Iran University of Science and Technology
Deep Learning (Fall 2019)

Music Generation

Moein Mirzaei
Department of Computer Engineering
Iran University of Science
and Technology
mo_mirzaei@comp.iust.ac.ir

Soheil Rastegar
Department of Computer Engineering
Iran University of Science
and Technology
soheil_rastegar@comp.iust.ac.ir

Abstract

An exciting application of the recent advance in AI is Artificial Music Generation. Can we reproduce artists' creativity through AI? Can a Deep Learning model be an inspiration or a productivity tool for musicians?

The first goal of researches in this area is to artificially generate human sounding music or a more specific artist sounding music. In this case-study we want to generate music automatically using Recurrent Neural Network.

In this paper we used ABC notation for music notes and we've encountered several challenges, Then we made problem easier by not considering human-readable notation and just using note sequences as they are presented in MIDI files.

1 Introduction

Music is the art of ordering tunes and sound in right places. A musical note is the smallest part generated by a musical instrument which have a specific frequency. Musical notes have relativity with each other and each note is a factor of other notes. Each twelve note (that we only use seven of them in each part of music) make an octave. Note names is each octave repeated as same as other octaves and their frequency get doubled by every octave, we move higher.

From ancient times people enjoyed listening to music which is generated by human and instruments. As we move forward people believed that art of music generation comes from human souls. But in recent years efforts have been made to make music generation automatic.

In recent years with advances in computer powers and deep learning good results have been achieved in music generation but still far from music generated by human in quality. It's a real problem because the automated system should have the capability to remember the sequence of notes and relations between them. Between all methods that have been proposed to encounter this challenge, the Recurrent-Neural-Networks (RNN) have achieved best performance by far. Special kind of RNN block which is called Long-Short-Term-Memory (LSTM) have been in peak of attention because of it resistant to gradient vanishing problem [2].

The input to a LSTM layer is a sequence of data. In the field of music, we want to make this sequence as clear as possible to have more chance of achieving higher accuracy that is why we can not feed it with raw wave data or sound signal. Input data must get converted to a musical notation which can be fed to LSTM layer. Other methods used ABC notation for describing a piece of music [3] although it gets higher benchmarks than simple raw data, but it still has some complexities for make it readable to humans which is not necessary in our case of neural network training. That's why we used improved version of music notation which uses MIDI format [4] as input.

One of the other challenges that we've encountered is training on Iranian music notes. The problem is that Iranian music notes is somehow different from MIDI notation. In which it has a lot of extra notes comparing to western music notes which is not supported by MIDI notation. Another problem is that there is not acceptable data for Iranian music notes. Except raw signal format that have to get converted to notes which requires lots of signal processing.

In this paper we have proposed several methods for solving these challenges. In first part we've reviewed related work and background of this subject. Next, we proposed our method and evaluate our suggestions in compare to other methods. At the end, we review the result and discuss about other ways that can be implemented to improve results and some other future works.

2 Related work/Background

Music generation have been studied in detail so far. They are several approaches to make automatic music generator which can be divided into two categories, first method uses traditional algorithms to generate music and second method use autonomous systems which learn from previous data to generate music.

For the first method we can refer to Drewes et al. [5] which algebra for computing possible note sequences and generating music for the second method we can refer a variety of proposed approaches after the break through of AI and introducing deep learning [6].

3 Proposed method

A. Data Gathering

We gather our ABC notation data from databases available for this notation which includes lots of music transformed in this format [7]. Also, our midi dataset is downloaded from an internet database of several classical composers [8]. For Iranian instrument data, we've used two large "se tar" solos from Iranian famous musician Jalal Zolfonon. We download it in mp3 format from links available [9].

B. Data Preprocessing

The ABC notation data is self-contained and so doesn't need more preprocessing. We put it in our model without any changes.

For MIDI files, we read it and toss away unusable information. A MIDI file contain several channels which play's as a whole, note description such as time to activate and time to deactivate, the instrument to play the note (which unfortunately doesn't contain Iranian musical instruments) the loudness of each note, and many other information. We only extracted the main channel notes without their time or loudness or other characteristics, just keeping frequencies and their order of playing. This only keeps note sequences and thus have enough information for model to learn notes sequence preference and arpeggios, it may reduce our quality of music by a amount, but makes the task of learning much easier for the model.

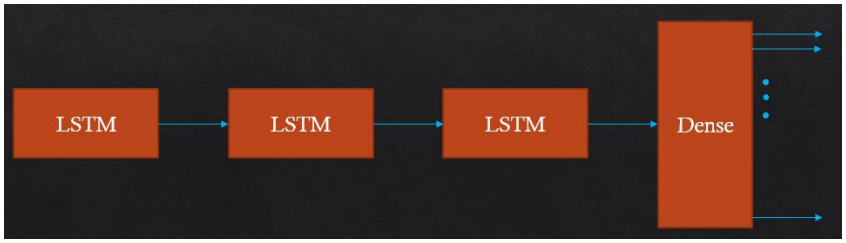
For Iranian music in mp3 format, we cannot expect a high result and good quality of generated music. As in this part we have to face two problems of signal processing (to converting raw sound signal to notes) and learning note sequences. Which reduce the possibility of getting a good result by a large factor. Anyway we used some signal processing techniques by calculating signal fourier transform in small time steps and predict which note is played in this step. The rest part is same as before. We feed the data to our LSTM model. But after getting the result again we have the

challenge of converting these notes to sound as simple note to wav converters doesn't support Iranian notes.

C. Define and train the model

We define a model which uses three stacked LSTM layers and passes the result of last layer to Time Distributed Dense layer, which returns probabilities of characters then it selects the character with the most probability.

We split the data to batches, and the batches contain data from the whole data. After that we train and test the model on these batches. We used categorical cross entropy as loss function since the model tries to predict the next character based on a sequence of input characters.



D. Generate a sample

After training the model, we've achieved the optimal weights. Then we use them to take a sample, by feeding a character from dataset to the model and get the next character, then we feed back this newly generated character to model again and so on.

4 Results

After the training the model with several epochs, we see these results. For ABC notation it takes 3 hours to learn on google Colab environment with train accuracy of 89 percent. But with MIDI notation it takes only 20 minutes in the same environment to reach 94 percent of accuracy. The quality of music cannot get measured by an algorithm, but it was yet a little hard to distinguish between real and generated music.

For Iranian music we don't compare accuracy as a metric with western music, because as described above, it takes more challenges to generate that western music doesn't have.

5 Discussion

This works showed us that it is possible to achieve a decent quality music generation. Although it is not comparable with human musician experts work, but it can be a help tool for music generation industry, to not completely generate music, but to edit human generated music and made suggestions for improvement. or give simple little ideas to begin work with by humans.

Future works on our explanations and methods can be building a powerful raw signal wav to midi generator using signal processing techniques that also support Iranian musical notes. Then this system can be concatenate with music generator LSTM model to generate better quality Iranian music.

More studies on transfer learning for LSTMs specifically for music generation can be very useful as for different genres, learning a model from begging on the language that all music genres have in common (which is use of notes) seems very inefficient.

Also using of GRU blocks instead of LSTMs can be a good way to test their pros and cons in the subject of this field.

References

[1] Mangal, Sanidhya, Rahul Modak, and Poorva Joshi. “LSTM Based Music Generation System.” IARJSET 6.5 (2019): 47–54. Crossref. Web.

[2] Hochreiter, Sepp. 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 6(02):107–116.

[3] https://en.wikipedia.org/wiki/ABC_notation

[4] https://www.instructables.com/id/What-is-MIDI/

[5] F. a. H. J. Drewes, "An algebra for tree-based music generation," Springer, pp. 172-188, 2007.

[6] N. a. B. Y. a. V. P. Boulanger-Lewandowski, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," arXiv preprint arXiv:1206.6392, 2012

[7] http://abc.sourceforge.net/

[8] http://www.piano-midi.de/

[9] http://saziha.ir

[10] https://machinelearningmastery.com/return-sequences-and-return-states-for-lstms-in-keras/

[11] https://machinelearningmastery.com/stacked-long-short-term-memory-networks/

[12] https://machinelearningmastery.com/timedistributed-layer-for-long-short-term-memory-networks-in-python/

[13] https://towardsdatascience.com/how-to-generate-music-using-a-lstm-neural-network-in-keras-68786834d4c5

[14] https://medium.com/artists-and-machine-intelligence/neural-nets-for-generating-music-f46dffac21c0