Week 6 Project Update
DS 340

Kyle Salitrik
Tomoki Takasawa

February 15, 2018

1 Dataset

The dataset we obtained for use in this project is the Free Music Archive (FMA) dataset gathered by University of California, Irvine. The dataset contains nearly 1TB of music data consisting of over 100,000 tracks from 161 genres. It was downloaded from this GitHub Repo.

1.1 Dataset Subselection

Utilizing the entire 1TB of music data is both impractical — as training time would be extraordinarily long — and impossible for the amount of storage that is available. Because of this, we used a Python script to perform the following operations:

- Extract the following information for each track:
  - File Name
  - Artist Name
  - Album Title
  - Track Title
  - Genre(s)
- Using a dictionary, count the number of songs in each genre
- Randomly select 20 songs from each of the 155 genres with over 20 songs and create a list of file names.

For the 3100 songs chosen, a bash script was used to copy these songs into a new directory.

1.2 Features Extracted

Using Python, the PyDub library was used to chop the song up into 5 second increments. Then the FFT was computed and LibROSA was used to extract the Key and Tempo of each song clip.

2 Methods to Use

2.1 Neural Net Framework

The Neural Net framework that will be employed is:

- Input Layer for each audio clip consists of:
  - FFT and/or Raw track data
  - Key
  - Tempo
- One or more LSTM layers ending with a Many-to-One LSTM layer
- A single output node that provides the probability a user will like a song

2.1.1 LSTM

Traditional & Convolutional Neural Networks work well with recognizing time-independent information such as images. However, our project requires the ability to recognize the flow or pattern over time. Therefore, we decided to move forward with a Recurrent Neural Network, specifically Long Short Term Memory (LSTM).

As opposed to traditional Neural Networks, a LSTM is able to use previous state as an input of the current state, current state as an input of next state, and so on. As a result, it is more suited to recognize and/or analyze music, which consists of series of states.

2.2 Optional Data Preprocessing Networks

Unlike the recognition / analysis of music, analyzing which data components of an image influence the result is not highly time dependent. Therefore, there is no need to use LSTM. Instead, we decided to use Convolutional Neural Network (CNN) to reduce the dimension of the input vector. This CNN determines association between each dimension and output by looking at desired sized frame in waveform.