STAT578 - STATISTICAL LEARNING IN DATA SCIENCE

MUSIC INFORMATION RETRIEVAL
Project Proposal

March 16, 2018

Zhenye Na(zna2)
University of Illinois at Urbana-Champaign
Department of Industrial and Enterprise Systems Engineering

INTRODUCTION

The goal of this project is to build an automated music generator program that can create songs. The ability to produce songs is highly desirable by listeners and is a feature of live parties, radio shows, and on-line recordings.
A high level description of the project is as follows: I will use a collection of songs (both melody and lyrics), try how well each song matches with each other and analyze the emotional change in the lyrics. Then the neural nets algorithm will automatically create a generated music after training.
These are two main parts of the final project.

1. **Semantic Analysis of Song Lyrics**
   I will explore the use of song lyrics for automatic indexing of music. Using lyrics we apply a standard text processing technique to characterize their semantic content. Then determine artist similarity in this space. We found lyrics can be used to discover natural genre clusters.

2. **Music Audio Analysis and Generation**
   I will perform Music Audio Analysis in Python and will use some neural nets architecture to generate 'its' own music.

DATASET

1. **MSD dataset**
   The Million Song Dataset is a freely-available collection of audio features and metadata for a million contemporary popular music tracks.

   Its purposes are:

   - To encourage research on algorithms that scale to commercial sizes
   - To provide a reference dataset for evaluating research
   - As a shortcut alternative to creating a large dataset with APIs (e.g. The Echo Nest's)
   - To help new researchers get started in the MIR field

2. **Web Crawler**

3. **55000+ Song Lyrics**
   The dataset contains 4 columns:

- Artist
- Song Name
- Link to a web-page with the song (for reference). This is to be concatenated with http://www.lyricsfreak.com to form a real URL.
- Lyrics of the song, unmodified.