

Song Recognition via Feed-Forward Deep Neural Networks

Campbell Clarkson

Drew Merrill

December 2021



Darla Moore School of Business

Introduction - Machine Learning

- Machine Learning is Ubiquitous today
 - Can deal with searching through large datasets
 - Solve heavily analytical problems

What is an application where machine learning can leverage the two advantages above?

AUDIO RECOGNITION!



Shazam - Modern Solution

- “Industrial Strength Music Search”
 - Wang et al. (2003)
- Helps users . . .
 - Find songs, advertisements, and media
- Without . . .
 - Knowledge of Title, Album, or Artist



Is it Trivial?

- Apple acquires Shazam for \$400 Million
- Modern music platforms allow for 'DIY' Music Publication



Our Project - Mimic Shazam

- **Objective:** Curate data and build a model to identify and name a particular song
- We build a dataset and deploy a model with a base accuracy of 94-97%

Has anyone done this?

The current body of work examines the robustness of recognition models relating to the audio *signal* (pitch, frequency, key, complexity)

- Xiao 2018, Tao and Getachew 2020

We use a “fingerprint” style, like Shazam¹, but with metadata of the song.

¹<https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition>

Data

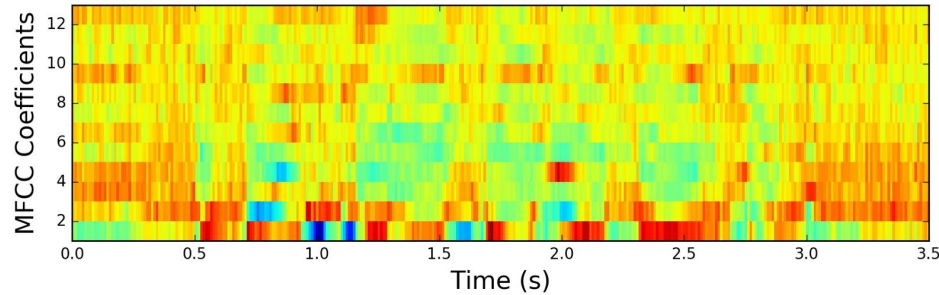
- 2000 Popular Songs Dataset
 - List of popular spotify Songs with related metadata
 - Danceability, tempo, energy, happiness, popularity
 - Mp3 files for select songs in model
- UrbanSound8k
 - Examination of robustness with noise added to model



Darla Moore
School of Business

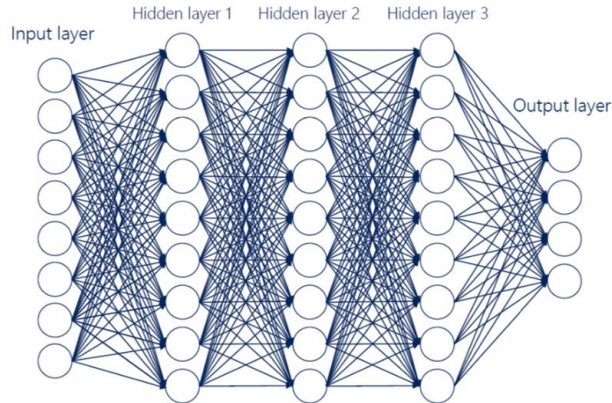
Data Processing

- We take a 20 song-subset of the 2000
 - Segmented into 3-second snippets
 - Snippets are stored according to class label
 - Obtain Mel Frequency Cepstral Coefficient for each segment
 - Logan (2000)



Model: Feed Forward DNN

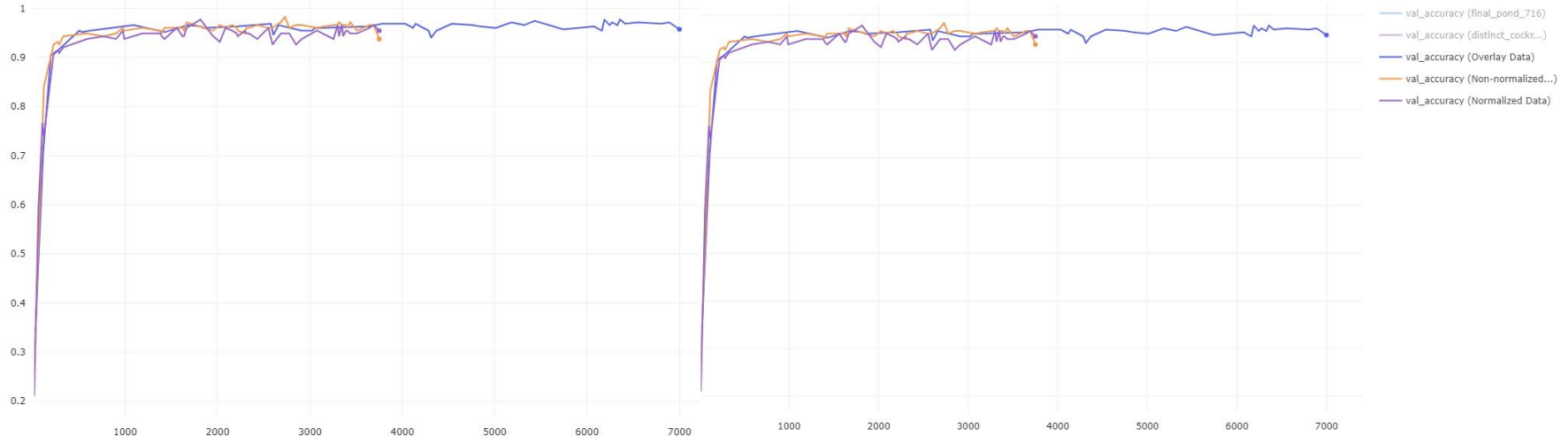
- We Implement a sequential DNN
- Mean MFCC Vector as input, model outputs prediction



Two Robustness Tests

- Normalizing the MFCC on the y-axis at each time period
- Do environmental noises overlayed on the music reduce effectiveness?

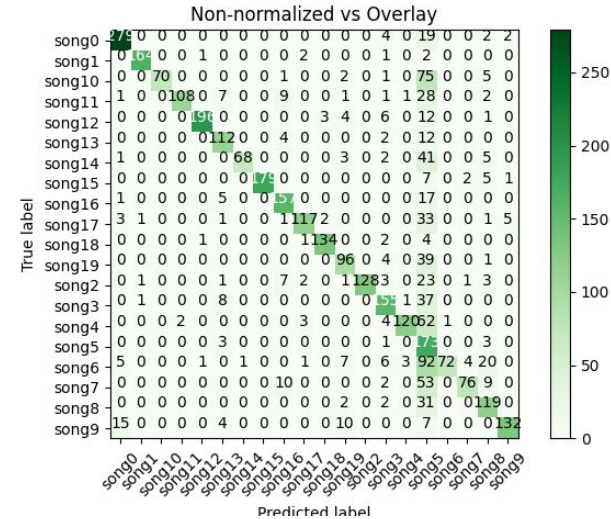
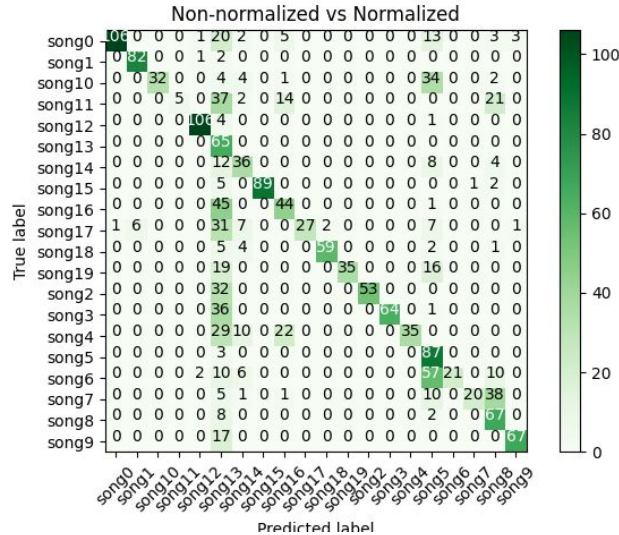
Final Results



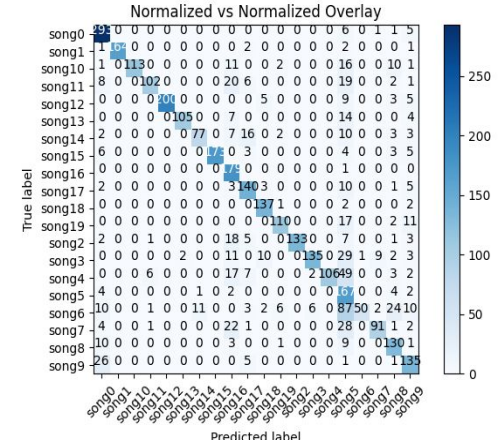
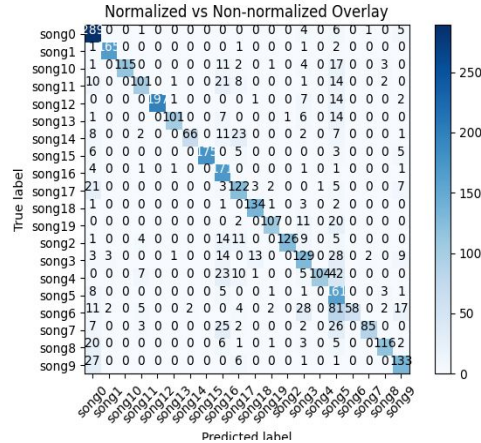
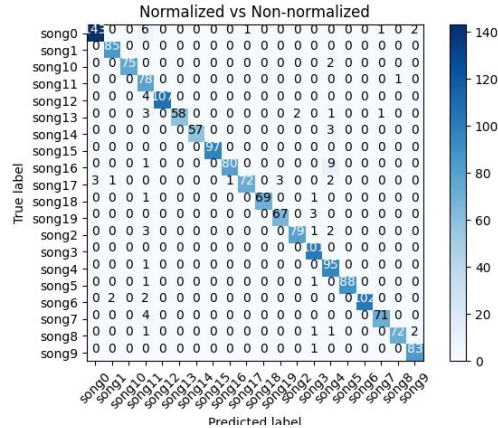
Test Data

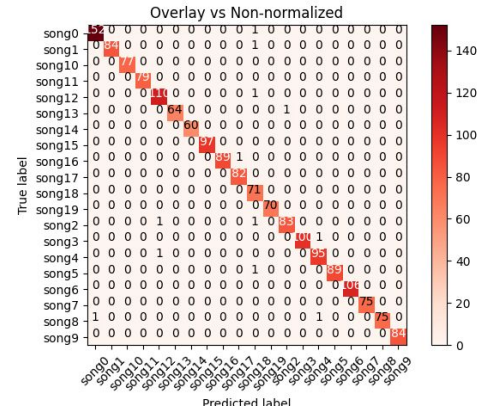
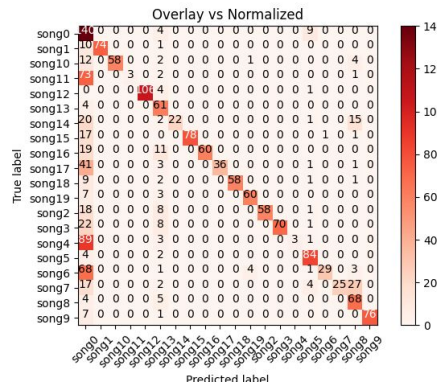
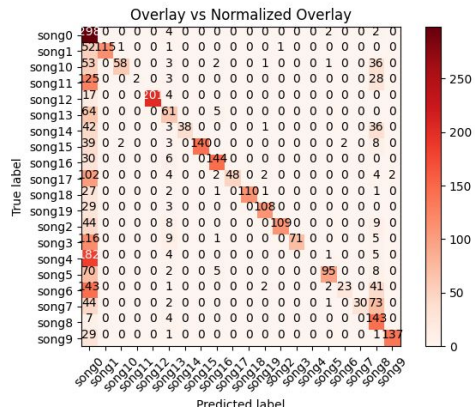
Validation Data

Non-normalized Model



Normalized Model





Challenges/Limitations

- We have not fully developed a usable application yet
 - Recording Audio is the key missing component
- Pre-processing audio data takes a very long time
 - 20 Songs = 1 hour
 - Scalability issues?

Conclusion

Audio Recognition is a useful tool with numerous practical implications

- We identify a model based on characteristics published by the music platform, as opposed to the signal
- The model is 94-97% accurate across three specifications

Thank You!



Darla Moore
School of Business