Scholar Publishing Group
International Journal of Art Innovation and Development
https://doi.org/10.38007/IJAID.2023.040110
ISSN 2789-7192 Vol. 4, Issue 1: 119-127

Multi-voice Music Generation System Based on Recurrent Neural Network

Fei Qiao[1, 2*]

[1]*Shanxi Technology and Business College, Shanxi, China*

[2]*Philippine Christian University, Manila, Philippine*

kawayijiangkou@163.com

[*]*corresponding author*

Keywords: Recurrent Neural Network, Multi-Voice Music Generation System, Artificial Intelligence, Automatic Composition

Abstract: Music is closely related to human life, and it is an important way for people to express their feelings and sing about life. With the rapid progress of artificial intelligence in recent years and its application in various fields, it has also brought great development to computer music, among which algorithmic composition is an important research branch of computer music. This paper aims to study the design of multi-voice music generation system based on recurrent neural network. This paper will take music audio as the research object, and propose a new algorithm for automatically synthesizing music based on recurrent neural network. The framework of automatic music synthesis with audio as the research object mainly includes the analysis of audio files, the audio features of music and the model applied to automatic composition. In the audio file analysis part, the structure of the audio file and the important parameters related to this experiment are introduced in detail, which lays the foundation for the successful experimental operation. In the part of music audio features, it introduces features including mel-frequency cepstral coefficients, linear predictive coding, zero-crossing rate, short-time energy value, etc. Among the models used in automatic composition, the circulating neural network, which is the most active artificial neural network in the field of automatic composition algorithm, and two variants of long-term and short-term memory model and gated circulating unit model are emphatically introduced, which are also the basic models studied in this paper. Secondly, the algorithm of automatic music and audio synthesis based on neural network is described in detail. Firstly, the problem of automatic music and audio synthesis is formally described, and the concepts of unit music, unit music vector, AI-generated music, etc. are put forward, which represents music creation as a processable problem. Then, the process of extracting the audio features of unit music is described in detail. After that, the prediction and synthesis process of music audio are described in detail, and the algorithm description is given. Finally, the audio mosaic synthesis part which directly affects the audience's

intuitive auditory experience is introduced, and the method of weakening and enhancing first is put forward for superposition mosaic, so as to achieve smooth mosaic effect. Finally, a series of experiments are carried out on the algorithm model, including the music and audio automatic synthesis experiment based on LSTM model, the human-computer interaction experiment and the music and audio automatic synthesis experiment based on GRU.

1. Introduction

Among them, the field of computer music has attracted the attention of many scholars and experts in various fields such as computer and signal processing, and even music experts. The purpose of computer research is to collect[1-2]. Record and analyze abstract symbols in music through computer media, and abstract human emotional tendency towards music through the algorithm of generating new songs. Nowadays, the new computer music addresses can be roughly divided[3-4]. In a narrow sense. A wide range of computer music also includes being able to use computers to help composers create music, and being able to use digital signals to store analog sounds instead of original storage methods. In the narrow sense, computer music mainly refers to the use of computers[5-6]. Which enables computers to use algorithms, artificial intelligence, neural networks and other technical solutions for analysis and creation[7-8].

In the design and research of multi-part music generation system based on recurrent neural network, many scholars have studied it, and achieved good results. For example, *Gao F* et al. used Mel cepstrum coefficient as[9]. The feature vector of audio classification, and classified audio into six categories: *Lu C* music, news, sports, advertisements, cartoons and movies, and used support vector machine to classify them through training [10]. The technology used by Lu is to conduct the Fourier series analysis of the audio spectrum to give the possible pitch distribution, and also provides some tools of EQ equalizer to gain or attenuate some frequency bands to assist users in collecting the spectrum work[11].

This paper introduces the research significance, background, research status at home and abroad in this field and the research content of this paper. This paper introduces the framework of automatic music synthesis, including audio file processing, feature extraction and application algorithm model. This paper describes the automatic synthesis algorithm of multi-voice with neural network, and formally expresses the problem of automatic music synthesis. Based on the recurrent neural network model, an automatic music synthesis algorithm is proposed, which automatically handles the problem of splicing music sequences into complete music.

2. Design and Research of Multi-Voice Music Generation System Based on Recurrent Neural Network

2.1. Automatic Synthesis of Music Audio Based on Neural Network

Based on the RNN-RBM model and the design idea of convolutional neural network, this paper designs a dual-axis LSTM network structure. This paper makes a great improvement on the aspect of keeping notes invariable, which is not available in the ordinary LSTM network, so that the notes of different voices in the same time step can transmit information instead of keeping independent of each other, and the training data after modulation can be identified without a lot of redundant

training processes. Two-axis LSTM network only focuses on the relative position of the whole note, not the absolute position, so it has great advantages over other methods of modeling multi-part music. At the same time, the process of composing music can generally be regarded as a process in which theoretical experience and accidental inspiration coexist. For the former, we can get it from the training of the data set by the artificial neural network mentioned above, while the latter needs to consider breakthroughs and innovations from different directions. In this paper, we use chaos theory to simulate this process. Chaos theory is the analysis of irregular and unpredictable phenomena and their processes, but it is obviously different from chaos and random phenomena. Similar to the explanation of "butterfly effect", any slight change in the parameters of chaotic dynamic system will cause infinite changes in the state of the system. Because of this characteristic of chaotic system, we can obtain melodies with different characteristics and complexities by changing system parameters and models. First of all, we experiment with a variety of different continuous and discrete dynamic systems as models, set the evaluation criteria of melody lines, and count and analyze the melody characteristics synthesized by various systems. After that, we form a control system with LSTM network. By adjusting the ratio of melody data set synthesized by chaotic system to preset data set, LSTM network can generate music pieces with more chaotic characteristics. In addition, the control system will evaluate the generated result, obtain its evaluation value, and compare it with the reference value. The deviation will act on the power system as a control signal to correct the system parameters [12]. We have designed the control system as shown in Figure 1.

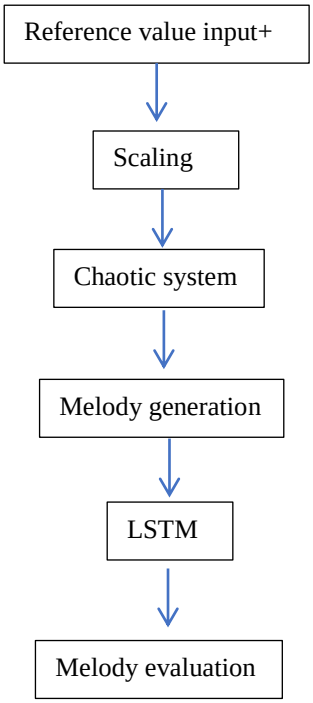

Figure 1. Structure block diagram of music generation system

2.2. Model Classification of Algorithmic Composition

Generally speaking, there is no universal way to classify different composing algorithms, but one of them can be investigated and summarized. One method is to distinguish between the ways in which algorithms are involved in the composing process, including the way in which the music

generated by the computer does not need any manual intervention, and the way in which the computer participates as an auxiliary function, such as the Arpeggiator, which is widely built in synthesizers and audio workstations [13]. In addition, they can be classified according to their system structure and data processing methods, as shown in Figure 2:

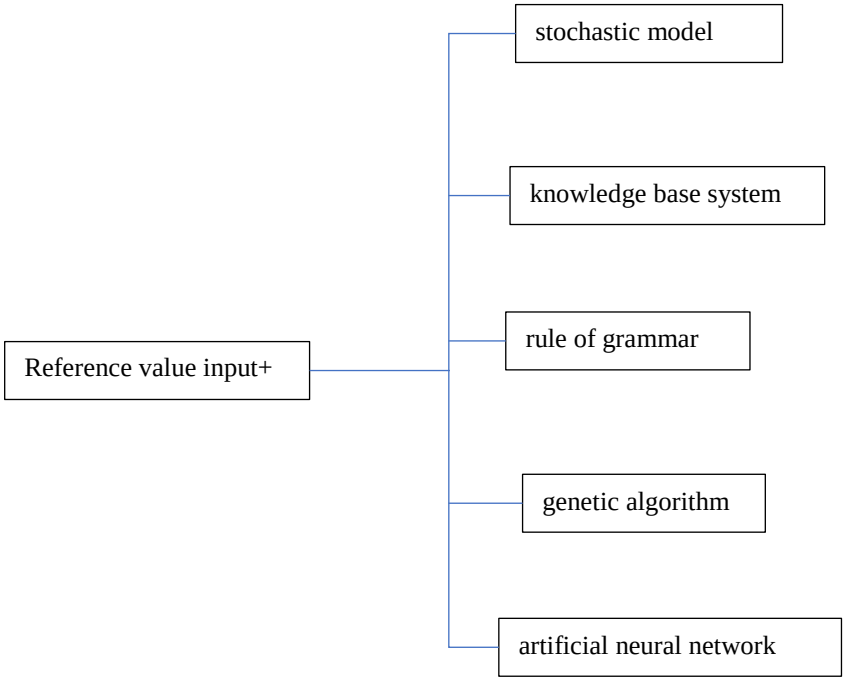

Figure 2. Classification of data processing methods

(1) The system based on mathematical model generally composes music by mathematical modeling based on equations and random events, and the most common way is random process. In the random model, a piece of music is combined by a non-deterministic method, and the combination process is only realized by the composer's choice according to the possibility weight of random events. The outstanding examples of random application of algorithms in composition are Markov chain and Gaussian distribution, and they are usually used together with other algorithms in the decision-making process.

(2) Knowledge base system. Knowledge base system usually contains a large number of music theory rules, and each style of music has its own independent rule base. However, the construction of knowledge rules and the search for exception rules are relatively difficult and complicated, so there are some limitations in application and innovation.

(3) Grammar rules of grammar-based generation system can also be used to generate music. Different from knowledge base system, grammar rules only describe macroscopic aspects of music, such as harmony and rhythm, so it is much easier to construct rules. As for the details of a single note, grammar rules can't be constrained, so it usually needs to be improved with other algorithms.

(4) The system based on evolutionary method almost always uses genetic algorithm to generate music. The way of music generation is the evolutionary process of genes or chromosomes, including recombination, mutation, crossover, inheritance and other ways to recombine small pieces of music. Through the continuous iteration of the algorithm, the disadvantaged individuals will be eliminated continuously, leaving the excellent individuals in the end. The key of the algorithm lies in the evaluation process, and the evaluation method is an important part of the algorithm to control the quality of works.

(5) Learning-based system Learning-based system usually adopts artificial neural network.

Because of the diversity of network structure, it usually has more expansibility than other models, and it is the most promising modeling method among all kinds of algorithmic composition models at present [14].

2.3. Algorithm Selection

The input signals xi, i = 1, 2, 3, ..., n are respectively multiplied by the corresponding weight values wk, where i represents the number of input signals and k represents the number of neurons. The input vk of the whole network is the sum of all input signals. In addition, the offset value (i.e., threshold) is also added to the sum of input signals to make neurons adapt to different specific situations [15].

Vk is calculated as follows:

$$u_k = \sum_{i=1}^{n} w_k x_i \tag{1}$$

$$v_k = u_k + b_k \tag{2}$$

The output of neurons is:

$$y_k = f(v_k) \tag{3}$$

The ideal activation function of is the standard step function, which maps the input value to the output value of 0 or 1, where 1 corresponds to the excited state of neurons and 0 corresponds to the inhibited state of neurons.

3. Design Research and Experiment of Multi-Voice Music Generation System Based on Recurrent Neural Network

3.1. System Design

The algorithm proposed in this paper uses the solution of the nonlinear dynamic system equation of chaotic dynamic system, and defines the quantization forms of scale and rhythm. The components of the solution are mapped to the pitch and rhythm values by using the Ernong mapping, and then the values are normalized and quantized discretely. Finally, a file conforming to MIDI standard is exported. We compare the synthesized melody results with a set of reference melodies, and get the factors that affect the melody characteristics by calculating the gradus suavitatis and changing the parameters of chaotic system, so as to control the complexity of the synthesized melody by changing the system parameters, and the whole dynamic system will keep running until the calculated melody falls within the predetermined range. Finally, the melody synthesized by chaotic dynamical system is used as the simulation of inspiration source in the composing process, and input into LSTM network together with other MIDI file data sets.

3.2. Experimental Design

Aiming at the design experiment of multi-voice music generation system based on recurrent neural network in this paper, firstly, the maximum likelihood estimation values of different neural model algorithms are analyzed to find a model with higher accuracy; secondly, the music generated by the model is compared with artificial music to judge the quality of the music generated by the model of the system constructed in this paper.

4. Design Research and Experimental Analysis of Multi-Voice Music Generation System Based on Recurrent Neural Network

4.1. Estimated Value

In this paper, the above MIDI file data sets are divided into training set, verification set and test set. The tonality of the data sets in Table 1 are all moved to C major and C minor, and the input data are all matrix patterns converted by Mido library. The vector dimension of each time step is 88, indicating the range of notes from A0 to C8 on the piano window. In this paper, the maximum likelihood estimation values of note sequences in different situations are compared among stochastic neural network model, RBM model, RNN-RBM model, LSTM-NADE model and biaxial LSTM model.

Table 1. Maximum likelihood estimates for different models under training data with uniform tone

	Stochastic model	LSTM	RNN-RBM	LSTM-NADE	biax LSTM
MuseData	-61.00	-11.56	-6.02	-5.06	-4.62
Piano-Midi	-61.00	-12.71	-7.10	-7.41	-5.12
Nottingham	-61.00	-6.57	-2.41	-2.13	-1.57

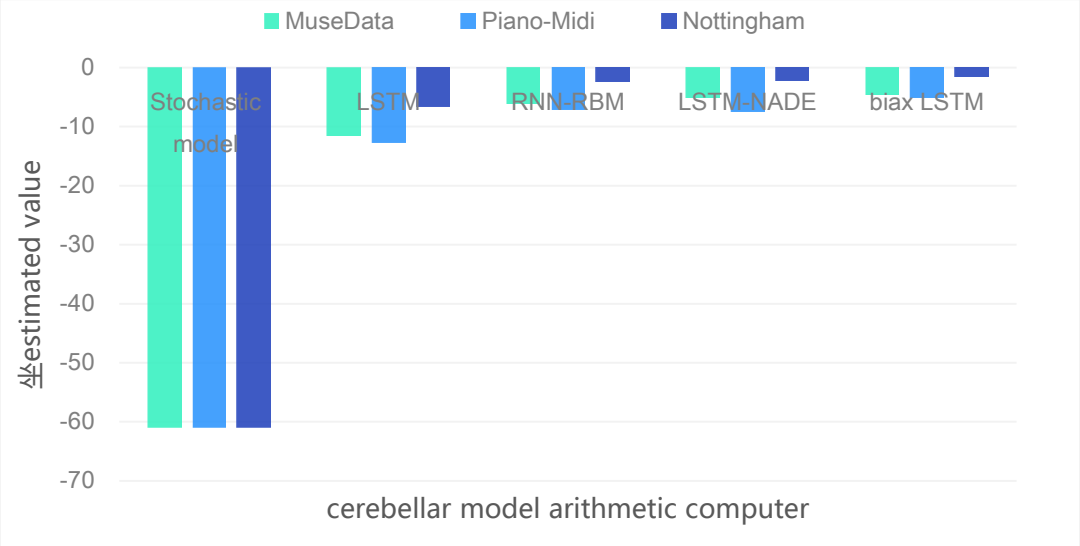

Figure 3. Comparison of the maximum likelihood estimates for the different neural network models

As can be seen from Figure 3, the accuracy of the combined RNN-RBM model is significantly higher than that of the simple LSTM model. At the same time, when the RNN unit is replaced by LSTM unit, the accuracy of the RBM model is increased after the NADE model is replaced. Compared with other models, LSTM network model has obvious advantages, and its accuracy is higher than other models.

4.2. Generate Music Contrast

Based on the comparison of time consumption, this paper conducts a subjective evaluation test on the generated music generated by the two systems. The test is divided into two parts. Firstly, 10

test audios are selected, 5 from human work songs and 5 from GRU composition model generation. In order to effectively control the variables of subjective test, 24 testers were invited to conduct this Turing test evaluation. At the same time, in order to avoid that the music that participated in the test in the original experiment will affect this evaluation because of the existing memory and subjective impression, we randomly selected the music that participated in this test after excluding the music that had already participated in the evaluation.

Table 2. Weweighted scores and ranking results based on GRU model

	one	2	three	four	five
People work song	97.5	95	92	seventy-eight	77.5
Model generation	93.5	eighty-nine	86.5	eighty-five	83.5

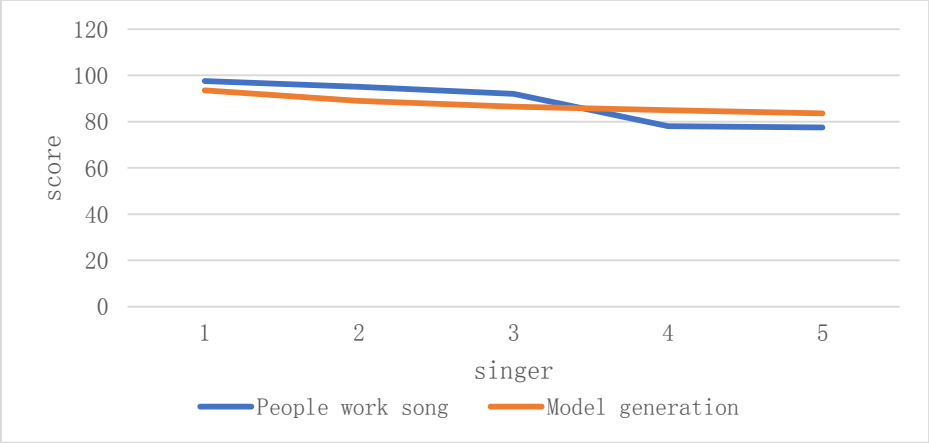

Figure 4. Comparison of human working song and model composition score

From Figure 4 and Table 2, we can clearly see that at present, there is still a certain gap between the generated music and the artificially generated excellent music, but on the whole, the music generated by the model is relatively stable and excellent. Therefore, it is suggested to increase music training data in the later period, and then further optimize the aesthetic taste of human beings.

5. Conclusion

With the development of computer science to artificial intelligence everywhere, artificial intelligence can be said to have penetrated into various fields, and music is an indispensable part of human entertainment life. Using artificial intelligence to compose music can achieve the purpose of enriching people's entertainment life, but how to formalize the process of music creation and achieve automatic synthesis is undoubtedly the key and difficult point in algorithmic composition. From the current development, There may still be a long way to go before computer-generated music can be commercialized or brought into life, but today, with the increasing development of science and academic progress, I believe this subject will get more development and progress. That is to say, this paper makes a discussion and research under this background, using audio files as the direct music research object, trying to understand music in the way of human senses based on human auditory experiments, and putting forward the automatic synthesis calculation of music audio based on neural network and making experiments. The main work of this paper is as follows: analyzing and separating audio parameters and data streams based on music audio files and

extracting music features. The concepts of unit music and music vector are put forward, and the problem of automatic music and audio synthesis is formally described, so that this kind of problem can be clearly expressed.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1]Huo J, Sun W, Dai H. Research on Machine Vision Effect Based on Graph Neural Network Decision. Journal of Physics: Conference Series. (2021) 1952(2): 022-050 (7pp).

[2]Zhengwen Li, Wenju Du, Nini Rao. Research on Classification Method Based on Inaccurate Image Dataset Cleaning. Journal of Signal Processing. (2022) 38(7):1547-1554.

[3]Tobing P L, Wu Y C, Hayashi T, et al. Voice Conversion with CycleRNN-based Spectral Mapping and Finely Tuned WaveNet Vocoder. IEEE Access. (2019) PP F(99):1-1.

[4]Lu Y, Yang Y, Wang L, et al. Application of Deep Learning in the Prediction of Benign and Malignant Thyroid Nodules on Ultrasound Images. IEEE Access. (2020) PP (99):1-1.

[5]Al-Rubaye W, Al-Araji A S, Dhahad H A. An Adaptive Digital Neural Network-Like-PID Control Law Design for Fuel Cell System based on FPGA Technique. University of Baghdad Engineering Journal. (2020) 26(9):24-44.

[6]Ota S, Taki S, Jindai M, et al. Nodding detection system based on head motion and voice rhythm. Journal of Advanced Mechanical Design Systems and Manufacturing. (2021) 15(1):JAMDSM0005-JAMDSM0005.

[7]Luo M, Ke Q, Li J. Research on Automatic Braking and Traction Control of High-speed Train Based on Neural Network. Journal of Physics: Conference Series. (2021) 1952(3):032048-.

[8]Dhinavahi A. Speech Recognition-based Billing System: A multi-model design and implementation. International Journal of Advanced Trends in Computer Science and Engineering. (2020) 9(2):1568-1573.

[9]Gao F, Meng C. Design of marine environmental noise signal generation system based on MATLAB, LabVIEW and FPGA. Journal of Physics: Conference Series. (2019) 1303(1):012068 (7pp).

[10]Lu C, Zhang Y, Zheng Y, et al. Precipitable water vapor fusion of MODIS and ERA5 based on convolutional neural network. GPS Solutions. (2023) 27(1):1-13.

[11]Borzov S M, Karpov A V, Potaturkin O I, et al. Application of Neural Networks for Differential Diagnosis of Pulmonary Pathologies Based on X-Ray Images. Optoelectronics, Instrumentation and Data Processing. (2022) 58(3):257-265.

[12]Al_Araji A, Al-Zangana S J. Design of New Hybrid Neural Controller for Nonlinear CSTR System based on Identification. University of Baghdad Engineering Journal. (2019) 25(4):70-89.

[13]Ahmad, Riaz, Ahmed, et al. Urdu Nasta'liq text recognition system based on multi-dimensional recurrent neural network and statistical features. Neural computing & applications, 2017, 28(2):219-231.

[14] Tobing P L , Wu Y C , Hayashi T , et al. Voice Conversion with CycleRNN-based Spectral Mapping and Finely Tuned WaveNet Vocoder. IEEE Access, 2019, PP(99):1-1.

[15] Seujski M , Suzic S , Pekar D , et al. Speaker/Style-Dependent Neural Network Speech Synthesis Based on Speaker/Style Embedding. Journal of Universal Computer Science, 2020, 26(4):434-453.