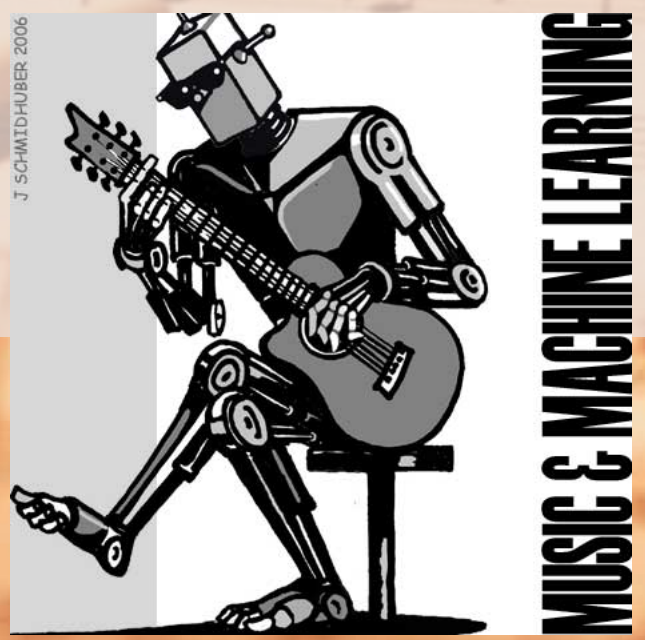
Music Genre and Feature Classification Using ML and NLP Techniques via Visual, Audio, and Linguistic Analysis

Gleb Alexeev

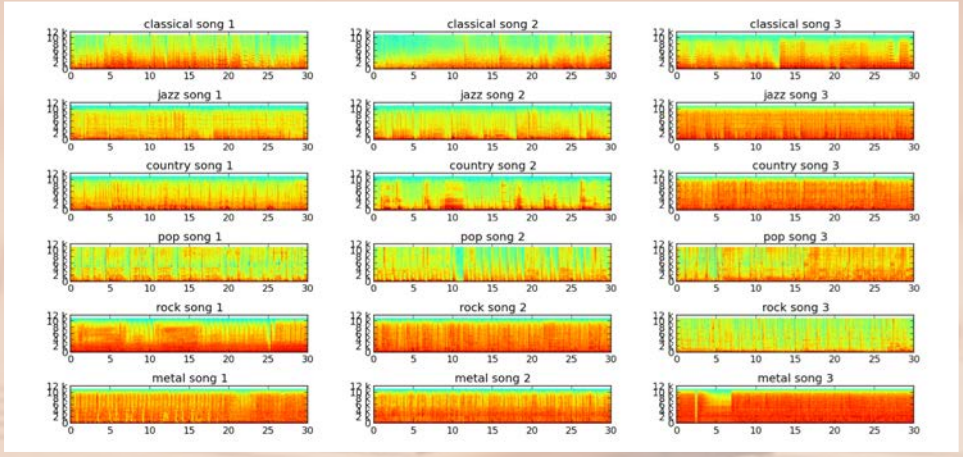


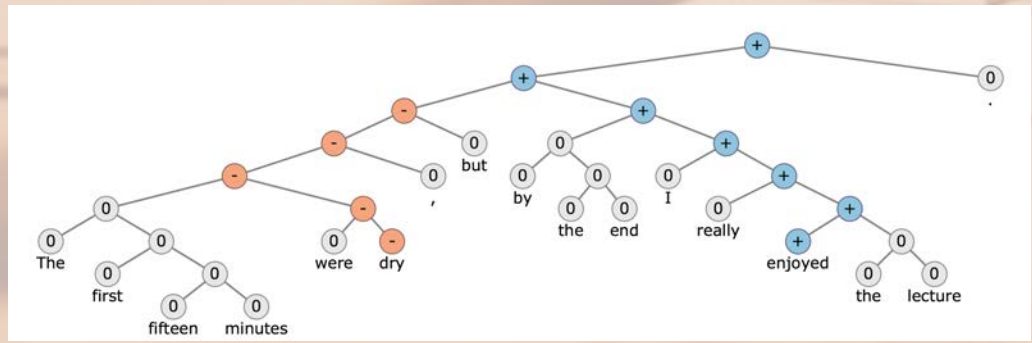
Introduction

- Audio is a very big part of NLP, and something that's a growing industry.
- Decided to focus on music as lots of interesting work there, but also can be applied to Speech recognition, beat recognition, emphasis, etc.
- Previous Work
  - https://github.com/jazdev/genreXpose
    - "proof of concept"
  - http://cs229.stanford.edu/proj2011/HaggbladeHongKao-MusicGenreClassification.pdf
    - Neural Networks and other algorithms
  - https://arxiv.org/pdf/1612.01840.pdf
    - FMA
  - A lot of previous work has been done on music classification, but not many NN approaches

Background

- Dataset
- Audio
  - Fourier Transforms
  - MFCC
- Image Classification
- Linguistics Approach
  - Features





Dataset

- FMA
  - 8,000 tracks of 30s, 8 balanced genres (GTZAN-like)
    - per track metadata such as ID, title, artist, genres, tags
    - common features extracted with librosa.
- Script to pull lyrics for each song snippet (classifying the whole song)

Audio

- Fourier Transform
  - Transform from time domain to frequency domain
  - Spectral analysis
- Mel-frequency cepstral coefficients (MFCC)
  - Encodes the power spectrum of a sound.
  - Calculated as the Fourier transform of the logarithm of the signal's spectrum.
  - Lower range focus, great for audio analysis in speech recognition
- Features to be extracted via scikit.learn, scipy, librosa, opensmile:)
- Use logistic regression for starters to classify between the 8 states, then use RNN
  - http://www.cs.colorado.edu/~mozer/Research/Selected%20Publications/reprints/Mozer1994.pdf
  - https://www-cs.stanford.edu/people/anusha/static/deepplaylist.pdf
- Beat Classification

Image Classification

- Write script to download google image files for each of the 8000 songs (either album or song, using name).
  - Less interesting
- Use Spectogram images and pass into a CRNN.
  - http://deepsound.io/music_genre_recognition.html
- Would be interesting to see difference between FFT and MFCC

Linguistics Approach

- Download lyrics of each song (if existent)
  - word2vec model
  - Classify via CRNN
- Features
  - Year
  - Artist
  - Title
  - Lyrics
  - Apply NLP to words

Evaluation

- Compare to state of the art music classification via deep ML
- Difference in that it is not a single classification of the entirety of the file, but rather a continuous output containing the network's belief of the genre in every point of time.
  - Expected higher error
- classifications:
  - Beat classification (continuous, then split as either "fast/slow")
  - Genre classification (discrete)
  - Potentially Year classification (discrete)
- Template based Natural Language generation

Questions? Recommendations?