Music Box : AI Generated Music

Somnath Purkayastha (psomnath@Stanford.edu)
CS230: Deep Learning, Fall 2018

Introduction

Music Box is an LSTM model, capable of generating original Piano score.

It was trained with total 50K music notes over 220 hours.

It was inspired by Magenta (from Google Brain team)

Input Data

2500 piano music (MIDI) files were collected for training

Musical notations (list of notes and chords) were extracted from these files

Each unique note/chord was coded with unique integer number

The list was then normalized (for calculation efficiency) and vectorized (to feed into the LSTM model).

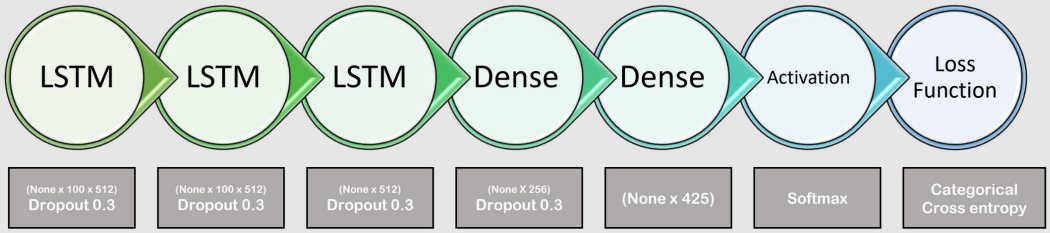
References

[1] Sigurður Skúli. How to Generate Music using a LSTM Neural Network in Keras. https://bit.ly/2IZtgm0

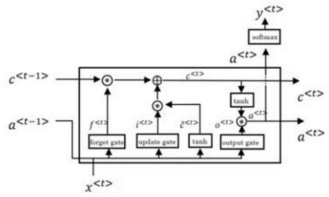
[2] Sample music generated by the model, described in this paper - https://bit.ly/2rzxklN

[3] Google Video Link for Poster Session – https://youtu.be/m8gsXd9iX2o

Model Architecture



LSTM Unit

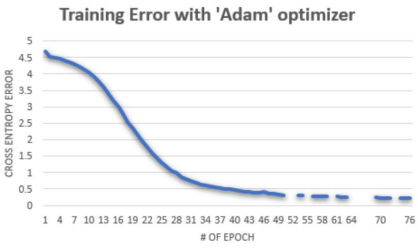


$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$
$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u)$$
$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f)$$
$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o)$$
$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$
$$a^{<t>} = \Gamma_o * \tanh c^{<t>}$$

Result

Values of hyper-parameters (like number of input notes, drop out rate, optimizers) were selected via random search and follow up trials.

With 'adam' optimizer, the error rate went down below 0.1% after 76 epochs.



Conclusion

A relatively simple LSTM model, with moderate input data and processing power (only CPU was used), gave impressive output.

In future we will plan to use more computation power (GPU), longer duration of training and more complex model to produce output with multiple instruments.

We will also try to capture the velocity information in the input data, for better output quality.