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. Their accuracy using lyrics only was 36.2% and went up to 37.9% with the inclusion of chords. 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.

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:

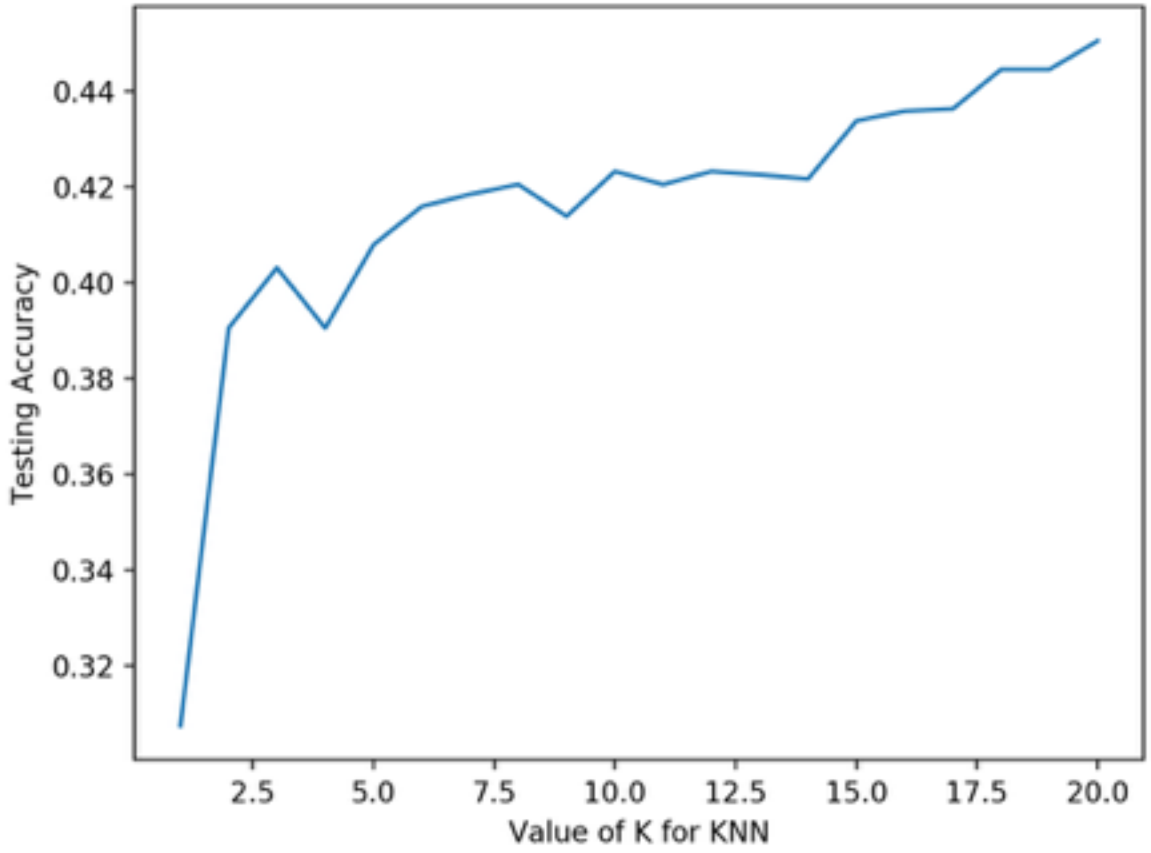
In total Soul and Reggae had 1,055 tracks,
Hip-Hop had 92, Classical had 52, Jazz and Blues had 387, Metal had 1,251,
Dance and Electronica had 588, Pop had 776, Folk had 3,869,
Punk had 1,021, and Classic Pop and Rock had 8,404.

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:

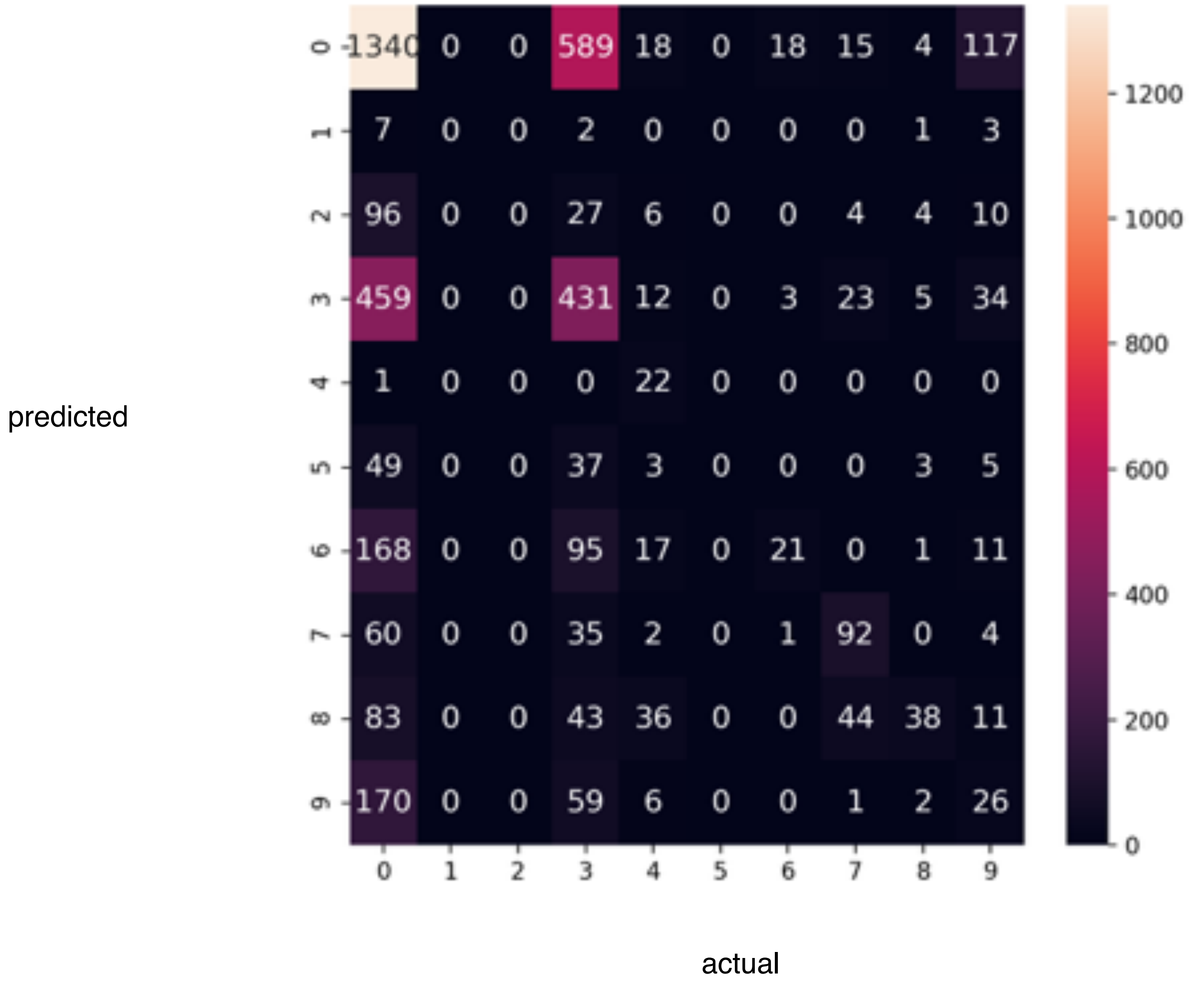
```
1:  Accuracy: 0.3074988568815729      11: Accuracy: 0.4204389574759945
2:  Accuracy: 0.3904892546867855      12: Accuracy: 0.4231824417009602
3:  Accuracy: 0.403063557384545       13: Accuracy: 0.4224965706447188
4:  Accuracy: 0.3904892546867855      14: Accuracy: 0.4215820759030636
5:  Accuracy: 0.407864654778235       15: Accuracy: 0.4336991312299954
6:  Accuracy: 0.41586648376771834     16: Accuracy: 0.4357567443987197
7:  Accuracy: 0.41838134430727025     17: Accuracy: 0.4362139917695473
8:  Accuracy: 0.4204389574759945      18: Accuracy: 0.4444444444444444
9:  Accuracy: 0.41380887059899407     19: Accuracy: 0.4444444444444444
10: Accuracy: 0.4231824417009602      20: Accuracy: 0.4503886602652035
```



```
                          precision    recall  f1-score   support

0 classic_pop_and_rock       0.55      0.64      0.59      2101
1            classical       0.00      0.00      0.00        13
2 dance_and_electronica      0.00      0.00      0.00       147
3                 folk       0.33      0.45      0.38       967
4              hip_hop       0.18      0.96      0.30        23
5       jazz_and_blues       0.00      0.00      0.00        97
6                metal       0.49      0.07      0.12       313
7                  pop       0.51      0.47      0.49       194
8                 punk       0.66      0.15      0.24       255
9      soul_and_reggae       0.12      0.10      0.11       264

            accuracy                           0.45      4374
           macro avg       0.28      0.28      0.22      4374
        weighted avg       0.44      0.45      0.42      4374
```



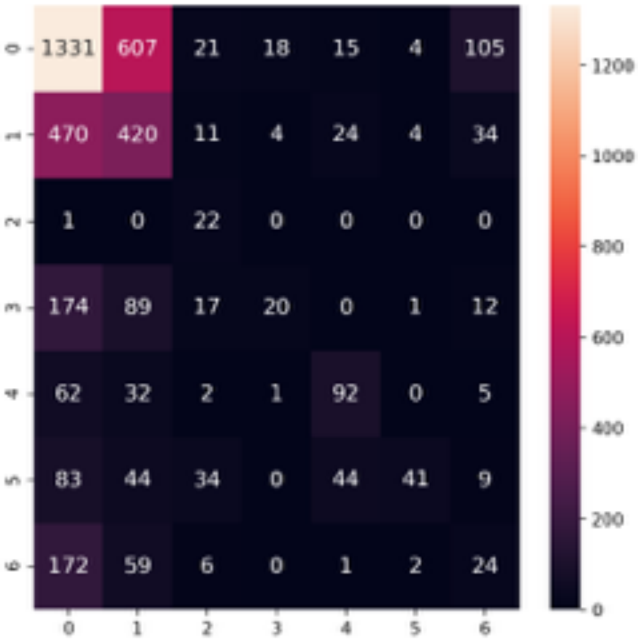
	0	1	2	3	4	5	6	7	8	9
0	1340	0	0	589	18	0	18	15	4	117
1	7	0	0	2	0	0	0	0	1	3
2	96	0	0	27	6	0	0	4	4	10
3	459	0	0	431	12	0	3	23	5	34
4	1	0	0	0	22	0	0	0	0	0
5	49	0	0	37	3	0	0	0	3	5
6	168	0	0	95	17	0	21	0	1	11
7	60	0	0	35	2	0	1	92	0	4
8	83	0	0	43	36	0	0	44	38	11
9	170	0	0	59	6	0	0	1	2	26

```
WITHOUT JAZZ, ELEC, ClASS
----------------------------------------------------------------------
-----------------
(12351, 4868)
(4117, 4868)
Accuracy: 0.47364585863492836
                      precision    recall  f1-score   support

0 classic_pop_and_rock       0.58      0.63      0.61      2101
1                 folk       0.34      0.43      0.38       967
2              hip_hop       0.19      0.96      0.32        23
3                metal       0.47      0.06      0.11       313
4                  pop       0.52      0.47      0.50       194
5                punk        0.79      0.16      0.27       255
6     soul_and_reggae        0.13      0.09      0.11       264

            accuracy                           0.47      4117
           macro avg       0.43      0.40      0.33      4117
        weighted avg       0.49      0.47      0.46      4117
```

predicted



	0	1	2	3	4	5	6
0	1331	607	21	18	15	4	105
1	470	420	11	4	24	4	34
2	1	0	22	0	0	0	0
3	174	89	17	20	0	1	12
4	62	32	2	1	92	0	5
5	83	44	34	0	44	41	9
6	172	59	6	0	1	2	24

actual result

Here I remove the genres classical, dance and electronica, and jazz and blues. All of these genres largely had songs without lyrics or very few, so it would make sense to remove them so the classifier is not confused by the lack of data relative to the classification. It seemed to improve the accuracy from 45 percent to 47 percent.

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. For example, When sifting through each of the datasets (the bags-of-words and the genre tags) to match the songIDs and ultimately have the genre/lyric dataset for training the classifier, the program took longer than several hours. It would also likely help to try more instances of the amount of neighbors in the k-NN classifier or even several different types of classifiers. Even further, it would have been interesting and possibly insightful to run an unsupervised algorithm on the initial bags-of-words dataset. I am also sure that my lack of musical/musical theory background might have hindered the project in some way.