

Research Article

Recognition and Classification Model of Music Genres and Chinese Traditional Musical Instruments Based on Deep Neural Networks

Ke Xu 

School of Art and Design, Qingdao University of Technology, Qingdao 266033, Shandong, China

Correspondence should be addressed to Ke Xu; xuke@qtech.edu.cn

Received 19 April 2021; Revised 31 May 2021; Accepted 7 June 2021; Published 22 June 2021

Academic Editor: Shah Nazir

Copyright © 2021 Ke Xu. 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.

The teaching of ideological and political theory courses and daily ideological and political education are two important parts of education for college students. With the iterative update of information technology, the individualized development of students, and the reform and innovation of ideological and political education, higher goals and requirements have been put forward for ideological and political education. Some universities have developed new paths in the teaching model, but they have not considered the evaluation module and paid little attention to their own development. They only paid attention to the fact that it injected fresh blood into the reform of education model and ideological education but ignored the improvement of their own quality. Therefore, with these limitations, the learning effect is not satisfactory. Keeping in view these issues, this article defines the concept of deep learning and ideological and political education of college students as the starting point and then analyzes the new precise and personalized concepts, new forms of intelligent teaching and evaluation, and new models of intelligent learning that deep learning brings to college students' ideological and political education. This is a new path of intelligent linkage with the subject, object, and mediator. It can deepen the reform of the education and teaching mode of individualization, accuracy, interactivity, and vividness of college students' ideological and political education and improve the evaluation and management of college students' ideological and political education. The experimental results of the study showed the effectiveness of the proposed study.

1. Introduction

Music is an abstract art that uses sound as a means of expression to reflect human emotions in real life [1]. Music, as an important component of human spiritual life, has occupied an important position in human daily life. Music can improve concentration, relieve people's pressure on work and study, and be good for physical and mental health [2]; music can bring people aural pleasure and spiritual enjoyment [3], help get rid of bad emotions such as sadness, loneliness, and sadness, and make people full of energy and passion. With the rapid development of Internet technology and digital multimedia technology, digital media resources represented by audio and video have obtained good transmission channels and convenient storage media, and digital music resources and Internet music entertainment

consumer users have shown explosive growth. The era of digital music has arrived.

Music information extraction [4] has become a popular research direction in the field of computer science. Music genre classification is an important research content in the field of music information extraction. Music genre is a notable label that distinguishes music, and it is also the category that listeners pay most attention to and retrieve the most. In the past, the classification of music genres mostly used manual labeling methods. Manual labeling is to ask professionals with a professional background in music and higher musical literacy to label music works by category. With the continuous emergence of music creation and online uploads, the digital music resource library on the Internet has become increasingly large, and manual labeling methods have gradually failed to meet the needs. To classify a

large digital music resource library, if manual labeling is used, it will consume a lot of manpower and time, and the labeling results are more subjective, and the labeling standards cannot be completely unified, which is limited by different professionals who label music. Therefore, the automatic classification of music [5–7] has gradually become a research hotspot for researchers. The automatic classification of music genres can effectively solve the problem of high cost and time-consuming human labeling. Through the algorithm, a unified classification standard can be formulated, and the algorithm can be continuously optimized, and a highly accurate and objective classification result can be obtained.

Due to the limited application of manual extraction of music features [8], the robustness is poor, and it is difficult to describe the deep features and timing characteristics of music. Moreover, in the current music genre classification tasks, traditional machine learning classifiers are mainly used, including BP neural networks, support vector machines, and nearest neighbor algorithm classifiers. Due to its shallow structure, the classifier limits the learning of music features, and it is difficult to extract more effective features to represent music, which affects the accuracy of classification. In recent years, deep neural networks [9–12] have achieved good results in natural language processing, computer vision [13–16], and other research fields. The deep neural network model can automatically learn deeper features from the shallow features and can reflect the local relevance of the input data. Deep learning provides a new solution for the automatic classification of music.

Therefore, this study first studied the music genre recognition [17, 18] and classification algorithm [19, 20] based on deep neural network and improved the algorithm. Compared with the classic algorithm that directly extracts the acoustic features or music features of music and trains with a classifier to obtain the recognition and classification results, this algorithm improves the accuracy of the recognition and classification of music genres. At the same time, for the recognition and classification of musical instruments, this study proposes a Chinese traditional musical instrument recognition and classification algorithm based on the deep belief network in deep learning. The deep belief network is used in the feature extraction task of traditional Chinese musical instrument music [21], which reduces the work of manual extraction and identification of features. At the same time, the recognition and classification effect has also been improved compared with the classic algorithm. The following are the main innovation points of this study:

- (i) Combining Bi-GRU and attention mechanism, a novel music genre classification model is proposed, which can learn more significant music features, thereby improving the accuracy of classification.
- (ii) A Chinese traditional musical instrument recognition and classification algorithm was proposed based on a deep belief network. The deep belief network is used in the feature extraction task of Chinese traditional musical instrument music, which limited the recognition and classification effect.

This study is structured as follows. Section 2 shows the background of the study. The methodology section of the study is given in Section 3 with details in the subsections. Section 4 briefly explains the experiments and results of the study performed. The study is ended in Section 5 which is the conclusion.

2. Background

With the rising and advancements in information technology, the individualized development of students, and the reform and innovation of ideological and political education, higher goals and requirements have been put forward for ideological and political education. The following are the details of this section.

2.1. Blues. It originated from the amateur music of poor black slaves in the south of the United States in the past. It had no accompaniment, but a solo singing with emotional content, and later combined with the European chord structure to form music of singing and guitar alternately. The blues are based on the pentatonic scale, which is composed of five scales arranged in pure fifths.

2.2. Classical. It is the traditional musical art of Western music, and it is music created under the background of mainstream European culture. The most prominent feature of classical music is that its works generally use notation to record the score, so that the rhythm and pitch can be recorded in detail, and it is also conducive to the direct coordination of multiple performers. Many types of musical instruments are used in classical music, including woodwind, brass, percussion, keyboard, bowed, and plucked stringed instruments.

2.3. Jazz. It originated from the blues, combining and absorbing classical music, folk music, and other musical styles on the basis of African music traditions, and gradually formed today's diverse jazz music.

2.4. Country. It originated in the southern United States. It is a kind of popular music with ethnic characteristics. The main characteristics of country music are its simple tune, steady rhythm, being mainly narrative, and a strong local flavor, mostly in the form of ballads, with body, two-part, or three-part form. Country music is mostly solo or chorus, with harmonica, guitar, violin, and other accompaniment. The themes of country music are generally love, country life, cowboy humor, family, God, and country.

2.5. Rock. It originated in the mid-1950s and was developed under the influence of blues and country music. It is characterized by prominent vocals and played with guitar, bass, and drum accompaniment, and keyboard instruments such as electronic organs, organs, and pianos are often used. Rock music has a strong beat, centered on various guitar sounds.

2.6. Metal. It is a kind of rock music, which was developed in Britain and the United States in its early days. Metal music has the characteristics of high explosive power, weight, and speed. Its weight is reflected in the low scale of electric guitars and point bass. The speed is reflected in the beat, the beat of metal music can reach more than 200 BPM, and the beat range of general pop music is only 80–130 BPM. The core instruments of metal music are electric guitar, electric bass, and drums, which control the rhythm and melody.

2.7. Disco. It is a kind of electronic music, which originated from African American folk dance and jazz dance. In rhythm, the characteristics of rock music, jazz, and Latin American music are mixed. As ballroom music, disco is characterized by a strong sense of rhythm, arranged by lively string music. Disco is generally 4/4 shots, and every shot is strong, about 120 BPM.

2.8. Pop. It originated in Britain and the United States in the mid-1950s. Popular music is eclectic, often borrowing elements of other styles of music. But pop music also has its core elements: its structure is relatively short, usually about three minutes.

2.9. Hip-Hop. It originated in New York, USA, when it was popular among African Americans and neighborhood gatherings. Hip-hop consists of two main components: rap and DJing. The performer sings in the way of saying words according to the rhythm of the instrument or synthesis.

2.10. Reggae. It is derived from the popular music of Ska and Rock Steady, which evolved in Jamaica. It is the general term for various dance music in Jamaica.

2.11. Electronic. It is a kind of music made using electronic musical instruments and electronic technology. In electronic music, a variety of genres are often combined, and they are modulated into unique timbres through electronic musical instruments and synthesizers to form a unique style. Commonly used electronic musical instruments include electric guitars, electric basses, synthesizers, and electronic organs.

2.12. Punk. It is simple rock music derived from Garage Rock and pre-punk rock, consisting of three chords and a simple main melody.

3. Methodology

3.1. Feature Sequence Extraction of Music Segment. The process of extracting the feature sequence of the music segment is shown in Figure 1. First, analyze the music file to extract the note feature matrix, then perform the main melody extraction and segment division based on the note feature matrix, and then combine the time points of the segment division and the main melody of the music to

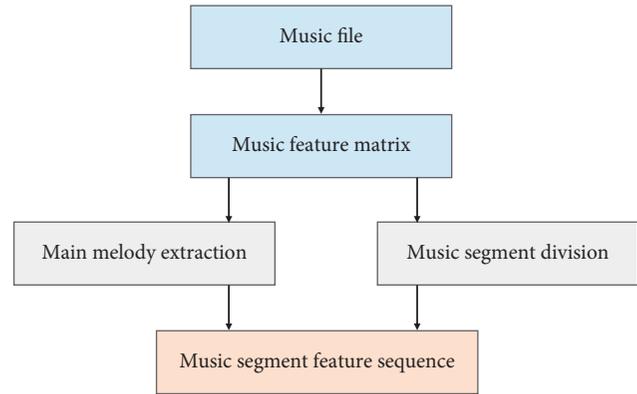


FIGURE 1: The specific process of extracting the feature sequence of music segment.

extract the feature vector based on the main melody for each segment. It composes the feature sequence of the music segment and serves as the input of the later classifier.

3.1.1. Main Melody Extraction. Listening to a piece of music, the perceptual information that people mainly obtain from the sense of hearing is the main melody of the music. The main melody is the soul of music and interprets the theme of music. The main melody of music is the key to music classification and an important basis for distinguishing music genres. This section studies and implements a fast and effective Skyline main melody extraction algorithm for extracting the main melody from music files.

We define the relevant attributes of the notes. Let n_i and n_{i+1} denote two adjacent notes, s_i and s_{i+1} denote the start time of these two notes, respectively, p_i and p_{i+1} denote the pitch of these two notes, respectively, and e_i and e_{i+1} denote the end of these two notes, respectively.

The input of the Skyline algorithm is the note feature matrix. The following describe the specific steps of the Skyline algorithm:

- (1) Arrange the note vectors in the note feature matrix in ascending order of their starting time, and remove the note vectors of channel 10 percussion instruments.
- (2) Traverse the note feature matrix. For note vectors with the same starting time, keep the note vector with the highest pitch and discard other note vectors.
- (3) For two adjacent note vectors n_i and n_{i+1} , if $s_i < s_{i+1}$, $e_i > s_{i+1}$, and $p_i < p_{i+1}$ are satisfied, let $e_i = s_{i+1}$.

3.1.2. Music Segment Division. Firstly, the sound file is sampled, framed, and coded, and the piano roll matrix is used to model the music playing; then the similarity between any two frames is calculated by Euclidean distance to generate a self-similar matrix, and a special Gaussian is constructed. The convolution kernel convolves along the diagonal of the self-similar matrix to generate a novelty curve. The novelty curve is a time-series curve describing the changes in musical performance. Finally, the peak points are

extracted from the novelty curve and segmented. The core idea of the algorithm is to estimate the instantaneous music novelty by analyzing the local self-similarity of music playing; at a significant novel point in time, the music played in the past or the future at that point in time is within a short period of time. It has a high degree of self-similarity, and there is a fairly low cross-similarity between the past and the future at this point in time. Simply put, in a short period of time before this point in time, the musical style of playing is similar. After this point in time, the musical composition is changed to another style of playing. The artistic style of playing music has undergone major changes, and the emotions and themes expressed have also changed, so the music segments can be divided.

3.2. Attention Mechanism. When humans are observing visual images, the human brain quickly scans the images that appear in the field of view and controls the line of sight to fall on the area that you want to focus on. The human brain will allocate different attention to observation according to different areas in the field of view image. For the areas that the field of view focuses on, the human brain will allocate more attention resources to observe carefully to obtain more details of the target area. Information will be ignored for other useless areas of view. The attention mechanism [22, 23] in deep learning is similar to this. It is also a mechanism of attention resource allocation. It can filter out key information that is more conducive to deep learning tasks from a large amount of information, thereby improving the performance of deep learning tasks such as detection [24], prediction [25], and recognition [26, 27].

Figure 2 shows a simplified schematic diagram of the encoding and decoding model that introduces the attention mechanism. The codec model with the attention mechanism can effectively improve its limitations. The encoder no longer converts all the information of the input sequence into a fixed-length context vector. For different outputs, it will focus on finding significant useful information related to the current output from the input data, and calculate different context vectors. Allow the model to better learn the alignment of input and output.

Taken separately, the attention mechanism can be understood as a query calculation process. Figure 3 is a generalized structure diagram of the attention mechanism. x is the input sequence data, and y is the query. First, input y , calculate the attention score of y and each input x_i through the function f , and then map the probability distribution between 0 and 1 through the softmax function. Finally, the probability distribution and each input are correspondingly weighted. Then, calculate the output value of the attention mechanism.

The calculation equation of the attention mechanism is as follows:

$$\text{attention} = \sum_{i=1}^n \text{soft max}(f(x_i, y)) * x_i. \quad (1)$$

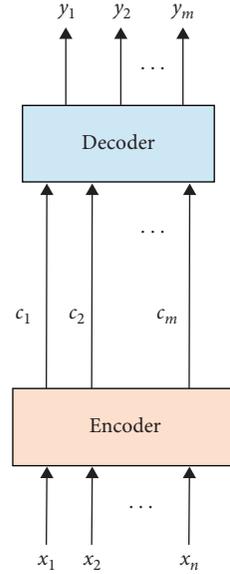


FIGURE 2: Schematic diagram of the codec model with the introduction of the attention mechanism.

3.3. Classification Model. Compared with 2D convolutional networks, 3D convolutional networks can better model time information through 3D convolution and 3D pooling operations. In a two-dimensional convolutional network, the process of convolution and pooling is completed in space. In a three-dimensional convolutional network, they perform in time and space. In the introduction of 3D convolutional network above, it was proposed that images should be output when 2D convolutional network is processing images, and images should also be output when multiple images (which are regarded as different channels) are operated. Therefore, the time information of input data will be lost after each convolution operation in the two-dimensional convolutional network. Only three-dimensional convolution can preserve the time information of the input signal and produce the output quantity. The same principle can be applied to 2D pooling and 3D pooling.

Figure 4 is the network model structure diagram of the classification of music genres in this article. The classification network model designed in this study can be divided into three parts according to different functions, namely, the input layer, the hidden layer, and the output layer. The input of the input layer is a sequence of musical segment features extracted from music. The main function of the hidden layer is to learn the final feature representation of music. The hidden layer is composed of Bi-GRU, attention mechanism, and fully connected layer.

In the attention mechanism, this article uses the following formula to calculate the attention score corresponding to each feature vector:

$$e_t = \tanh(WH_t + b), \quad (2)$$

where e_t is the attention score of the feature vector H_t at the time t in H .

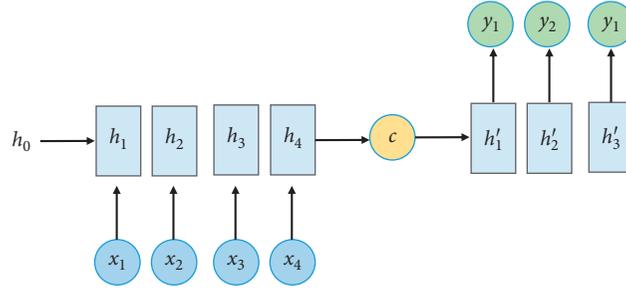


FIGURE 3: Attention mechanism.

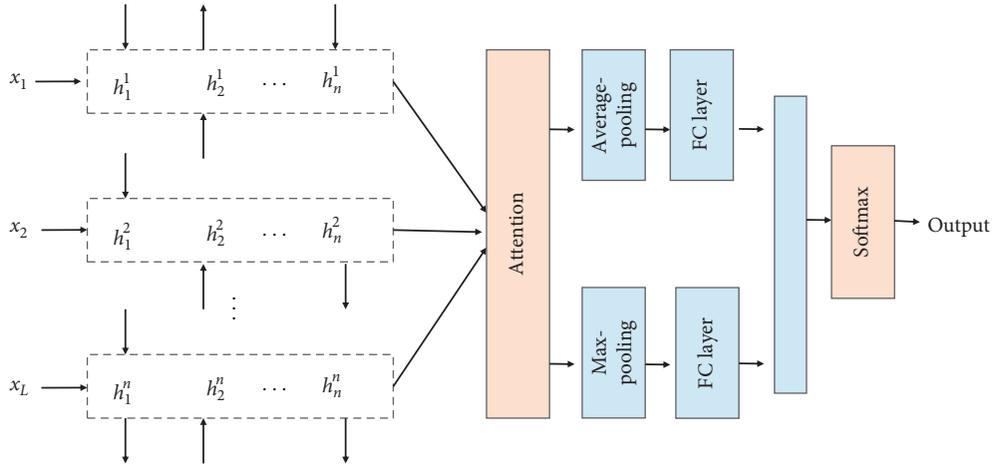


FIGURE 4: Classification model.

Then, the calculated attention score is mapped to the value range (0, 1) through the softmax function, and the attention probability distribution of each feature vector is obtained:

$$a_i = \text{soft max}(e_i) = \frac{\exp(e_i)}{\sum_{k=1}^L \exp(e_k)}. \quad (3)$$

The calculated attention probability distribution and each feature vector of the feature representation H are weighted and summed to obtain the feature vector representation v of the music file:

$$v = \sum_{i=1}^L a_i H_i. \quad (4)$$

The combination of Bi-GRL and the attention mechanism network allows the model to effectively learn the value information of the different weights of the genre classification of each piece of music, including forward and backward value information, and more accurately from the input music. The useful information learned from the segment feature sequence is helpful to improve the accuracy of classification.

At the end of the hidden layer, the music feature vector powder extracted by the attention mechanism of the fully connected layer is used to calculate the confidence score of each genre. The output layer uses the softmax function to

map the output of the hidden layer to the probability of each genre label to which the music file belongs. Finally, the genre tag with the highest probability is selected as the genre tag of the music file.

3.4. Instrument Recognition. This study uses the music signal characteristics of traditional Chinese musical instruments to identify musical instruments. We regard the 2-second-segment musical instrument music signal as a sample, use the MFCC of the sample as the input feature, and input it into a deep belief network with H hidden layers (as shown in Figure 5), and through the output layer, the softmax layer outputs the predicted label of the musical instrument.

4. Experiments and Results

4.1. Cross-Entropy Cost Function. The essence of neural network training is to continuously iterate to minimize the loss function and the process of model parameter convergence. This study uses the cross-entropy loss function to describe the difference between the predicted value output by the network model and the target expected value. The output layer of the network model calculates the probability of each genre through the softmax function and then calculates the cross-entropy loss function. The definition of the cross-entropy loss function is as follows:

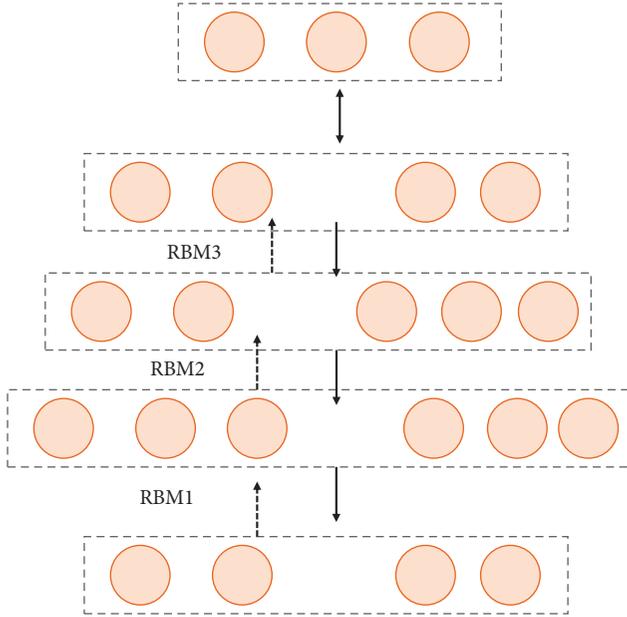


FIGURE 5: DBM-based recognition and classification network structure diagram of traditional Chinese musical instruments.

$$C = \frac{1}{n} \sum_x y \ln(a) + (1 - y) \ln(1 - a), \quad (5)$$

where C represents the loss, n is the number of samples, x is the input sample, a is the output predicted value of the network model input x , and a is the target expected value of the network model input x .

4.2. Adam. In the process of training the network model, the size of the learning rate has an important impact on the improvement of the model's performance, and the learning rate is one of the hyperparameters that are difficult to set. This article uses Adam optimization algorithm as the optimization method of the network model. Adam algorithm is an adaptive learning rate algorithm, which has excellent performance in practice and is widely used. The Adam algorithm designs independent adaptive learning rates for different parameters by calculating the first-order moment estimation and the second-order moment estimation of the gradient. The calculation formula for adjusting the network parameters is as follows:

$$\begin{aligned} v_t &= k_1 v_{t-1} + (1 - k_1) g_t, \\ s_t &= k_2 s_{t-1} + (1 - k_2) g_t^2, \\ \hat{v}_t &= \frac{v_t}{1 - k_1^t}, \\ \hat{s}_t &= \frac{s_t}{1 - k_2^t}, \\ \Delta\theta &= \alpha \frac{\hat{v}_t}{\sqrt{\hat{s}_t + \epsilon}} \end{aligned} \quad (6)$$

4.3. Evaluation Environment. In order to carry out the experiment smoothly, we prepare the experimental data in advance. This article downloads genre-labeled MIDI music files from the Internet dedicated to sharing music, constructs a real data set, and collects a total of 2000 music files. There are 5 genres in the data set, including classical, country, dance, folk, and metal. The number of 1VIIDI music files of each genre is shown in Table 1.

4.4. Experimental Results. A special Gaussian convolution kernel is used to convolve along the diagonal of the self-similar matrix to obtain the novelty curve. After smoothing the novelty curve, the peak point is extracted from it and used as the time point for segment division. The smoothed novelty curve and the extracted peak points are shown in Figure 6.

According to the experimental settings, 6 groups of comparative experiments were carried out. The brief description of the experimental settings is shown in Table 2. By comparing Experiment 1 and Experiment 2, it can be concluded that the classification effect of the extracted feature set input to BP neural network for classification experiment is far lower than the classification effect of the 11 features explored and selected in Experiment 2 according to the genre classification task input to BP neural network. It can be seen that the extracted music features are not suitable for the classification task of music genres in this study, which indicates that feature extraction is not easy to be universal, and feature sets usually need to be constructed according to the actual classification task. Meanwhile, the validity of the feature sets selected in this study in the classification task of music genres is verified.

Comparing Experiment 2 and Experiment 3, we can obtain that, in Experiment 3, we divide the MIDI file into sections, use the section as the analysis unit, extract the features of the section with the same feature combination, form the section feature sequence, and input it into the classification network. Bi-GRL can learn the deeper expression of music about time sequence and semantic information from the input music segment feature sequence, which can effectively improve the accuracy of music classification, and the classification effect is better than the traditional 1VIIDI music classification method based on BP neural network.

Comparing Experiment 4, Experiment 5, and Experiment 6, the music segment is divided into different methods, and the extracted music segment feature sequence will affect the final classification performance. In Experiment 5 and Experiment 6, the music was divided into segments with equal time intervals of 5 seconds and 10 seconds, and the final classification accuracy of the experiment was lower than that obtained in Experiment 4 using the segment division method introduced in this article. The possible reason is that the development of music melody is a process of repetition and change, and there is a certain transition boundary. The division of music with equal duration does not take into account this music characteristic, and the extracted segment feature sequence cannot describe the

TABLE 1: Number of music files in five genres.

Classical	Country	Pop	Rock	Metal	Total
400	386	375	440	399	2000

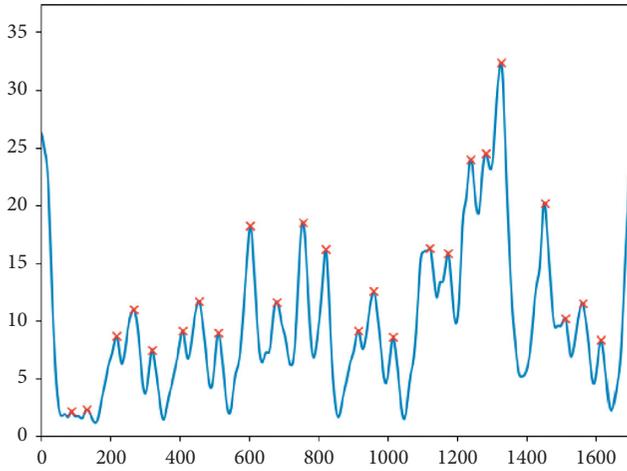


FIGURE 6: Novelty curve and peak point.

TABLE 2: Comparison of experimental results of music classification.

Experiment number	Category	Acc
1	BP neural network + local features	0.75
2	BP neural network + global features	0.86
3	Bi-GRU + dense	0.88
4	Bi-GRU + attention + dense	0.91
5	Bi-GRU + attention + dense (5 seconds)	0.87
6	Bi-GRU + attention + dense (10 seconds)	0.88

music well, so it affected classification performance. In Experiment 4, this study finds the mutation points of music playing to divide the music segment, which can achieve a higher classification effect. The experimental results verify the effectiveness of the music segmentation method used in this study.

5. Conclusion

In this study, we propose a method of music genre classification based on deep learning. According to the characteristic sequence of the input music segment, the cyclic neural network and attention mechanism are studied, and the Bi-GRU and attention mechanism are used to design the classification network model. Bi-GRU is good at processing sequence data. It can learn the contextual semantics and deep features of music from the sequence feature sequence. The attention mechanism is added to automatically assign different attention weights to the features learned by Bi-GRU from different segments and learn more significant music features, thereby improving the accuracy of classification. In

addition, this study also proposes a recognition and classification algorithm for traditional Chinese musical instruments based on deep belief networks. The experimental results of the study have achieved credible results.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author has no conflicts of interest regarding the publication of this study.

References

- [1] J. A. Sloboda and P. N. Juslin, "Psychological perspectives on music and emotion," *Music and Emotion: Theory and Research*, pp. 71–104, Oxford University Press, Oxford, UK, 2001.
- [2] A. Kresovich, M. K. Reffner Collins, D. Riffe, and F. R. D. Carpentier, "A content analysis of mental health discourse in popular rap music," *JAMA Pediatrics*, vol. 175, no. 3, pp. 286–292, 2021.
- [3] T. Eerola, J. K. Vuoskoski, H.-R. Peltola, V. Putkinen, and K. Schäfer, "An integrative review of the enjoyment of sadness associated with music," *Physics of Life Reviews*, vol. 25, pp. 100–121, 2018.
- [4] Y. Li, W. Hu, and Y. Wang, "Music rhythm customized mobile application based on information extraction," in *Proceedings of the 4th International Conference on Smart Computing and Communication*, pp. 304–309, Birmingham, UK, October 2019.
- [5] F. Medhat, D. Chesmore, and J. Robinson, "Automatic classification of music genre using masked conditional neural networks," in *Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM)*, pp. 979–984, IEEE, New Orleans, LA, USA, November 2017, In press.
- [6] S. Vishnupriya and K. Meenakshi, "Automatic music genre classification using convolution neural network," in *Proceedings of the 2018 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–4, IEEE, Coimbatore, India, January 2018.
- [7] S. Shetty and S. Hegde, "Automatic classification of carnatic music instruments using MFCC and LPC," *Data Management, Analytics and Innovation*, Springer, Singapore, pp. 463–474, 2020.
- [8] Y. T. Chen, C. H. Chen, S. Wu, and C. C. Lo, "A two-step approach for classifying music genre on the strength of AHP weighted musical features," *Mathematics*, vol. 7, no. 1, 19 pages, 2019, In press.
- [9] R. Liu, X. Ning, W. Cai, and G. Li, "Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification," *Mobile Information Systems*, vol. 2021, Article ID 9962057, 15 pages, 2021.
- [10] C. Yan, G. Pang, X. Bai, Z. Zhou, and L. Gu, "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, 2021.
- [11] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE Journal*

- of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4561–4572, 2021, In Press.
- [12] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, “Polynomial fitting algorithm based on neural network,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.
- [13] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, “Feature refinement and filter network for person re-identification,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [14] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, “Remote sensing image recognition based on multi-attention residual fusion networks,” *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [15] X. Zhang, Y. Yang, Z. Li, X. Ning, Y. Qin, and W. Cai, “An improved encoder-decoder network based on strip pool method applied to segmentation of farmland vacancy field,” *Entropy*, vol. 23, no. 4, p. 435, 2021.
- [16] X. Ning, X. Wang, S. Xu et al., “A review of research on co-training,” *Concurrency and Computation: Practice and Experience*, 2021.
- [17] D. Bisharad and R. H. Laskar, “Music genre recognition using convolutional recurrent neural network architecture,” *Expert Systems*, vol. 36, no. 4, Article ID e12429, 2019.
- [18] S. Iloga, O. Romain, and M. Tchuente, “A sequential pattern mining approach to design taxonomies for hierarchical music genre recognition,” *Pattern Analysis and Applications*, vol. 21, no. 2, pp. 363–380, 2018.
- [19] S. Oramas, F. Barbieri, O. Nieto, and X. Serra, “Multimodal deep learning for music genre classification,” *Transactions of the International Society for Music Information Retrieval*, vol. 1, no. 1, pp. 4–21, 2018.
- [20] H. Bahuleyan, “Music genre classification using machine learning techniques,” 2018, <http://arxiv.org/abs/1804.01149>.
- [21] R. Yang, L. Feng, H. Wang, J. Yao, and S. Luo, “Parallel recurrent convolutional neural networks-based music genre classification method for mobile devices,” *IEEE Access*, vol. 8, pp. 19629–19637, 2020, In press.
- [22] W. Cai and Z. Wei, “Remote sensing image classification based on a cross-attention mechanism and graph convolution,” *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020, In Press.
- [23] W. Cai, B. Liu, Z. Wei, M. Li, and J. Kan, “TARDB-Net: triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification,” *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 11291–11312, 2021.
- [24] Z. Chu, M. Hu, and X. Chen, “Robotic grasp detection using a novel two-stage approach,” *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [25] W. Sun, P. Zhang, Z. Wang, and D. Li, “Prediction of cardiovascular diseases based on machine learning,” *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.
- [26] L. Sun, W. Li, X. Ning, L. Zhang, X. Dong, and W. He, “Gradient-enhanced softmax for face recognition,” *IEICE Transactions on Information and Systems*, vol. E103.D, no. 5, pp. 1185–1189, 2020.
- [27] Y. Zhang, W. Li, L. Zhang, X. Ning, L. Sun, and Y. Lu, “AGCNN: adaptive gabor convolutional neural networks with receptive fields for vein biometric recognition,” *Concurrency and Computation: Practice and Experience*, Article ID e5697, 2020, In press.