



Assessing Music Similarity Through Deep Learning

Nicholas Allen
naallen@stanford.edu

Stanford University, CS230: Deep Learning

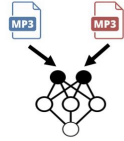
Introduction

Motivation for Deep Learning-Based Music Recommendation

- Music recommenders are systems that aim to recommend music to users based off existing preference
- Often use collaborative filtering techniques
- Types of data include reviews, listening habits, and music characteristics such as genre and tags
- However, these systems require existing usage data, or artist metadata

Goal: Determine artist similarity solely using audio data

- **Input data:** Audio track, or set of audio tracks, from two artists (more accurately, mel-spectrograms)
- **Output data:** Similarity score, from 0 to 1



Dataset

Similarity Metric

- Similarity metric was constructed using Last.FM listener data
- 17 million records relating user IDs to artist IDs, with number of plays
- Metric constructed by counting co-occurrences between all pairs of artists and normalizing

Audio Data

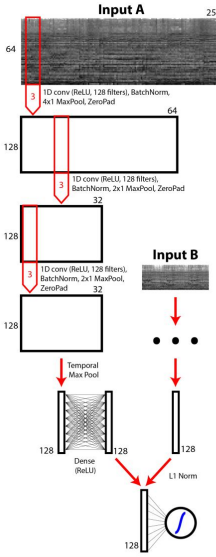
- Artists were matched to artist profiles on Spotify, with iTunes as a fallback
- 30 second track samples of top 3 most popular tracks scraped
- Converted to mel-spectrograms with librosa, size 256x64 (256 timesteps, 64 frequency bins)



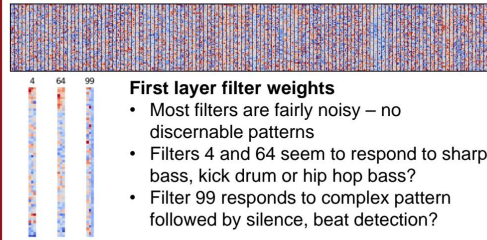
Methods

Neural Network Design

- Two inputs, one output – Siamese networks well suited
- Model performs several 1D convolutions, with interspersed BatchNorm, MaxPool, ZeroPad layers
- Temporal max pool along time axis – removes time-dependence of features
- Dense layer, then L1 norm of the difference between the two outputs from each siamese network
- Final neuron calculates output
- MSE loss used
- Input uses pairwise combinations of audio samples for two artists



Model Visualization



Results

Accuracy Metric

- MSE is good as a loss function, but more relevant accuracy metric needed
- Metric designed which uses model to estimate artist rankings, then comparing with ground truth using unordered accuracy and edit distance

n	Edit Distance	Unordered Distance
3	0.1224 (0.0748)	0.1293 (0.0748)
5	0.1469 (0.1143)	0.1796 (0.1388)
10	0.1796 (0.1653)	0.3163 (0.3020)
25	0.1633 (0.1543)	0.5731 (0.5543)

Sample Suggestions

Ground Truth	Model	Baseline
The Beatles		
Radiohead	The 69 Eyes	Massive Attack
Coldplay	Leonard Cohen	DJ Shadow
Bob Dylan	Cat Stevens	Klaus Badelt
The Rolling Stones	Death Cab for Cutie	Era
Led Zeppelin	Built to Spill	Enya
In Flames		
Children of Bodom	Rhapsody of Fire	Immortal
Slipknot	Creed	Soilwork
Disturbed	Mudvayne	Converge
Dark Tranquillity	Marilyn Manson	Dark Tranquillity
Iron Maiden	Cradle of Filth	Trivium
Tupac		
Nas	Akon	Orbital
Snoop Dogg	Jay-Z	Flying Lotus
50 Cent	50 Cent	Mr. Scruff
Notorious B.I.G.	Lil Wayne	Shpongle
Dr. Dre	The Pussycat Dolls	King Crimson

Future Work

- Design better similarity metric – based more on audio similarity, or even ratings
- Collect more audio data for each artist

Acknowledgements



Shervine Amidi and the C230 team