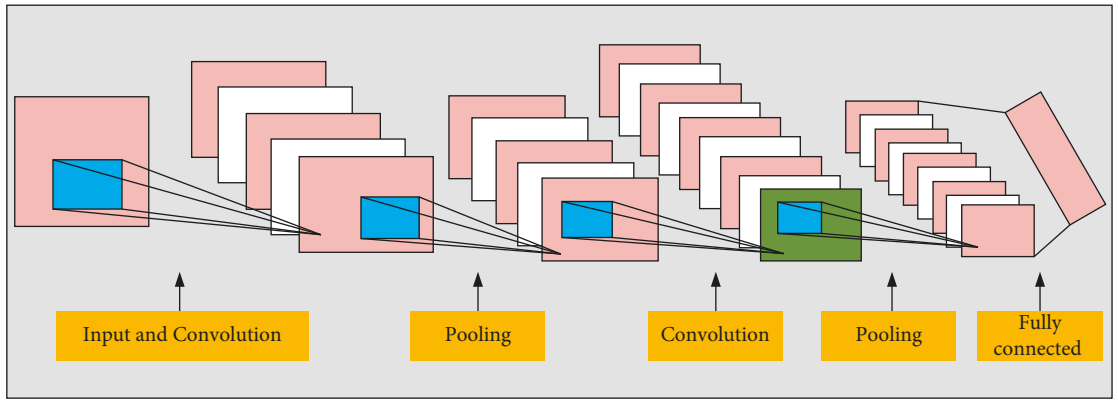

Figure 1: Schematic diagram of CNN.

$$x_i = \mathrm{act}\left(x_{i-1} \otimes k_i + b_i\right). \tag{1}$$

Then, the mathematical expression of sigmoid function is

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{2}$$

The mathematical expression of tanh function is

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{3}$$

The mathematical expression of ReLu function is

$$f(x) = \max(0, x). \tag{4}$$

The mathematical expression of LeakyReLu function is

$$f(x) = \begin{cases} x, & x \geq 0, \\ \alpha x, & x < 0. \end{cases} \tag{5}$$

Therefore, the efficiency of the entire network operation can be improved to a certain extent.

The output layer adopts softmax function to normalize the output value, and the probability value in the corresponding category is shown in the following formula:

$$\begin{aligned} h_{w,b}(x_i) &= \begin{bmatrix} p(y_i = 1 | x_i; w, b) \\ p(y_i = 2 | x_i; w, b) \\ p(y_i = 3 | x_i; w, b) \\ \cdots \\ p(y_i = n | x_i; w, b) \end{bmatrix} \\ &= \frac{1}{\sum_{j=1}^{n} e^{w_j x_i + b_j}} \begin{bmatrix} e^{w_1 x_i + b_1} \\ e^{w_2 x_i + b_2} \\ e^{w_3 x_i + b_3} \\ \cdots \\ e^{w_n x_i + b_n} \end{bmatrix}. \end{aligned} \tag{6}$$

In classification tasks, it is a common method to use cross-entropy loss function to evaluate the gap between predicted and true values. The cross-entropy formula is as follows:

$$\mathrm{loss} = -\frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{n} y_{ji} \log\left(\hat{y}_{ji}\right). \tag{7}$$

The error from the cross-entropy function needs to be calculated by backpropagation, so as to realize the updated backpropagation of model parameters. The original form of the gradient descent method is shown below:

$$\theta := \theta - \alpha \frac{\partial}{\partial \theta} J(\theta). \tag{8}$$

In the experiments of the following sections, this study also verifies that the use of Adam has faster convergence than SGD (Stochastic Gradient Descent). The mathematical expression of a common Adam optimizer is as follows:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \end{aligned} \tag{9}$$

Therefore, the updating rule of gradient descent is as follows:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{v_t + \varepsilon}} m_t. \tag{10}$$

4. Experimental Results and Analysis

4.1. Experimental Data Collection and Introduction. In order to evaluate the method of music prediction, this study uses 2,903 music from MIREX music genre library, plus 3,397 music downloaded according to the information label in Baidu Music, and we use 5,000 music among them as a training set and others as a test set.

In the music information label, in order to verify the general adaptability of the proposed method to different music information, we used the two-dimensional attribute of genre-emotion as the music label in the experiment. Rock-passion and dance-joy were two-dimensional similar labels,

Figure 2: The dataset of music recommendation.

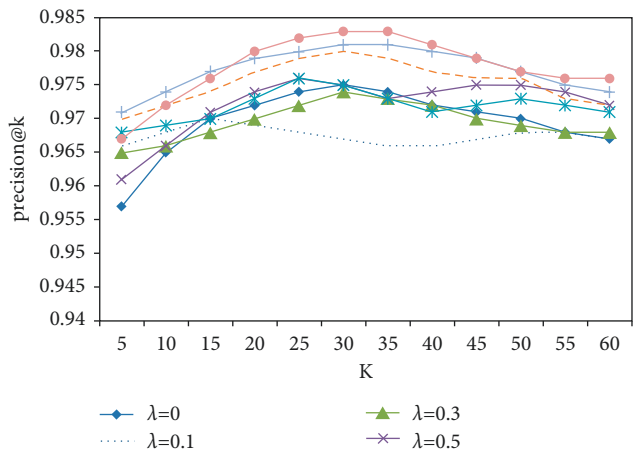

Figure 3: Music recommendation accuracy under different λ.

and these two groups of control labels were used to verify the recommendation effect of the proposed method for similar labels in the actual prediction. For each music file in the experiment, a notation method was adopted, as shown in Figure 2.

4.2. Experimental Results' Analysis. First, in order to determine the proportion of the recommendation results of the three attributes of song title, singer, and album, the laboratory was conducted to analyze the accuracy of their recommendation results and analyze the results. Figure 3 shows that when the value of λ keeps increasing, the proportion of recommendation methods for song attributes decreases, and the accuracy of mixed recommendation increases by λ is 0.7, 0.8, and 0.9. The accuracy of recommendation results is relatively high, indicating that the proportion of attributes is small. Item attribute is to capture the public's preference for an attribute. Even if the user has not heard the song, the user's liking for the song can be analyzed from the user's interest model. Therefore, the recommendation method of social tag is based on association rules, which has a large proportion. When λ equals to 0.9, the mixed recommendation is more accurate.

Figure 4 shows the recommended performance of each model at different noise levels. We tested FCN, Musicnn, sample level + SE, self-attention model, harmonic CNN, and short-block CNN models. We preprocessed the raw input data for each model. Among the four different perturbations considered, dynamic range compression t and white noise addition are the most critical. Musicnn is robust in time extension, but relatively weak in pitch variation. Therefore, harmonic CNN and short-block CNN are the two best original data models. In addition to adding white noise (0.4), harmonic CNN shows better generalization ability for input deformation.

In order to further demonstrate the effectiveness of the proposed method, the music recommendation results of user-oriented feature evaluation are presented in Figure 5. It can be seen from the figure that the music recommendation accuracy of a single user is significantly higher than that of multicategory users under average preference characteristics. The main reason is that the carrier of piano music is relatively small, and the recommendation probability of random piano music is also relatively small. However, the classification features used in the CNN model proposed in this study contain all the music spectrum features, so they can be matched with the title, description, and other text

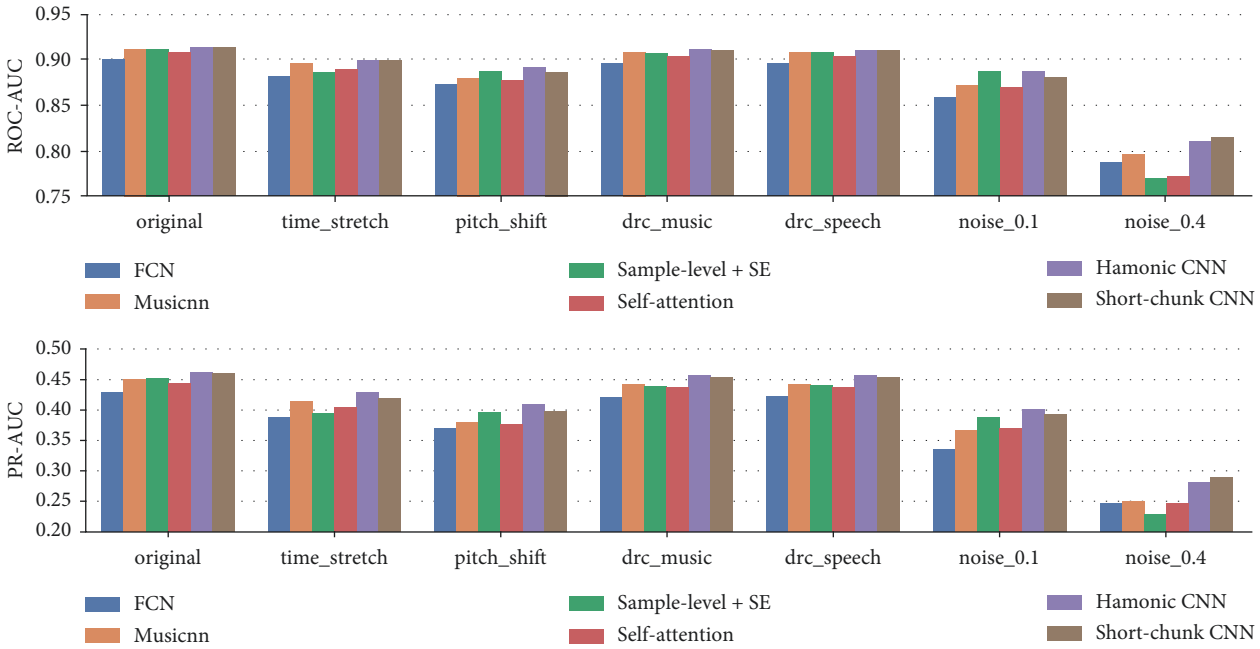

Figure 4: Influence of different noise levels on recommendation accuracy.

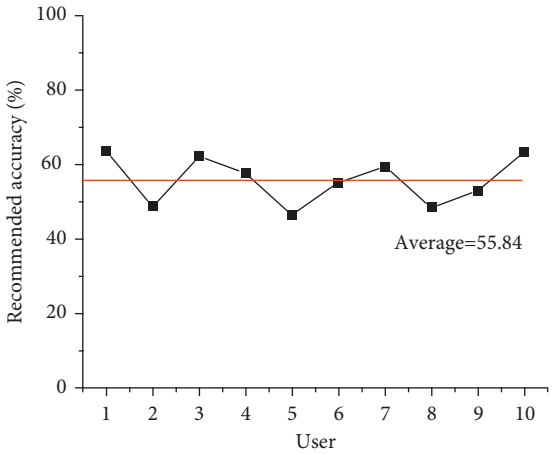

Figure 5: User-specific music recommendation results.

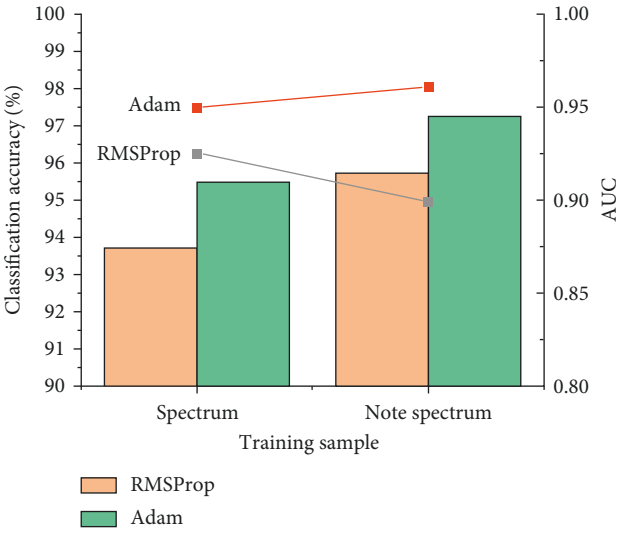

Figure 6: Music recommendation accuracy of spectrum and note spectrum samples.

features of music, so as to form more accurate music recommendation accuracy.

When the learning rate of the model is 0.001, the validity of the model is verified by training and testing samples from the music recommendation dataset under the condition of maintaining two identical gradient descent methods. The specific test results are shown in Figure 6. It can be seen from the figure that, under the same conditions, the training results using fusion feature training samples are significantly better than those using nonfeature music data as training samples. Under RMSProp and Adam optimal controllers, the classification accuracy is improved by 2.0% and 1.25%, respectively. Therefore, music features can improve the music recommendation performance of the CNN model to a certain extent.

In Figure 7, Krepresents the recommended songs. In abscissa K, different values are taken for the recommended songs, and in ordinate precision@k (which means the prediction accuracy in abscissa K), different accuracy rates are recommended according to different values of K. With the single attribute of song, the recommendation accuracy of singers is the highest. With the increase of the number of recommended songs, the recommendation accuracy decreases because the more recommended songs, the greater the probability of error. Attribute album has little influence on users' interests. According to laboratory data calculation, when the value is 1, it means social label recommendation

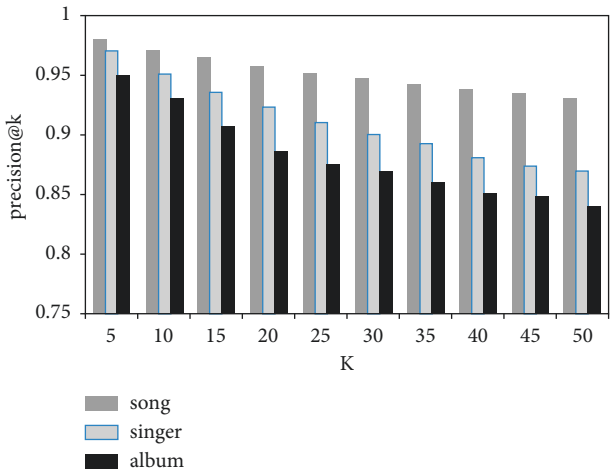

Figure 7: Recommended results for different attributes.

based on association rules; when the value is 0, it means filtering recommendation based on item attributes, which is between 0 and 1, and the optimal solution is obtained through experiments.

5. Conclusions

In this study, we propose a user music recommendation algorithm based on the CNN model, which integrates the advantages of deep schools and considers the music preferences of different users and finally achieves better music recommendation and prediction effect. After user preferences have been obtained, a group of recommended music lists are first obtained by combining existing user data, manual labels, and other recommendation algorithms and then adjusted and sorted according to the user preference classification of each music to achieve more accurate recommendation effect.

Data Availability

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

Conflicts of Interest

The author declares that are no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Lee, H. S. Choi, D. Cho, and H. Lee, "Can digital consumption boost physical consumption? The effect of online music streaming on record sales," *Decision Support Systems*, vol. 135, Article ID 113337, 2020.

[2] M. Sukhavasi and S. Adapa, "Music theme recognition using CNN and self-attention," 2019, https://arxiv.org/abs/1911.07041.

[3] C. Hewitt and H. Gunes, "Cnn-based facial affect analysis on mobile devices," 2018, https://arxiv.org/abs/1807.08775.

[4] S. Allamy and A. L. Koerich, "1D CNN architectures for music genre classification," in *Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 01–07, IEEE, Orlando, Florida, December 2021.

[5] L. Prétet, G. Richard, and G. Peeters, "Learning to rank music tracks using triplet loss," in *Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 511–515, IEEE, Barcelona, Spain, May 2020.

[6] P. Singh, D. Bachhav, and O. Joshi, "Musical instrument recognition using CNN and SVM," *Int. Res. J. Eng. Technol.(IRJET)*, vol. 6, no. 3, pp. 878–912, 2019.

[7] J. Zhang, M. Karkee, Q. Zhang et al., "Multi-class object detection using faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting," *Computers and Electronics in Agriculture*, vol. 173, Article ID 105384, 2020.

[8] H. Wang, X. Zhu, P. Chen, Y. Yang, C. Ma, and Z Gao, "A gradient-based automatic optimization CNN framework for EEG state recognition," *Journal of Neural Engineering*, vol. 19, no. 1, Article ID 016009, 2022.

[9] X. Chen, "The application of neural network with convolution algorithm in Western music recommendation practice," *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, pp. 1–11, 2020.

[10] X. Li, A. S. Young, S. S. Raman et al., "Automatic needle tracking using Mask R-CNN for MRI-guided percutaneous interventions," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 10, pp. 1673–1684, 2020.

[11] S. Song, H. Huang, and T. Ruan, "Abstractive text summarization using LSTM-CNN based deep learning," *Multimedia Tools and Applications*, vol. 78, no. 1, pp. 857–875, 2019.

[12] M. Taenzer, J. Abeßer, S. I. Mimilakis, and C. Weiss, "Investigating CNN-based instrument family recognition for western classical music recordings," *International Society for Music Information Retrie*, vol. 14, pp. 612–619, 2019.

[13] G. Kim, I. Choi, Q. Li, and J Kim, "A CNN-based advertisement recommendation through real-time user face recognition," *Applied Sciences*, vol. 11, no. 20, p. 9705, 2021.

[14] H. Tang and N. Chen, "Combining CNN and broad learning for music classification," *IEICE - Transactions on Info and Systems*, vol. E103.D, no. 3, pp. 695–701, 2020.

[15] D. Yang, J. Zhang, S. Wang, and X. D. Zhang, "A time-aware CNN-based personalized recommender system," *Complexity*, vol. 20, no. 1, pp. 7–19, 2019.

[16] H. C. Ceylan, N. Hardalaç, A. C. Kara, and F. Hardalac, "Automatic music genre classification and its relation with music education," *World Journal of Education*, vol. 11, no. 2, pp. 36–45, 2021.

[17] S. Cheng and Y. Liu, "Eye-tracking based adaptive user interface: implicit human-computer interaction for preference indication," *Journal on Multimodal User Interfaces*, vol. 5, no. 1-2, pp. 77–84, 2012.

[18] M. Dong, "Convolutional neural network achieves human-level accuracy in music genre classification," pp. 1–16, 2018, https://arxiv.org/abs/1802.09697.

[19] D. Ghosal and M. H. Kolekar, "Music genre recognition using deep neural networks and transfer learning," *Interspeech*, vol. 2, pp. 2087–2091, 2018.

[20] D. Kluver, M. D. Ekstrand, and J. A. Konstan, "Rating-based collaborative filtering: algorithms and evaluation," *Social Information Access*, vol. 19, pp. 344–390, 2018.

[21] L. Song and X. Wang, "Faster region convolutional neural network for automated pavement distress detection," *Road Materials and Pavement Design*, vol. 22, no. 1, pp. 23–41, 2021.
[22] L. Zhu, H. Li, and Y. Feng, "Research on big data mining based on improved parallel collaborative filtering algorithm," *Cluster Computing*, vol. 22, no. S2, pp. 3595–3604, 2019.
[23] Q. Hu, Z. Han, X. Lin, Q. Huang, and X Zhang, "Learning peer recommendation using attention-driven CNN with interaction tripartite graph," *Information Sciences*, vol. 479, pp. 231–249, 2019.
[24] J. M. Standley, "Music research in medical/dental treatment: meta-analysis and clinical applications," *Journal of Music Therapy*, vol. 23, no. 2, pp. 56–122, 1986.
[25] B. Xiao, X. Yin, and S. C. Kang, "Vision-based method of automatically detecting construction video highlights by integrating machine tracking and CNN feature extraction," *Automation in Construction*, vol. 129, Article ID 103817, 2021.
[26] G. Gessle and S. Akesson, "A comparative analysis of CNN and LSTM for music genre classification," *Journal of Metals*, vol. 33, no. 3, pp. 245–257, 2019.
[27] C. N. N. Cnn, "Speech emotion recognition using convolutional neural network (CNN)," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 8, pp. 1–20, 2020.
[28] Y. H. Cheng, P. C. Chang, D. M. Nguyen, and C. H. Kuo, "Automatic music genre classification based on CRNN," *Engineering Letters*, vol. 29, no. 1, pp. 17–32, 2020.
[29] F. Ortega, B. Zhu, J. Bobadilla, and A Hernando, "CF4J: collaborative filtering for java," *Knowledge-Based Systems*, vol. 152, pp. 94–99, 2018.
[30] Z. Song, H. Sui, and L. Hua, "A hierarchical object detection method in large-scale optical remote sensing satellite imagery using saliency detection and CNN," *International Journal of Remote Sensing*, vol. 42, no. 8, pp. 2827–2847, 2021.
[31] D. Wadikar, N. Kumari, R. Bhat, and V. Shirodkar, "Book recommendation platform using deep learning," *International Research Journal of Engineering and Technology IRJET*, vol. 7, pp. 6764–6770, 2020.
[32] Y. Wu and W. Li, "Automatic audio chord recognition with MIDI-trained deep feature and BLSTM-CRF sequence decoding model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 2, pp. 355–366, 2019.
[33] 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.
[34] J. Van den Bergh, V. Chirayath, A. Li, J. L. Torres-Perez, and M Segal-Rozenhaimer, "NeMO-net – glabeling of multimodal reference datasets to support automated marine habitat mapping," *Frontiers in Marine Science*, vol. 8, p. 347, 2021.
[35] A. Ferraro, X. Favory, K. Drossos, Y. Kim, and D Bogdanov, "Enriched music representations with multiple cross-modal contrastive learning," *IEEE Signal Processing Letters*, vol. 28, pp. 733–737, 2021.
[36] Y. Dai, S. Yan, B. Zheng, and C Song, "Incorporating automatically learned pulmonary nodule attributes into a convolutional neural network to improve accuracy of benign-malignant nodule classification," *Physics in Medicine and Biology*, vol. 63, no. 24, Article ID 245004, 2018.