Matthew Lowery
CS366: Computational Linguistics
Final Project Paper

Musical Genre Classification using Song Lyrics

1. Description of the problem and my motivations

This project seeks further insight as to what types of traits determine a song's musical genre. This classification certainly appears to be more or less subjective as compared to strict and formulaic, with cross-listings (a song being labeled as more than one genre) among most songs and an ever-increasing list of such musical genres. From classical to electronica and dance music, how can we use machines to do the work for us when we have trouble deciding for ourselves? Billboard determines genre by "key fan interactions with music, including album sales and downloads, track downloads, radio airplay and touring as well as streaming and social interactions on Facebook, Twitter, Vevo, Youtube, Spotify and other popular online destinations for music" The main goal of this project is to see whether or not lyrics can be used to help decide what type of genre a song is and ultimately if any measurable aspects of a song that can be used for machine learning type algorithms would help in such classification. I am undertaking this project mostly as a way to entertain my curiosity and interest in music because plays such a large role in my life as well as in the majority of societies and cultures around the world.

2. Review of Existing Work

Musical genre classification has been a topic studied quite a bit amongst NLP research. With hoards of new music being broadcasted on a daily basis, it makes sense that companies like spotify, apple music, pandora, etc would want some automated mechanism for classifying a songs genre. A lot of previous research on genre classification has been done, but a lot of it lacks including lyrical content mostly because it is difficult to collect large scale lyrical data as well as for countless copyright issues. As such, lyrical datasets were quite small up until musiXmatch and the million song dataset joined forces to release lyrics for 237,662 tracks in a clean bags-of-words format (~2011). In 'Using Shared Vector Representations of Words and Chords in Music for Genre Classification' (https://www.isca-speech.org/archive/SMM_2019/pdfs/SMM19_paper_19.pdf) chords as well as lyrics from the musiXmatch and the million song dataset's bag of words were combined to predict musical genre. Instead of using the genre dataset that MSD released using musicbrainz tags (also in 2011), they scraped the genre tags from billboard's website, that included cross-listings, but did not have a large dataset ~3,000 genre tags from 5 genres: Latin, Country, Pop, Rock, RnB/Hip-Hop, whereas the MSD genre tag dataset (from http://millionsongdataset.com/blog/11-2-28-deriving-genre-dataset/) had ~59,600 genre tags. In combining musiXmatch and MSD's bags-of-words lyrics dataset and the MSD genre dataset, I was able to match 17,495 songs with their lyrics to specific genre tags from 10 genres: classic pop

and rock, folk, dance and electronica, jazz and blues, soul and reggae, punk, metal, classical, pop, hip-hop. Other features that are often considered in genre classification have been loudness, tempo, time signature, key, mode, duration, timbre, but we focus on lyrical relevance only

http://www.ee.columbia.edu/~dliang/files/FINAL.pdf

3. Describing the data, algorithms and methods used

This project mostly came down to properly curating the datasets for use in classification algorithms. There are two datasets that have been combined for such use. The first one was the musiXmatch dataset, which provided an official lyrical collection of the Million Song Dataset (which is a freely-available collection of audio features and metadata for a million contemporary popular music tracks). This dataset was a bags-of-words format, with each track being described as the word-counts for a dictionary of the top 5,000 words across the entire set of 237,701 tracks, where they have performed stemming and other normalizations. The second dataset was also found through the Million Song Dataset, where they released a dataset of ~59,600 songs with their respective genre tags from classic pop and rock, folk, dance and electronica, jazz and blues, soul and reggae, punk, metal, classical, pop, hip-hop. I combined these two datasets using the Million Song Dataset IDs that each song was labeled with. As such, I ended up with 17,495 songs with the bag-of-words lyrical format and genre tags. The format of the bags-of-words were something like: 1:2, 3:4, 130:6, 4000:50. So the number prior to each of the colons represents the associated word from the dictionary of

the top 5,000 words across the tracks and the number after the colon represents the amount of times that word appears in the song. I then made a vector representation of the word counts after splitting the training and testing data (75-25) for each individual genre as they were very skewed in representation:

In total Soul and Reggae had 1,055 tracks,
Hip-Hop had 92, Classical had 52, Jazz and Blues had 387, Metal had 1,251,
Dance and Electronica had 588, Pop had 776, Folk had 3,869,
Punk had 1,021, and Classic Pop and Rock had 8,404.

The training set had 75% of the songs for each genre respectively and was then input into a K-nearest-neighbors classifier. A k-NN classifier was used for the sake of its quick training time, ease of use, and functionality for classification.

4. The Results

In deciding the number of neighbors for the classifier I ran the accuracy for the amount of neighbors from 1-20:

```
1:  Accuracy: 0.3074988568815729
2:  Accuracy: 0.3904892546867855
3:  Accuracy: 0.403063557384545
4:  Accuracy: 0.3904892546867855
5:  Accuracy: 0.407864654778235
6:  Accuracy: 0.41586648376771834
7:  Accuracy: 0.41838134430727025
8:  Accuracy: 0.4204389574759945
9:  Accuracy: 0.41380887059899407
10: Accuracy: 0.4231824417009602
11: Accuracy: 0.4204389574759945
12: Accuracy: 0.4231824417009602
13: Accuracy: 0.4224965706447188
14: Accuracy: 0.42158207590306357
15: Accuracy: 0.4336991312299954
```

```
16: Accuracy: 0.4357567443987197
17: Accuracy: 0.43621399176954734
18: Accuracy: 0.4444444444444444
19: Accuracy: 0.4444444444444444
20: Accuracy: 0.45038866026520347
```

5. Analysis of shortcomings and ideas for the future

This project has a lot of potential in terms furthering the classifiers accuracy and functionality. As there are so many aspects of a song to consider in classifying its genre, so too are there many features for the classifier to consider. Including attributes like loudness, tempo, time signature, key, mode, duration and timbre would likely improve the classifiers accuracy, but for times sake, I was unable to encompass all of these. When sifting through each of the datasets (the bags-of-words and the genre tags) to match the