УДК 004.8

MUSIC NOTES SEQUENCES GENERATION BASED ON MACHINE LERNING APPROACH

Slonskiy O.V., Sorokin M.V.
ICS professor phD. Arsirii O.A.
Odessa national polytechnic polytechnic university, UKRAINE

ABSTRACT. In this work, music notes sequences generation method with usage of combined recurrent and feed-forward neural networks is suggested. Conducted research presents potential of such models in simple melody-like sound sequences generation.

Introduction. For a long time, music composition among other art activities was exclusively human trait. During the technology progress, there were many attempts to mimic this skill. Scientists have been trying to achieve this goal using various deterministic and stochastic methods. In this paper, we would like to present our attempt to generate melody sequences using such machine learning technique as neural networks.

Goal. The main goal of this work was development of approach for music notation processing and structure of neural network that will be capable of generation of structure, which can be processed in melody-like sound sequence.

Main part. Main problem in working with music sequences is fact, that music sequences consist of time stamps, which are scattered through undetermined period of time. This forces to use approaches of time-series analysis. Another problem appears from fact, that current international music notation fit extremely poor with existing data analysis approaches. To overcome these problems developed system implements two modules.

First module solves problem of poor music notation. Current music notation model has three dimension: pitch (vertical position on musical staff), time stamp (horizontal position on musical staff) and note duration (symbol, representing the note). As you can see, there are two parameters (time stamp and note duration), that characterize note in time space. This causes unneeded inconsistences. Therefore, the decision was to reduce given space to two dimension, where one of the dimension represents both timestamp and note duration. This allowed to create new music notation, where on of the axis still represents pitch and other axis represents if the note sounds in a current time stamp (as time stamp was taken 1/128 part of tact. A presentation of transformed format looks like a sequencer (music tool) view. It is aggregated on Figure 1.

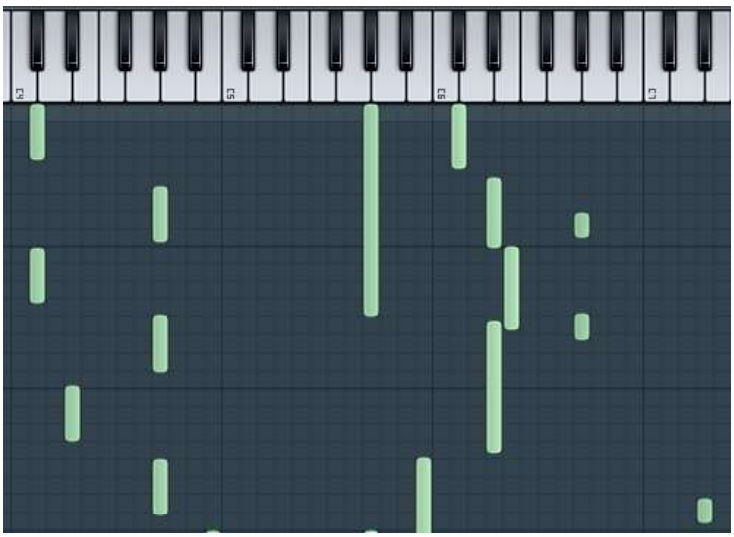

Figure 1: Presentation of formatted music notation (vertical axis – time, step – 1/128 part of tact)

The described module takes a data of music composition in standard format and transforms it to a table with columns, corresponding to every possible music notes (e.g. C0, F4) and rows, corresponding to 1/128 part of musical tact. If in certain part presents some note, it is marked as True, if not, then as False.

Second module is responsible for training and generation of sound sequences. Main part of this model is LSTM neural network. Common approach in time-series analysis using neural network is RNN (recurrent neural network) – class of artificial neural network where connections between units form a directed graph along a sequence. Because basic RNN suffers from “long term memory loss”. In presented approach was used enhanced RNN called LSTM (long-short term memory) [1]. In LSTM is implemented inner mechanism that prevents long-term memory loss, which is extremely useful when for processing long time series.

In general, one time stamp consists of few active notes; this raises a problem of sparse feature space. Regular time stamp consists of 96 places for active note and only from 0 to 6 of them will be active. To overcome this problem, it was decided to use autoencoder for feature space reduction [2]. Autoencoder structure is presented on Figure 2.

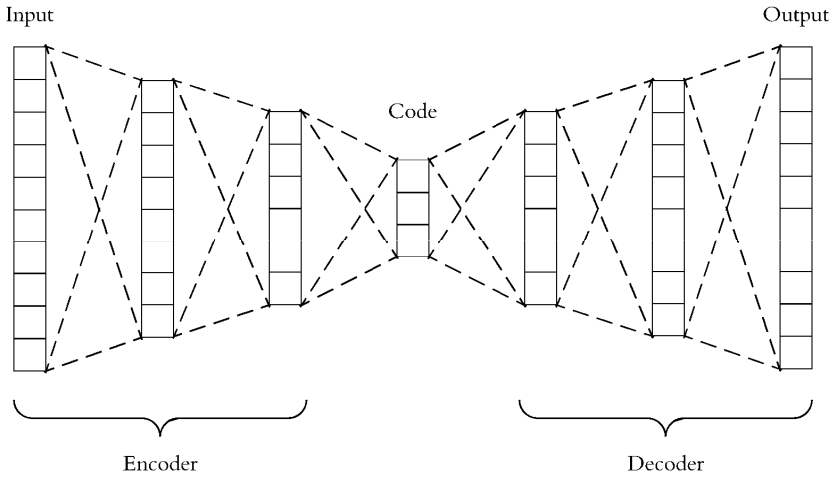

Figure 2: Detailed visualization of an autoencoder

Reduced space (code in Image 2) was used as input parameters to LSTM. To achieve such combination of neural networks, autoencoder was trained separately and cut in the middle. In general, training process consisted of following steps:

2. Format notes to special music notation.
3. Using random batches of timestamps to train autoencoder.
4. Cut Autoencoder in half, concatenate to input, and output layers of LSTM.
5. Train LSTM on music sequences.

As result, we got a system, which is capable to generate a simple note sequences by inputting some timestamp of given model, and predicting next output step-by-step. This note sequences can be played by musicians or synthesized with special tools. In this way, we can make an original note sequence, which is based on some set of music compositions, and form a played music record of it.

REFERENCES

1. Bourlard, H.; Kamp, Y. "Auto-association by multilayer perceptrons and singular value decomposition". / Bourlard, H.; Kamp, Y – Biological Cybernetics, 1988 – p. 291–294.
2. Sepp Hochreiter. «Long short-term memory». / Sepp Hochreiter - Neural Computation, 1997 C.1735–1780.