

## Music Generation System Based on LSTM

Yuan Xiao<sup>1,a</sup>, Shengwu Xiong<sup>1,2</sup>, Pengfei Duan<sup>1,2</sup>

<sup>1</sup> School of Computer Science & Technology, Wuhan University of Technology, Wuhan 430063, China

<sup>2</sup> Hubei Key Laboratory of Transportation Internet of Things, Wuhan 430070, China

<sup>a</sup>email: 1214118459@qq.com

**Keywords:** LSTM Model; Automatic Music Generation System; Musical Notation

**Abstract.** Traditional music generation is manual, so it's of great significance to generate music automatically by machines. This paper proposes a method of using the Long-Short Term Memory (LSTM) Unit model to format music file, extract characters and generate music. And then constructing an automatic music generation system through the formalization processing and neural network supervised learning. The experimental results show that the system can effectively generate a new music which is similar to the original one.

### Introduction

With the rapid development of science and technology, computer music tools have been widely used and the demand for music has also become more diverse. Mobile phone users like personalized ring tone; the entertainment needs transform multiport music, while theaters and games requires a lot of rhythm and beat. Though traditional manual composition can ensure the quality of music, it reveals many problems in supply and adaptability. While the appearance of the powerful tool of computer has brought much impact on traditional music, and are leading the direction of modern music. In increasingly mature of computer automation technology and artificial intelligence today, some researchers and artists at home and abroad have done an attempt to make computers automatically produce music that is in the sense of beauty, promoting the progress of the music industry.

With the use of the existing music, there are two models widely used for feature extraction: Genetic Algorithm (GA) model [1-3] and probabilistic model [4]. Each of these two models have their own advantages and disadvantages. As for the Genetic Algorithm model, it can emphasize the strong rhythm in each fragment and combine them into distinct pieces of music. But it's with low efficiency, because every iteration process of it has a delay. In addition, due to the lack of context, it's difficult to get the coherence and deep-seated rhythm information. In terms of the probabilistic model, the rhythm is regarded as a disorder of notes, and it ignores the grammar and even the order of notes. This model has been proved to be simple and effective in music classification, but its shortcomings are also very obvious. It also cannot solve the problem of context correlation.

In view of the problems existing in the above models, this paper will use LSTM [5-8] to solve the problem of automatic music generation. LSTM not only can shorten the music generation time, but also can solve the consistency of music context, which illustrates the advanced nature and representative of deep learning in music generation. The main works of this paper are:

- 1) Structured processing music file, automatically generating music with LSTM model in computer.
- 2) Building a high efficiency integrated music generation system combining knowledge of software engineering with LSTM model.
- 3) Using LSTM to solve the problem of context correlation in music generation.
- 4) Evaluating the system performance by building an open survey like Turing test.

This paper introduces LSTM and analyzes its characteristics firstly, then use LSTM neural network model combined with the MIDI [9] (Musical Instrument Digital Interface) digital interface specification to produce music and develop a music generation system. Finally, evaluates the system efficiency. The experimental results show that, automatically generated music requires less

time with this method, for which audibility and continuity meet the basic music aesthetic seem to be a feasible solution.

### LSTM Model

Feed Forward Neural Networks (FNN) [10] has no memory function so that it cannot remember and use history information. For the defects of FNN, Jordan [11] and Elman [12] respectively proposed Recurrent Neural Networks (RNN) in 1986 and 1991, which is a kind of loop feedback network. The core idea of RNN is the network-hidden layers with loop that can use historical information to assist in dealing with current data through such connect. Expansion of traditional RNN is equivalent of a multilayer FNN. The number of historical data corresponds to the network, leading to historical information loss and gradient disappear (explosion) during the parameters training with a large amount of data. Thus historical information that traditional RNN can use is limited in practice [13]. A new solution is needed to solve the problem of RNN and carry forward the advantage of memory function.

Hochreiter and Schmidhuber proposed the LSTM unit to solve the problem of gradient disappear in traditional RNN. After years of development, the most widely used LSTM network is shown in Fig.1.

LSTM unit dedicated to the use of memory cell to store historical information, the update and use of which is controlled by 3 door-input gate, forget gate, output gate. Those three gates use a sigmoid activation function and the input gate used tanh to convert. LSTM cell can be defined with the following formula:

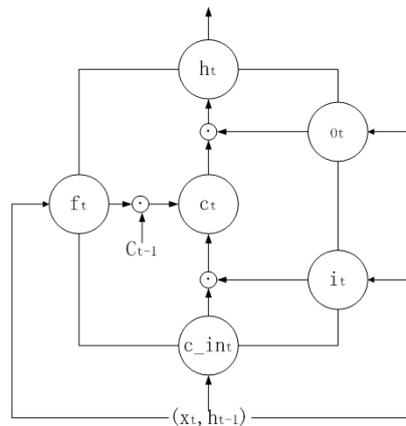


Fig. 1. Structure of LSTM unit

Calculating the value  $i_t$  of input gate, the input gate represents the influence degree of current data input on state of memory cell:

$$i_t = g(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (1)$$

Calculating the value of forget gate, which controls the impact of historical information on the current state:

$$f_t = g(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2)$$

Calculating the value of output gate, which controls the output of state value of control memory state unit:

$$o_t = g(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (3)$$

$c_{in_t}$  transforms the input, the value of candidate memory cell in current moment:

$$c_{in_t} = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_{c_{in}}) \quad (4)$$

The status update of  $c_t$  and  $h_t$  :

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c_{in_t} \tag{5}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{6}$$

Calculation of all gates include the current input data  $x_t$  and unit output value  $h_{t-1}$  in last moment, and connect  $c_t$  and  $h_t$  to prepare for next update. Due to control of gate, LSTM Cell can keep work information for a period of time and without internal interference in training. For example, the memory cell can keep information for a long time and use them through output gate if the forget gate doesn't work and won't be affected by the subsequent input when input gate is closed.

### LSTM Music Generation System

LSTM has been proved to be successfully applied to Natural Language Processing (NLP) [14] in practice, such as speech tagging, sentence legal check and word vector expression. Music as a special carrier for human language, it can bring unexpected results putting into LSTM network for training after special processing.

This paper extracts information from music files with MIDI format, and then puts them into music system framework, with which to analyze music and put into LSTM network for training.

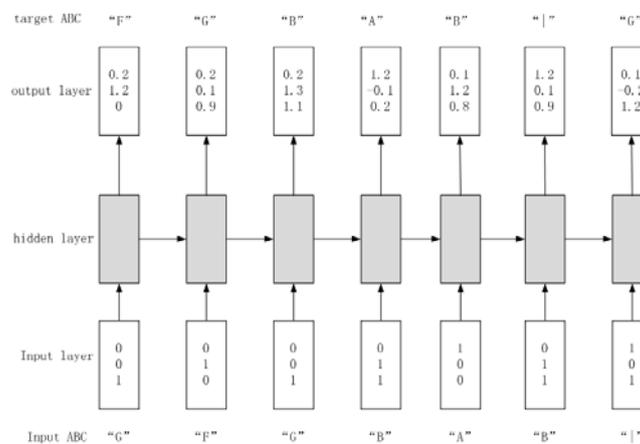


Fig. 2.Cyclic shift output

#### The process of building the system

1. Processes the MIDI file into ASCII character set which can be accepted by music system in accordance with the ABC notation.
2. Converts the character of ABC notation into vector, according to the idea of natural language processing (NLP) in word vector (Word Representation, Word embedding).
3. Train module: accepts vector input, output and basic parameters LSTM needed. Adopting "Back" principle, given a preceding character vector of the sequence, using the next one to present output of current input, the whole process is shown in Fig.2.
4. Test module: receives vector of the test data input module and input to the LSTM module trained by step 3. Then output vector set of prediction results.
5. Converts output vector of the LSTM network test module into the character of ABC notation. It's the reverse process of step 2.
6. Converts ABC notation data into MIDI file, the output file can be played directly by music player.

## Evaluation

To verify the availability of the system, this paper analyzes our works in three aspects: the generation efficiency, frequency spectrum, and artificial evaluation.

LSTM network parameters is shown in Table 1, the training set is 90% of the input vector set and the test set is 10% of the input vector set. For example, input set is {1,2,1,3,4,3,5}, where {1,2,1,3,4,3} is training set and {5} is test set.

Table 1. Network training parameters

Parameter type	Value
Network layer (hidden layer)	2
RNN size	128
Learning rate	$2e^{-3}$
Batch size	50
Maximum training times	200
Proportion of training set	0.90
Proportion of test set	0.10

Time consuming of the whole system is mainly focused on the LSTM network training module. In efficiency, the system improves its production to a great extent so that no complex training is needed for generating music after the training in large amounts of data set.



Fig. 3. Music Score

After 150 times of training in network, Fig.3 is the music score output by test module. It obtains rules of the original music on the whole and magnifies some notable features, which shows the ability of LSTM network to capture context relationship and score in learning rules. What's more, there are subtle changes but not original excerpt of the new generated music.



Fig. 4. The original waveform



Fig. 5. The generated waveform

Compared with the original music(Fig.4), the generated music(Fig.5) is in accordance with the original one in waveform on the whole, and cramped in part, which is consistent with results in gammagram.

Style imitation can be consider to be an artificial-intelligence, and thus a partial Turing test can be used for references [15]. The evaluation is divided into “good”, “general” and “unacceptable” when evaluating the appreciation degree of generated music by human, and we selected 3 pieces of music that are trained by different times for this survey. More than 1000 students are invited to audition and rate the music mounted on website: <http://xiaoyuan.science:8088/survey/>. As shown in

Fig.6 , 88.7% of respondents accept the music generated by our system , and the effect become better and better with increasing the training times, although the effect improves few after the training times increased to a certain level. It's seen that the system is useful and the generated music can satisfy sensory needs of most people even if with shortcomings in details.

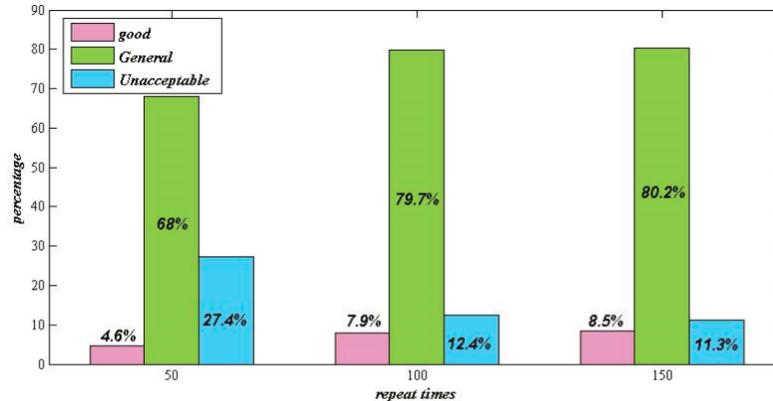


Fig. 6. The survey results

## Conclusions

With the development of computer music technology, people has made lots of efforts in the field of computer automatic music generation. Combined with related knowledge of deep learning, this paper uses music LSTM model to deal with music, and has generated a set of efficient automatic music generation system, which has been achieved good results in the whole, while also show some shortcomings in some features. The conversion of MIDI music to ABC code has lost some melodic information, thus making music compact in part. Therefore a deeper research to extract more information of the tune is needed in the future.

## Acknowledgement

This work was supported in part by Science & Technology Pillar Program of Hubei Province under Grant No.2014BAA146, Nature Science Foundation of Hubei Province under Grant No.2015CFA059, Science and Technology Open Cooperation Program of Henan Province under Grant No.152106000048, the Fundamental Research Funds for the Central Universities under Grant No.(2016-zy-049), Hubei Key Laboratory of Transportation Internet of Things under Grant No.2015III015-B03 and CERNET Innovation Project under Grant No.NGII20151006.

## References

- [1] Holbrook G J. Optimizing Future Perfect: A Model for Composition with Genetic Algorithms[J]. Dissertations & Theses - Gradworks, 2015.
- [2] Jacob B L. Composing With Genetic Algorithms[J]. International Computer Music Association, 2010:452--455.
- [3] Roig C, Tardón L J, Barbancho I, et al. Automatic melody composition based on a probabilistic model of music style and harmonic rules[J]. Knowledge-Based Systems, 2014, 71:419-434.
- [4] Yu Ligong, Bu Jiajun, Chen Chun. Rhythm generation based on inside-outside algorithm[J]. Journal of Zhejiang University(Engineering Science), 2005, 39(12):1969-1972
- [5] Zheng S, Jayasumana S, Romera-Paredes B, et al. Conditional Random Fields as Recurrent Neural Networks[J]. Computer Science, 2015:1529-1537.
- [6] Williams R J, Zipser D. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks[J]. Neural Computation, 1998, 1(2):270-280.

- [7] Sak, Haşim, Senior, Andrew, Beaufays, Françoise. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition[J]. Eprint Arxiv:1402, 2014.
- [8] Gers F. Long Short-Term Memory in Recurrent Neural Networks[J]. *Epl*, 2001, 9(8):1735 - 1780.
- [9] Yang D H. The Application of MIDI Model Music Collection to Digital Piano Group Instruction[J]. *Journal of Xiaogan University*, 2004.
- [10] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[J]. *Journal of Machine Learning Research*, 2010, 9:249-256.
- [11] Jordan M I. Chapter 25 Serial order: A parallel distributed processing approach[J]. *Advances in Psychology*, 1997(97):471-495.
- [12] Elman J L. Finding Structure in Time[J]. *Cognitive Science A Multidisciplinary Journal*, 1990, 14(2):179-211.
- [13] Zhang Liang, Huang Shuguang, Shi Zhaoxiang, Hu Ronggui. CAPTCHA Recognition Method Based on RNN of LSTM[J]. *Pattern Recognition and Artificial Intelligence*, 2011, 24(01):40-47.
- [14] Hu Xuegang, Dong Xuechun, XieFei. Method of Chinese text categorization based on the word vector space model[J]. *Journal of Hefei University of Technology(Natural Science)*, 2007, 30(10):1261-1264.
- [15] Saygin A P, Cicekli I, Akman V. Turing Test: 50 Years Later[J]. *Minds & Machines*, 2000, 10(4):2000.