Research on Artificial Intelligence Composition based on Recurrent Neural Network

Shuang Xu, Shuguang Fang

Shandong Agricultural University, College of Information and Engineering, China.

Abstract

This paper attempts to use audio as the carrier of audio and audio, and uses music signal processing to view the music track as a sequence of music segments with time series characteristics, and uses LSTM as a training model to generate training. The model not only generates new music sequences, but also smoothly splicing the music pieces into a complete audio. It is a good attempt to compose music on the carrier, based on TensorFlow and the Keras framework. The notes extracted from the MIDI format music files are trained by the neural network to generate predictive MIDI files and converted into MP3 music formats, realizing the use of random notes intelligent composition, and establishing a low-cost, more accurate composition application.

Keywords

TensorFlow; Artificial intelligence composition; Recurrent neural network.

1. Introduction

The connection between music and mathematics is inseparable, and all forms of music can be described by mathematical expressions to achieve the transformation between notes and numbers. With the rapid development of artificial intelligence technology, its application has become more and more extensive, and artificial intelligence to meet the emotional needs has emerged. In this paper, artificial intelligence technology is used to generate the least loss of music creation through a series of extraction and training, saving manpower and satisfying people's pursuit of art.

2. Audio prediction

2.1 Data set processing

MIDI files are an important form of music. Using the characteristics of MIDI files, you can read Notes (notes) and Chord (chords) from MIDI files. In order to improve efficiency, convert Chord to integer processing, and finally extract random files. The notes are returned as a list. Among them, the input shape is set before entering the first layer LSTM, and the Dropout layer is designed to prevent over-fitting and discard some of the neurons. Dense is the fully connected layer. Finally, the SoftMax layer calculates the different pitch percentages to take the highest value as the newly generated tone.

2.2 Feature processing

Music conveys emotional information by affecting people's auditory feelings. Experiments show that human auditory perception changes linearly with changes in pitch. The MFCC reflects the pitch-hearing characteristics of the human ear by transforming the logarithmic relationship between frequency and pitch. Research results of music emotion and scene classification problems based on audio

The MFCC can efficiently recognize the pitch and frequency on the music signal and can be used as a feature of audio classification. Therefore, this paper takes MFCC as a feature of unit music. The common MFCC is 39-dimensional. It consists of 13-dimensional static coefficient, 13-dimensional first-order differential coefficient and 13-dimensional second-order scoring coefficient. The differential coefficient represents the dynamic characteristics of music, and the 13-dimensional static coefficient is composed of 1D energy characteristics. The 12-dimensional coefficient is composed.

3. Neural network training

Use the fit method to train the model, store the weights in the HDF5 file, and save such a file at the end of each round to generate the music with the file with the lowest loss rate. The neural network model is generated by using the best parameters obtained by neural network training. The predicted data is mapped into intones by setting appropriate offsets, generating MIDI files, and then converting them into MP3 files.

The model objective function is set to the tanh function, and the LSMMRNN model music prediction problem can be expressed as a function construction problem of the parameter set θ=(W, U):

$$F(pre(m_i);W,U) = h_i$$

$$h_i = o_i \tanh(c_i)$$

Let V_i denote the music vector V(pre(mi)) of the pre-order information pre(mi) at the i-th time, then:

$$o_i = \varphi(W_o V_i + U_o h_{i-1})$$

Let I_i、F_i denote the input gate and forgetting gate of the LSTM model, respectively, and C_i denote the memory unit of the LSTM, then the memory cell C_i per unit time i in the LSTM passes through the input gate I_i And Forgetting Gate F_i is adjusted to the sum of new content and early memory content:

$$c_i = f_i c_{i-1} - I_i \tilde{c}_i$$

$$\tilde{c}_i = \tanh(W_c V_i + U_c h_{i-1})$$

The input gate and the forgetting gate respectively control the input of new content and the forgetting of old content:

$$I_i = \varphi(W_I V_i + U_I h_{i-1})$$

$$f_i = \varphi(W_f V_i + U_f h_{i-1})$$

After the memory unit is updated, the hidden layer calculates the current hidden layer H_i according to the calculation result obtained by the current input gate, as shown in the above formula. At this point, when W and U are determined, the constructor F is uniquely determined.

4. Experimental results

Figure 1 and Figure 2 show the amplified loudness values when the time values are 1s and 2s, respectively, and the joints in the rectangular frame. It can be seen from Fig. 2 that when the time value is 2s, the range of change is slightly longer, and the feeling of fading and fading can still be clearly felt, so that the two songs are loosely connected tightly, so that the auditory can clearly distinguish not one. For the first piece of music, when the track loudness analysis is performed, the track loudness map of the data portion obtained after processing can be seen that there is a significant long-segment weakening portion at the audio joint, which is quite different from the original music frequency.

At the time of 1 s shown in Figure 1, the sudden convergence of the music interface is weakened and the linear change is not obvious. The smoother part has a better result, the auditory recognition is not obvious, and the smooth part is better. process result.

Conclusion

This paper uses TensorFlow framework to realize artificial intelligence composition application, and combines its upper framework Keras, which simplifies the implementation process and creates some music with random input, which has good effect. However, the composition of the model is long and short, and the results are also uneven. The quality of the composition depends on the quantity and quality of the audio material. After adding interactive calculations, some improvement in the quality of the generation is obtained, but how to get the generally higher quality music and The adaptability aspect of the algorithm needs to be improved.

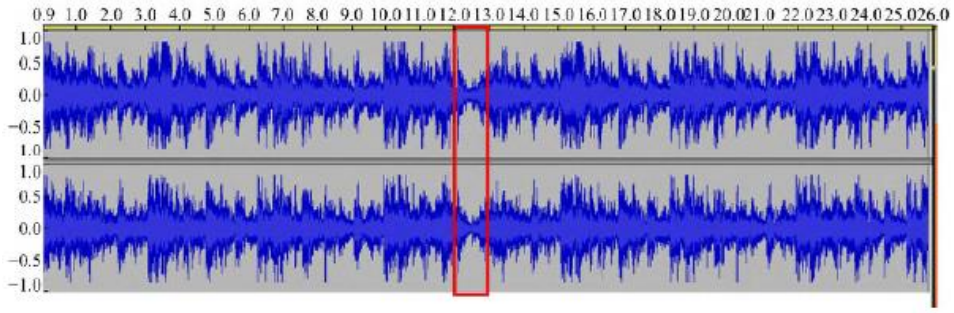

Figure 1 Track loudness diagram when t=1

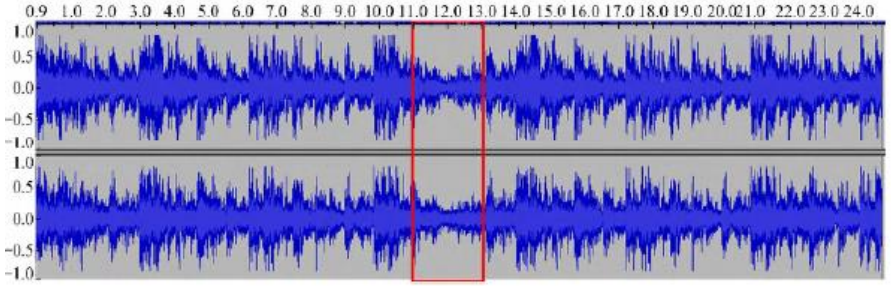

Figure 2 Track loudness diagram when t=2

References

[1] Liu Yuquan. The Third Mode of Composing——On the New Thinking of Computer Music Creation[J]. Chinese Music, 2006(03): 51-54.

[2]Turkalo D M.All music guide to electronica(book Review)[J].Library Journal,2001,126(13):90.

[3]Hiller L A, Isaacson L M.Experimental music/Composition with an electronic computer[M]. NewYork:McGraw,1959.