A Music Genre Classification Method Based on Deep Learning

Qi He

1Arts, Hubei University of Education, Wuhan 430205, China 2UCSI University Faculty of Social Sciences and Liberal Arts, Cheras 56000, Kuala Lumpur, Malaysia

Correspondence should be addressed to Qi He; 1002164199@ucsiuniversity.edu.my

Received 20 January 2022; Revised 16 February 2022; Accepted 12 March 2022; Published 29 March 2022

Copyright © 2022 Qi He. 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.

Digital music resources have exploded in popularity since the dawn of the digital music age. The music genre is an important classification to use when describing music. The function of music labels in discovering and separating digital music resources is crucial. In the face of a huge music database, relying on manual annotation to classify will consume a lot of cost and time, which cannot meet the needs of the times. The following are the paper’s primary research findings and innovations: to better describe the music, this article will be divided into multiple local musical instrument digital interface (MIDI) music passages, playing style close by analyzing passages, passages feature extracting, and feature sequence of passages. Extraction of note feature matrix, extraction of topic and segment division based on note feature matrix, research and extraction of effective features based on segment theme, and composition of feature sequence are all part of the process. Because of the shallow structure of standard classification methods, it is difficult for classifiers to learn temporal and semantic information about music. This research investigates recurrent neural networks (RNN) and attention using the distinctive sequence of input MIDI segments. To create data sets and conduct music categorization tests, collect 1920 MIDI files with genre labels from the Internet. The method for music classification is validated when it is combined with the experimental accuracy of equal length segment categorization.

1. Introduction

Music carriers are constantly replaced, and physical albums are gradually replaced by digital music [1]. Because different people have their own preferences for the music genre, music genre recognition technology can further help all kinds of music software and accurate and efficient music genre recognition is the cornerstone of audience management, collection, and recommendation system. Considering that different music has different styles, accompaniment instruments, and singing characteristics, Sturm proposed that characteristics of music can be summarized and extracted according to different characteristics of music, and then it can be divided into different schools [2].

Bogdanov et al. used a neural network algorithm to provide new ideas for music genre classification through computer training and learning [3], but it has defects such as slow convergence speed and falling into local optimal. Previous scholars have done a lot of work. Dannenberg et al. proposed to improve particle swarm optimization algorithm through a genetic algorithm to optimize BP neural network [4], which significantly improved the classification accuracy, but it was imperfectly improved in the aspect of music feature extraction. Considering that more accurate extraction of music features also has a significant impact on the accuracy of classification, this paper extracts the information of music data from the basic elements of music.

Under the new background of fierce competition in the digital music market, massive digital music resource library stored in the cloud, and increasing digital music users, the importance of music labels becomes prominent. Music label refers to a phrase that can summarize music content at a high level and plays a good role in identifying and dividing digital music resources. Music labels divide music with similar characteristics such as musicology, region, and era into broader categories [5]. Common music label classification includes music genre classification, music writer classification, music emotion classification, music applicable scene classification, and so on. Taking music genre classification as an example, digital music resources are classified into genres.
It can be divided into Pop, Classical, Rock, Jazz, Dance, Country, Blues, Folk, and other music genres. Music of the same genre has similar artistic styles.

Automatic classification of music, in today’s Internet era and digital music era, has important and extensive application fields:

1.1. Storage Management. Music labels can significantly divide the huge digital music resource library on the Internet, facilitate organization and management, and facilitate distributed storage of music resources, fast and convenient resource positioning, and download.

1.2. Search Engines. The correct classification of music work types brings great convenience to Internet music search engines, music lovers, or audience users. Music search engines provide faster and more accurate search results based on the type of music works; users can also according to the type of music works, rapid retrieval, access to the required resources and information.

1.3. Recommendation System. Music streaming media service provider according to music users to listen to music take advantage of user interest preference, music property, content, categories, mining the user’s preferences and needs, actively refer users to its interest or you need to type of music, to improve the user’s subscription will pull the economic growth of digital music market.

1.4. Music Creation. According to the composer input by the user, music genre, emotion, and the keys pressed, the automatic composition can be realized by using the technology of human intelligence, and auxiliary suggestions of composition and arrangement can also be provided for the creators.

The recognition and classification of music genres play an important role in music information retrieval. Many music users have a strong interest in specific genres of music, and the function of the music genre identification classification system is to divide music into different types according to the wind and to recommend music based on the user’s interest so that users can enjoy music for quick retrieval and efficient management. Most music pieces are sung by people and accompanied by various musical instruments. In addition, the structural characteristics of musical genres vary from genre to genre, and even the same person can sing different sounds in different ranges when performing different musical genres. Many factors make it difficult for people to extract features of music signals, which makes it difficult to improve the accuracy of music genre identification and classification. Therefore, in recent years, the recognition and classification of music genres have been widely concerned and developed rapidly. How to further improve the accuracy and efficiency of music genre recognition and classification has become the focus of current research in this field [6].

Most music songs contain the corresponding musical instrument audio, so the effective recognition and classification of musical instruments can provide strong support for music information retrieval. Therefore, the recognition and classification of musical instruments is also an important part of music information retrieval, which is of great significance to music retrieval. For people, the recognition and classification of musical instruments are relatively easy to complete, as long as the person has a strong musical accomplishment, it can be more accurate recognition and classification of musical instruments, but most of the ordinary music audience does not have this ability, so it is necessary to teach the computer how to automatically identify and classify musical instruments. From an acoustic point of view, the timbre of an instrument is the main basis for distinguishing one instrument from another, accurately marking the unique characteristics of each instrument. The timbre of a musical instrument is mainly determined by the vibration of the articulating part of the instrument. Different vibration states lead to different overtones and waveforms, and the proportion of harmonics in overtones determines the timbre of musical instruments. However, the proportion of overtones in the variable, the same instrument using different playing techniques, in the timbre can also show significant changes, can also be mistaken for the sound of other instruments. This will also lead to difficulties in the feature extraction of musical instruments from musical signals, which will lead to low accuracy in the recognition and classification of musical instruments. Although there is little research on musical instrument recognition classification, how to improve the accuracy of musical instrument recognition classification will become the focus of research in the field of music information retrieval.

The cultural importance of technological applications in music is significant. Instruments’ spiritual significance, the affective force of amplification, the cultural repercussions of recording distribution, and the creative and expressive possibilities given by numerous mediums are just a few examples. The results of the classification can be used to support sociological and psychological research into how people construct the sense of musical similarity and form musical groupings, as well as how this relates to the “objective” truth generated by machine classifiers.

The two emphases of the research on music genre and musical instrument recognition and classification are the extraction of relevant features and the design of the used classifier. There is no unified standard of what kind of characteristic quantity to choose to do recognition and classification effect. In order to improve the accuracy of recognition and classification, some researchers start from the principle of signal generation to find new and effective features, and some researchers integrate various existing single feature quantities for recognition and classification. Although the above method is to a certain extent improve the recognition accuracy, they do not have unity, and sometimes doing some general classification tasks do not even know what are the required characteristics, then using artificial feature extraction methods to identify point’s class task will be extremely difficult, and the superiority of using
deep learning algorithm was obvious. Different from other machine learning models, deep learning can simulate the structure of the human brain, store, and process a large amount of information, and mine the correlation of data’s internal information, that is, extract more essential features of data, so as to improve the recognition and classification performance.

At present, deep learning is widely used in the field of image processing but rarely used in the field of audio, especially in the field of music information retrieval. Therefore, this paper first studies the music genre recognition and classification algorithm based on a deep confidence network in deep learning and improves the algorithm. Compared with the classical algorithm, which directly extracted the acoustic features or musical features of music and trained the classifier to obtain the recognition and classification results, the algorithm improved the recognition and classification accuracy of music genres. At the same time, in the field of musical instrument recognition and classification, especially in the field of traditional musical instrument recognition and classification, which is unique to China, there is no paper that uses deep learning for musical instrument recognition and classification. Therefore, this paper also proposes the Chinese traditional musical instrument recognition and classification algorithm based on a deep confidence network in deep learning. In the feature extraction task of traditional Chinese Musical Instruments, the deep confidence network is applied to reduce the workload of manual feature extraction and recognition and classification, and the recognition and classification effect is improved compared with the classical algorithm.

The research is organized as follows: the related work is presented in Section 2. Section 3 analyzes the MIDI segment division. Section 4 discusses the music genre classification experiment in detail. Finally, in Section 5, the research work is concluded.

2. Related Work

Most scholars choose to use MPEG Audio Layer-3 (MP3), Waveform Audio File Format (WAV), and other audio files that store waveform for research and relatively little research on the classification of MIDI music. MIDI files are structured data formats that store events instead of waveforms, including event time information, event type, event content, etc.

2.1. Five Characteristics of Music. Under the background of music culture of different schools and different emotions, religious themes are expressed through five basic elements of music, such as jazz rhythm, extremely complex disco rhythm, strong metal music rhythm, exciting happy music rhythm, and lively music scheduling, while our common tones, lonely and sad music tones, are usually deep and chaotic.

2.1.1. Rhythm. Different genres of music can be distinguished by the speed and strength of music rhythm. For example, pop and rock have stronger and faster rhythms than jazz and blues and have stronger regularity. Beats per minute can reflect the speed of music rhythm. A pulse sequence of music signal can be regarded as a signal of a fixed number of beats. The pulse sequence corresponding to each certain number of beats is obtained. Through cross-correlation calculation between each known pulse sequence and the measured signal, the corresponding metronomic value of the pulse sequence with the largest calculation result is selected as the metronometric number per minute of the measured musical signal. To pick up the music signal envelope, the envelope marked contour peak, to calculate the whole period of music envelope spikes per second, as peak frequency, used to measure the strength of the music rhythm, high peak frequency shows that the characteristics of the music have a strong beat, generally speaking, this kind of music to rock a new class.

2.1.2. Tunes. Tunes said tones of change, from the perspective of the audio signal, the pitch is shown by the frequency of the sound signal, namely the vocal cord vibration frequency, for Hertz, this literary Grace use frequency-domain expects to represent tones, convert data by the Fourier transform to frequency domain signal, to deal with the noise of the data, get the music frequency domain average, if the music of the mean is larger, it indicates that the tone of the song is higher; conversely if the mean value is small, it indicates that the tune is lower, which is used to indicate the level of the tune of the music.

2.1.3. Harmony. Harmony is the sound combination of two or more sounds produced at the same time. It contains two forms of musical instrument chord and human voice overlap. Harmony has the color function of regulating the thick and light, thick and thin music. Based on the above acoustic knowledge, this paper adopts the time-domain variance of music information to represent harmony. When the time-domain variance is large, it indicates that the energy information of sound changes greatly, and it is judged that instrument ensemble or human voice overlap occurs at this time. When the time domain variance is small, it indicates that the energy information of the sound changes little, and it is judged that the sound is played by a single instrument or a single person at this time. Based on the above analysis of the time-domain mean, the variance in the time domain is considered mathematically to characterize whether harmony occurs.

2.1.4. Sound Strength. Also called loudness, volume, unit of decibels. In this paper, the short-time energy feature of music information is used to characterize the sound intensity of the music. By calculating the short-time energy feature in the music message frame, the size of sound intensity is represented. The larger the short-time energy feature is, the larger the energy contained in this time interval is, the larger the corresponding sound intensity is; conversely, the smaller the short-time energy feature is, the smaller the sound intensity is.
2.1.5. Strength. Since the timbre of the singer and the timbre of the instrument are different in different schools of music, they can be distinguished by timbre elements. Malta Fairs and convention center (MFCC) [7] is a commonly used feature in speech recognition.

2.2. Research Status. Music genre recognition and classification is a very important part of music information retrieval. It has been studied extensively since the 1990s. In 1995, Li and Tzanetakis [8] and others put forward the music signal frequency domain analysis method, the fast Fourier transform was carried out on the audio data, and then a log scale transforms to get data as characteristics of the data, input into contains two hidden layers neural network training, the last two music genre of classical and pop music for identification.

In recent years, with the deep learning technology in the field of image recognition, speech recognition has made excellent achievements. More and more scholars begin to explore the application of deep learning technology to music information retrieval. In 2009, Xu et al. [9] applied a deep confidence network to music genre classification for the first time, and experimental data results show that the performance of this method is superior to traditional machine learning methods. Tao [10] will limit the Boltzmann chance use music genre, constructs a 5 layer limit Boltzmann machine, but this method has an obvious defect is only to be more than 50% in four music genre classification accuracy, along with the increase in the number of music genre types of classification, the classification accuracy rate goes down.

In 2006, Tao Li et al. proposed a new feature extraction method for music genre recognition and classification based on wavelet transform theory. They used statistical methods to calculate the statistical value of wavelet coefficients and then obtained features. At the same time, they used different pattern recognition and classification methods, such as support vector machine (SVM), Gaussian mixture models (GMM), linear discriminant analysis (LDA), and nearest-neighbor (KNN). In 2010, Panagakis et al. [11] proposed an unsupervised dimension reduction non-negative multilinear principal component analysis (NMPCA) method for third-order tensors and compared it with other commonly used multicomponent quantum space analysis methods for recognition and classification, and the results proved that NMPCA could extract more essential features. In 2011, Seo and Lee [12] proposed a high-order moment feature in the recognition and classification of music genres, that is, the third-order statistical features of the spectrum are fused to form segel-level features. Although this feature can improve the accuracy of recognition and classification, it increases the dimension of feature vectors and the calculation becomes much more complicated [5].

Compared with foreign countries, the research on music genre recognition and classification started late in China, but it has also made some progress. In 2009, Zhen Chao et al. used a forward feature selection algorithm to screen the underlying features, including root mean square energy, zero-crossing rate, strongest beat, strongest beat intensity, spectrum center value, spectrum flow, spectrum variability, cepstrum coefficient, and linear prediction coefficient, when identifying and classifying music genres. In addition, a multimodal music genre recognition and classification method integrating music labels were proposed, which completed the recognition and classification tasks of eight music genres including blues, country, classical, pop, jazz, electronic reggae, and rock, achieving a recognition rate of nearly 87% [13]. In 2011, McKinney and Breebaart [14] used the two-dimensional feature map of pitch and rhythm to calculate feature vectors to represent the melody of music, so as to realize the recognition and classification of classical, jazz, folk music, rock, and pop music genres, and the recognition rate reached 81% [15, 16].

In addition to the above-listed classical algorithms at home and abroad, there are actually much effective music genre recognition and classification algorithms. The research focus of various algorithms lies in how to extract as many features as possible that can reflect the essential attributes of music signals and how to design a classifier with better performance to achieve the purpose of optimizing the recognition and classification performance. The core idea of these methods is to manually extract as many low-level or high-level features of music as possible and use a variety of classifiers for recognition and classification and select the most suitable features and classifiers by comparison. One drawback of manual extraction of music features is that different algorithms are needed for different features. The algorithms are not universal, and the proposal of new features is also a big difficulty. In this paper, the music genre recognition and classification algorithm based on a deep communication network is studied [17]. As long as the bottom characteristics of the input music signal, such as multilinear principal component (MPC), fast Fourier transform (FFT), Malta Fairs, and convention center (MFCC), the network will automatically start to learn and analyze the input characteristic data, so as to get more appropriate abstract characteristics reflecting the essence of each genre of music. In this way, the complexity of feature extraction is greatly reduced, and the performance of music genre recognition and classification is better and more intelligent [18].

According to the literature retrieved at present, similar to the recognition and classification of music genres, research on musical instrument recognition and classification mainly focus on feature extraction algorithms and classifier design. However, deep learning, as a new popular feature extraction technology, has not been widely applied in the field of musical instrument recognition and classification, let alone in the research of Chinese traditional musical instrument recognition and classification. Therefore, this paper proposes a Chinese traditional musical instrument recognition and classification algorithm based on a deep confidence network in deep learning [19].

The label of musical instrument is not only of great significance to the classification of music type but also can be used to predict the emotion and music scene contained in the music, so the identification and classification of the musical instrument also play an important role in the field of
MIDI file sampling, frame, coding, generation of piano curtain matrix

Calculate the similarity between any two frames and generate the self-similarity matrix

A special Gaussian convolution kernel is constructed and convolved along the diagonal of the self-similar matrix to generate the novelty curve

The time point of extracting peak value of novelty degree curve is used as the segment division point

**Figure 1:** Basic flow chart of MIDI segment division.

Music information retrieval. If we know which instrument plays a particular piece of music, then we can optimize the performance of automatic music classification according to the characteristics of the instrument used. Because the study of musical instrument recognition and classification can help people to study the recognition and classification of other music information retrieval fields, the research work of musical instrument recognition and classification has attracted more and more attention of scholars recently.

### 3. MIDI Segment Division

This paper studies and draws on Chen Delong’s method of thinking to divide MIDI music into segments. This method draws lessons from Foote’s music segmentation method based on local self-phase and is applied to MIDI music. The time point of the music playing mutation is calculated, and the MIDI music is divided into several segments that play similar parts and express certain emotions and themes.

Figure 1 is the basic flow chart of the MIDI segment algorithm. A specific process for the first MIDI file sampling, frame, and coding, piano curtain matrix for music playing modeling, and then the Euclidean distance is used to calculate the similarity between any two frames and generate the self-similarity matrix. Then a special Gaussian convolution kernel is constructed and convolved along the diagonal of the self-similar matrix to generate the novelty curve, which is a time-series curve describing the variation of music playing. Finally, the peak points of the novelty curve were extracted and segments were divided. At a significantly novel time point, the musical performance has a high degree of self-similarity in the past or the future, and a low degree of cross-similarity between the passing of the time point and the future. The artistic style of music playing has changed greatly, as well as the emotion and theme expressed. Therefore, MIDI can be divided into sections. Through the algorithm introduced in this section, MIDI music can be divided into several segments.

#### 3.1. Piano Curtain Matrix

In computers, the performance of MIDI music can be effectively described by a Piano Roll. On the horizontal axis is time, on the vertical axis is the pitch of notes, and each note played can be represented by a horizontal bar. Therefore, according to the piano curtain diagram, MIDI music sampling, frame, coding, generation of piano curtain matrix, MIDI music modeling, each column of the piano curtain matrix records the notes played at the current sampling time and the corresponding volume. In this paper, the specific steps of generating steel violin curtain matrix are as follows:

1. Input the note feature matrix of the MIDI file and sort the note vectors in ascending order according to the starting time.
2. MIDI music is sampled by $dt$ at a certain interval, and the MIDI music is divided into $M$ frames, and each frame corresponds to a sampling moment. Initialize the piano curtain matrix $G$, whose dimension is $128 \times M$, element values are all 0, and the serial number of the noted vector currently sampled is $l = 1$. Get the number of note vectors $N$ of the note feature matrix.
3. Divide the start time $t_{\text{start}}$ and end time $t_{\text{end}}$ of the $i^{\text{th}}$ note vector by interval $dt$, respectively, to obtain the start sequence number $n_{\text{start}}$ and end sequence number $n_{\text{end}}$:

\[
\begin{align*}
\quad n_{\text{start}} &= t_{\text{start}}/dt, \\
\quad n_{\text{end}} &= t_{\text{end}}/dt.
\end{align*}
\]

4. According to the $i^{\text{th}}$ note vector, the value of $n$ is at $n_{\text{start}}, n_{\text{start}} + 1, \ldots, n_{\text{end}}$ traversal, assign value to piano curtain moment $G$:

\[ G\{\text{pitch}\}[n] = \text{volume}. \]

5. Let $i \leftarrow i + 1$. If $i < N$, return to step (3); If $i = n$, then the generation of the piano curtain matrix is completed.

#### 3.2. Self-Similar Matrix

In the previous section, we obtained the piano curtain matrix of MIDI, which is a two-dimensional matrix with a dimension of $128 \times M$. Each column represents the coding array of each frame obtained from periodic sampling, and there are a total of $M$ columns. The encoding array of each frame can be regarded as a vector in 128-dimensional space.

Euclidean distance is used as similarity measurement to calculate the similarity of any two frames and finally generate the self-similarity matrix $S$. The self-similar matrix $S$ is a two-dimensional matrix with dimension $M \times M$. Since the Euclidean distance is used to calculate the similarity, the self-similarity matrix $S$ is a symmetric matrix with a diagonal similarity of 0. Therefore, we only need to calculate the similarity of the upper triangle part of the upper self-similarity matrix $S$, and all the similarities of the whole self-similarity matrix can be obtained by symmetry properties. Calculate their similarity:
$$D(i, j) = \sqrt{\sum_{k=1}^{128} (x_{ik} - x_{jk})^2}.$$ (3)

3.3. Novelty Function and Peak Point. Novelty function is a time series curve that describes the variation of music playing, and the value of each moment is novelty degree. Each novelty value represents the degree of change in local musical playing style before and after the time point. In this section, we introduce how to obtain the novelty function and get the segment partition points. In the previous section, we obtained the self-similarity matrix, which is the key to calculating the novelty of each sampling moment. First, we construct such a simplest $2 \times 2$ basic convolution kernel, which can be divided into two parts:

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (4)$$

In Formula (4), the former part is used to calculate the mutual similarity between the two regions before and after the time point, and the latter part is used to calculate the self-similarity within the respective regions of the two regions before and after the time point. The novelty degree of the music at the central time point is calculated. If the mutual similarity between the two regions before and after a certain time point is low, but the self-similarity within each region of the two regions is high, the center of the convolution kernel is located at this time point, and the calculated novelty value will be high, and this time point can be used as the segment division point.

Next, find out the novelty of each sampling time point:

$$N(i) = \sum_{m=-k}^{k} \sum_{n=-k}^{k} C(m, n)S(i + m, i + n). \quad (5)$$

It can be seen from Formula (5) that when the convolution calculation exceeds the boundary of the self-similar matrix, the novelty function $N(i)$ cannot be calculated. We can fill the self-similar matrix, but because the self-similar matrix represents the similarity of music at any two sampling moments, filling the self-similar matrix cannot meet this characteristic. Therefore, the improvement method adopted in this paper is to fill the piano curtain matrix and add $k$ column 0 values at the beginning and end of the piano curtain matrix. It can be understood that the music is silent and stops playing during the filling time of the beginning and end. After filling, the self-similar matrix is calculated and the novelty degree function is generated according to the above steps.

3.4. Generation of Segment Feature Sequences. In the first two sections, we extracted the main melody of MIDI music and divided the MIDI music into sections to obtain several sections, each section contains several main melody note vectors. However, the genre theme information reflected by each piece of music cannot be directly perceived by the computer, it can only be expressed by extracting the corresponding characteristic parameters. In this section, we extract the music feature parameters from the segments divided by MIDI and construct the segment feature sequence as the input of the deep neural network for music classification.

After dividing MIDI music into multiple sections, we combined the perceptual process of music, the way of expression of music, and the elements of grouping carried out experimental exploration, and finally selected some features with great distinction and good expressiveness to music genres. The combination of these features can better describe the passage, and these features are used to form the feature vectors of the passage. Finally, the feature vectors of each passage are formed into the feature sequence of the passage in chronological order. These features and extraction methods are introduced below.

3.4.1. Average Pitch. Pitch refers to the level of sound, which is determined by the vibration frequency of the sound source. Pitch is one of the components of musical melody. The average pitch describes the pitch level of a section and plays an important role in the performance of the musical melody.

$$p_{avg} = \frac{\sum p_i}{n}. \quad (6)$$

3.4.2. Stability of Pitch. The music fluctuates over time, and the themes and emotions expressed also change. To a certain extent, the stability of pitch can reflect the expression techniques of musical works. Pitch stability is calculated by the pitch and average pitch of each note in the passage.

$$p_{std} = \sqrt{\frac{1}{n} \sum_{i} (p_i - p_{avg})^2}. \quad (7)$$

3.4.3. Vocal Range. Register describes the pitch span of a passage, indicating the distance between the highest and lowest notes of a note in the passage.

$$p_r = \max(p) - \min(p). \quad (8)$$

3.4.4. Playing Speed. The speed of music performance is an important component of music and has an important influence on the classification of musical genres. This paper describes the performance speed of the passage by adopting the two features of the partition in the MIDI head block and the weighted average of beats per minute (BPM) of the main melody in the passage. It can be seen from the above that division defines the time resolution of MIDI and establishes the global speed benchmark of MIDI music. The weighted average of the playing speed of the main melody notes in the passage can be calculated as follows:
3.4.5. **Strength.** Intensity is the basic elements of melody, closely related to the emotion and thought expressed in music, and is an important means of expression in music works. This paper describes the dynamics information of music by means of the average dynamics and stability of the main melody notes played in the passage.

\[
\text{tempo} = \frac{\sum_i \text{BPM}_i \times D_i}{\sum_i D_i},
\]

\[
I_{\text{avg}} = \frac{\sum_i I_i}{n},
\]

\[
I_{\text{std}} = \sqrt{\frac{1}{n} \sum_i (I_i - I_{\text{avg}})^2}.
\]

### 4. Music Genre Classification Experiment

The musical techniques, the cultural setting, and the content and spirit of the subjects can all be used to identify a music genre or subgenre. Although a single geographical category will typically comprise a large variety of subgenres, the geographical origin is occasionally used to describe a music genre. In this section, we define experimental indicators, experimental settings, and analysis of the experimental result.

#### 4.1. Experimental Indicators

This section carries out a music flow classification experiment on MIDI music files of classical, country, dance, folk, and metal music genres, which is a multiclassification task with relatively balanced categories. In classification tasks, accuracy, accuracy, recall rate, and F1 value are often used to evaluate the performance of classification methods and models. Computations of these indices often require the construction of an obfuscation matrix. The confusion matrix can be used to visualize the classification results and represent the corresponding relationship between predicted classes and actual classes in the verification set or test set. The confusion matrix is a square matrix, which can be written as

\[
\frac{2 \times TP}{2 \times TP + FP + FN}.
\]

The abscissa represents the real class, the ordinate represents the predicted class, and each element \(M(I, j)\) represents the number of instance samples of the real class \(I\) predicted to be class \(J\).

### 4.2. Experimental Settings

This paper uses Python programming language and uses Keras to call Tensorflow background for MIDI music genre classification experiment. In the experiment, 80% of MIDI files of each music genre are selected as training sets, and the remaining 20% are selected as validation sets. Training sets and validation sets have no intersection independently. The number distribution of MIDI files of each category in the training set and verification set is shown in Table 1.

#### 4.3. Analysis of Experimental Results

For example, the experimental results are shown in Table 2.

**Table 1: MIDI music distribution.**

<table>
<thead>
<tr>
<th>Music genre</th>
<th>Classical</th>
<th>Rural</th>
<th>Dance music</th>
<th>Dance music</th>
<th>Dance music</th>
</tr>
</thead>
<tbody>
<tr>
<td>The training set</td>
<td>320</td>
<td>308</td>
<td>268</td>
<td>320</td>
<td>320</td>
</tr>
<tr>
<td>Validation set</td>
<td>80</td>
<td>78</td>
<td>67</td>
<td>80</td>
<td>79</td>
</tr>
</tbody>
</table>

**Table 2: Comparison of experimental results of music classification.**

<table>
<thead>
<tr>
<th>Number</th>
<th>Acc</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7526</td>
<td>0.7229</td>
<td>0.7369</td>
<td>0.7298</td>
</tr>
<tr>
<td>2</td>
<td>0.8646</td>
<td>0.856</td>
<td>0.8556</td>
<td>0.8558</td>
</tr>
<tr>
<td>3</td>
<td>0.8828</td>
<td>0.8797</td>
<td>0.8804</td>
<td>0.8801</td>
</tr>
<tr>
<td>4</td>
<td>0.901</td>
<td>0.8995</td>
<td>0.8991</td>
<td>0.8993</td>
</tr>
<tr>
<td>5</td>
<td>0.8724</td>
<td>0.8643</td>
<td>0.8652</td>
<td>0.8648</td>
</tr>
<tr>
<td>6</td>
<td>0.888</td>
<td>0.8852</td>
<td>0.8855</td>
<td>0.8854</td>
</tr>
</tbody>
</table>

**Table 3: Confusion matrix (%).**

<table>
<thead>
<tr>
<th>Pre</th>
<th>Dance music</th>
<th>Metal</th>
<th>Rural</th>
<th>The classical</th>
<th>Folk</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>Dance music</td>
<td>85.07</td>
<td>5.97</td>
<td>8.96</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Metal</td>
<td>3.8</td>
<td>94.94</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>15.38</td>
<td>0</td>
<td>82.05</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>The classical</td>
<td>2.5</td>
<td>1.25</td>
<td>2.5</td>
<td>92.5</td>
</tr>
<tr>
<td></td>
<td>Folk</td>
<td>1.25</td>
<td>0</td>
<td>0</td>
<td>3.75</td>
</tr>
</tbody>
</table>

**Table 4: Comparison of experimental results of music classification.**

<table>
<thead>
<tr>
<th>Number</th>
<th>Acc</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7526</td>
<td>0.7229</td>
<td>0.7369</td>
<td>0.7298</td>
</tr>
<tr>
<td>2</td>
<td>0.8646</td>
<td>0.856</td>
<td>0.8556</td>
<td>0.8558</td>
</tr>
<tr>
<td>3</td>
<td>0.8828</td>
<td>0.8797</td>
<td>0.8804</td>
<td>0.8801</td>
</tr>
<tr>
<td>4</td>
<td>0.901</td>
<td>0.8995</td>
<td>0.8991</td>
<td>0.8993</td>
</tr>
<tr>
<td>5</td>
<td>0.8724</td>
<td>0.8643</td>
<td>0.8652</td>
<td>0.8648</td>
</tr>
<tr>
<td>6</td>
<td>0.888</td>
<td>0.8852</td>
<td>0.8855</td>
<td>0.8854</td>
</tr>
</tbody>
</table>
network model, and the obtained prediction results are represented by a confusion matrix, as shown in Table 3. According to the analysis of the results in Table 3, metal music, classical music, and folk music have achieved good classification effects, reaching 94.94%, 92.50%, and 95.00% accuracy rates, respectively. Dance music and country music have some misclassification. Country music is often misclassified as dance music because some country music can be used as an accompaniment to country dances in a style similar to dance music, and dance music can also be misclassified as country music. There is a slight mismatch between dance music and metal music, possibly because they are similar in their emphasis on rhythm.

5. Conclusion

Panda digital music resources bring difficulties to management, and the traditional classification based on a manual annotation will consume a lot of manpower and cost that cannot meet the needs of the times. Automatic classification of music has gradually become a hot topic and has important research significance. Aiming at the limitation of feature extraction and shallow structure of classifier in traditional research methods, automatic classification will consume a lot of manpower and cost that cannot meet the needs of the times. Automatic classification based on deep learning is proposed. The main tasks of this paper are as follows:

1. In the process of feature extraction, MIDI music is divided into several sections with similar local playing styles, and the section features are extracted to form the section feature sequence with the section as the analysis unit to better describe the music. The specific process includes the extraction of note feature matrix, the extraction of theme and segment division based on note feature matrix, and the formation of feature sequence based on segment theme research and extraction of effective features.

2. In the classification method, a MIDI music classification method based on deep learning is proposed.

3. In order to verify the feasibility and effectiveness of the above content, the relevant experiment of MIDI music genre classification is carried out through programming.

First, the section division and theme extraction experiment. Then the music genre classification experiment is carried out, including the traditional music classification experiment based on BP neural network and the music classification experiment based on deep learning proposed in this paper. From the comparison of experimental results, it is found that the feature set extracted in this paper is suitable for the MIDI music genre classification task. In this paper, the feature sequences of segments are extracted through segment division. Combined with BI-GRU, the sequential characteristics of music can be described effectively, the context semantics and high-level features of music can be learned, and the classification performance is better than that based on BP neural network. On the basis of bidirectional gated recurrent units (BI-GRU), an attention mechanism was introduced to learn more prominent music features, and 90.1% accuracy was achieved in the validation set, proving the effectiveness of the proposed classification method. The validity of this method is verified by the experimental accuracy of equal length segment classification.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References