Generating polyphonic music using neural networks

Marek Židek | Supervisor: Mgr. Jan Hajič jr.
Faculty of Mathematics and Physics, Charles University in Prague, 2017

Introduction

With recent deep learning advances, music generation is getting more and more attention. Recurrent neural networks are most commonly used for this task and related works show respectable progress.

This experimental thesis explores new ways of generating unique polyphonic music sequences and features a robust evaluation method. We work with discretely represented classical music for piano as training data.

The main problems with related models that we try to address in this work are joint distribution of parallel notes, mistakes (i.e. out-of-tune notes), note axis invariance and overall coherence. Our final model *skip-longrange lstm* shows to be a solid improvement over baseline model in both automated metrics and survey evaluation.

We used a survey evaluation design unseen in related works. It was designed to reveal not only actual results, but also expectations and preconceptions about AI music.

This thesis also provides numerous discussions on subjects not mentioned in the related work (e.g. automated evaluation, over-fitting) and also a good and highly recent overview of recent related research with over 60 citations.

Improvements

* added residual skip connections through time

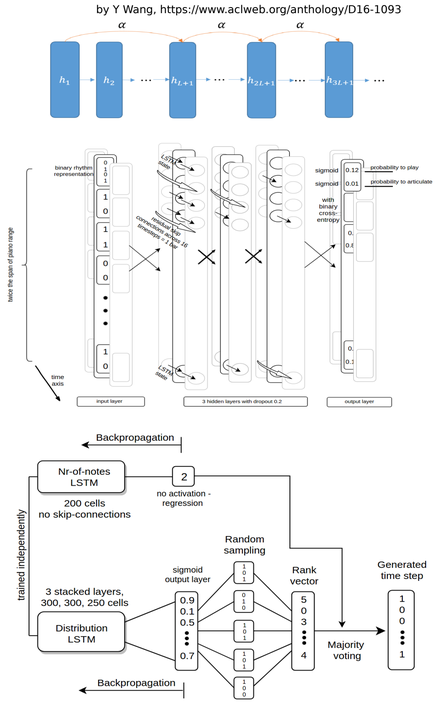
* experimented with the frequency, length and strength of the skip-connections

* added periodic temperature to skip connections

$$\alpha_t = \frac{1}{t \mod L}$$

$$h_{t+1} = tanh(c_{t+1}) \otimes o_{t+1} + \alpha_t h_t$$

* added generating with majority voting for less mistakes and better coherence



Evaluation

* automated (cross-entropy/perplexity), survey and authors impression

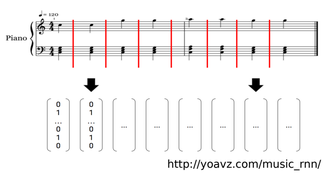
* 9 likert scale questions on 30 sec. music samples: e.g. is it computer generated, consistent, euphonic, boring, rhythmic, overally enjoyable?

* free text spaces for subjective opinions

* 4 different surveys
  * only generated
  * only real
  * mixed with turing test question (TT)
  * same mixed without TT and with a misleading introductory text

Basic method

* represent the notes in a chosen discrete time scale as a sequence of vectors.



* using an advanced RNN, predict a next time step based on the previous time step and the previous context (memory)

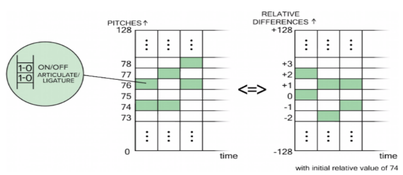
* instead of predicting a whole note configuration (2^88 softmax), predict each note indepenently (deal with the lost dependencies later)

* incorporate the most important additional information (e.g. ligatures, a beat, a velocity...)

Outlined ideas

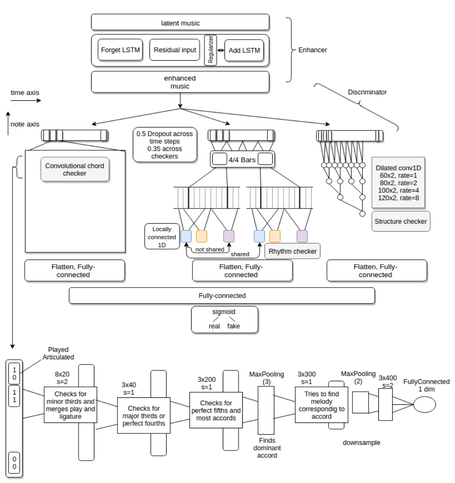
These are the experiments that were being explored, but not properly finished and will be in our future interest.

* relative interval representation training



* enhancing GAN architecture

  * generator has latent music on input instead of a random uniform number
  * experiments with WGAN for music generation



Results

Results were evaluated against a baseline LSTM model

* survey results show that longrange proposed model overally outperforms vanilla lstm, which is supported by our test set automated evaluation on Nottingham dataset and by our own impression.

* 47% of the responses marked longrange samples as real, which is a solid increase over 39% for vanilla lstm

* generated music is more close to modern avantgarde music

* comparing longrange and real modern samples, modern was considered more real and impressive, however, longrange music was more melodic, fluent, variable and less boring.

* variance characteristic of samples seems to be a major difference between real and generated music, however, respondents didn't use it for decision making i.e. it has the lowest and close to zero correlation to TT and overall impression

* people seem to be nicer on computers, meaning that once they guess a generated origin, they give higher rankings than if they don't know about a possible AI origin. (For a computer? Pretty good.)