Jack Pertschuk

Teaching Computer to Listen

There are few things as human as the ability to listen to and appreciate music. Differentiating songs which we perceive as having “merit” from sonic *garbage*. Implicit in the difficulty of this differentiation is the subjectivity of enjoying music; there is no state-independent neurological function for appreciating music, and to suggest that there is would be heretical.

Yet I propose that it’s possible to teach a computer to not only listen to, but appreciate music.

**Objective 1: Isolate certain instruments by causing them to activate specific output-layer neurons.**

I’ve settled on using a Python implementation of the network, using Theano for GPU optimized matrix operations.

**Applications:**

Improved mastering algorithms, improved agnostic recommendation algorithms, etc.

**Goal: Create interactive web tool for analyzing music using deep learning and visualizing the analysis in real time. Implement using python microservice, javascript browser visualization.**

Friday July 8th working on acquiring and building input sample set. Investigating script to pull 10k+ audio files with tags from hype machine. Also possibly contact hypem for access.

To implement with

**Plan:**

**Saturday 9 – Get hands dirty with RNN(LSTM impl) start getting data set**

**Sunday 10 – Finish acquiring dataset, hopefully start testing by end of day**

**Monday 11 – Finish Tweak RNN and start training**

**Tuesday 12 – While its training work on browser GUI**

**Wednesday 13 – Fix, retrain, optimize LSTM**

**Thursday 14 – Work on GUI**

**Friday 15 – Go home, work in car as much as possible**

**Saturday 16 – Register for apple event**

**Sources:**

byeon-PhD-thesis.pdf

**Sunday 10th:**

This morning I finished a working signal processor to create a 2d array for network input from an audio file. The program produces a detailed spectrogram, that clearly depicts audio features, although lacks polish to the human eye. Throughout the signal analysis process, I’ve prioritized keeping the most information and minimizing repetitive information (e.g. but not using overlapping frames for the Fourier transform). I ended up not using the imaginary results of the Fourier transform.