Classifying Songs by Genre Using Lyrics

Adam Gulick

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. author

1 Related Work

The question that this project stemmed from is not a novel idea. There has been plenty of work done with similar ideas of classifying music into genres. The largest variety in the studies on this topic comes from the models that are trained as well as the number of genres that the songs were classified into.

Even just looking at a few of the recent studies in regards to this question you can see many different models and approaches. Hu et al. (2009) looked into classifying 5,585 songs into 18 different mood categories using the songs lyrical text using Support Vector Machines and determined that Bag of Words were the most successful type. This problem was addressed again by Hu et al. (2017) when they looked at whether or not using a multimodal system made a significant difference in music mood classification.

In the study by Leszczynski and Boonyanit (2021), a LSTM model was trained using GloVe embeddings of the song lyrics, reaching a peak success rate of 68%.

Meenakshi et al. (2020) uses a Bag of Words method for tf-idf calculations and was able to acieve a 63% accuracy.

In a 2020 study, DeMasi (2020) ran this problem on twenty-four different models and found the most success using a Deep Neural Network trained on a tf-idf feature set.

Canicatti (2016) discusses the importance of choosing a proper set of genres to be classified into. For a classifier to have a high accuracy, you need genres to be general enough for each genre's lyrics can be accurately classified but specific enough for each genre to have a unique set of lyrics. A set of 5 distinct genres were established for this study.

A similar study was done on Filipino Music by Abisado et al. (2021), in which a Naive Bayes classifier was trained using the scikit-learn library to analyze and classify the moods of more than 200 songs into a binary mood of happy or sad.

A 2017 study by Dammann and Haugh (2017) looked at not only the lyrics of a song on Spotify, but also the song preview and the album artwork when attempt to classify the genre of a song. While the latter two are not relevant to this study, the model used to analyze the lyrics can be recreated for this study. A Naive-Bayes classifier was used specifically for the lyrics. When combining all three types of data, an accuracy of 91.75% was achieved.

Using two more recent language models, BERT and DistilBERT, Akalp et al. (2021) were able to achieve accuracies of 77.63% and 71.29%. A BERT model uses the contextual relationship between words to generate predictions.

Despite there being a wide variety of approaches and training models to this problem. They all follow the same rough pattern. Every study gets a large dataset consisting of songs across all gen-

res. This data is then tokenized and used to train a model. A new set of songs is then given to the model as test data, and the accuracy is recorded. The largest differences in the various studies comes from the number of genres that the songs were sorted into and the models that were trained. Some variance came from the different ways that the text could be tokenized.

2 Methods

The data for this project is rather simple. A set of songs and their lyrics from a wide variety of artists and genres will be needed to train and test the model. In attempts to avoid having to fetch my own data–which would be a long and cumbersome task– I looked at the other studies to see if there was a good pre-existing dataset. There was one particular dataset[1] from Kaggle was used in multiple of the studies so that is the dataset that will be used for this project.

For this project, five genres will be used to classify the songs. These genres are: Rock, Rap, Pop, Country, and Dance. Canicatti (2016) also used five genres. However, for that study, Jazz replaced Dance in the list of genres. The dataset that will be used did not have any songs denoted as Jazz, thus it is omitted and Dance will take its place. For this study, two different models will be trained to classify songs into genre. These two models will then be compared to determine the difference in success.

For both classifiers, the required text preprocessing will be the same for both. The lyrics will have to be taken from the data set, tokenized, and given to the classifier as training data with the correct genre. To be able to accurately compare the two models, the same set of songs will be used to train both models. Both models will then be tested on the same set of songs unique from those in the training data.

2.1 Naive Bayes Classifier

The first model that will be trained is a Naive-Bayes Classifier using the NLTK module. This is the same model that was trained in Akalp et al. (2021) as well as others. That study was able to reach a 41.67% success rate of correctly classifying the songs. Canicatti (2016) also used a Naive-Bayes classifier and was able to reach a 30.98% success rate across the five genres.

The training data will consist of a set of features obtained by separating all the songs by genre and giving the lyrics and the tag to the model. This data will then be used to train the model on the training data, which will then be used to determine the accuracy when ran on the test data.

[1] https://www.kaggle.com/datasets/neisse/scrapped-lyrics-from-6-genres?resource=download

Acknowledgements

References

Mideth Abisado, Mardyon Yongson, and Ma.Ian De Los Trinos. 2021. *Towards the Development of Music Mood Classification of Original Pilipino Music (OPM) Songs Based on Audio and Lyrics Keyword*, page 87–90. Association for Computing Machinery, New York, NY, USA.

Hasan Akalp, Enes Furkan Cigdem, Seyma Yilmaz, Necva Bölücü, and Burcu Can. 2021. Language representation models for music genre classification using lyrics. *2021 International Symposium on Electrical, Electronics and Information Engineering*.

Anthony Canicatti. 2016. Song genre classification via lyric text mining.

Tyler Dammann and Kevin P. Haugh. 2017. Genre classification of spotify songs using lyrics , audio previews , and album artwork.

Nick DeMasi. 2020. Can you hear me now? predicting song genre from song lyrics using deep learning.

Xiao Hu, Kahyun Choi, and J. S. Downie. 2017. A framework for evaluating multimodal music mood classification. *Journal of the Association for Information Science and Technology*, 68.

Xiao Hu, J. S. Downie, and Andreas F. Ehmann. 2009. Lyric text mining in music mood classification. In *ISMIR*.

Megan Leszczynski and Anna Boonyanit. 2021. Music genre classification using song lyrics.

K. Meenakshi, M. Safa, G. Geetha, G. Saranya, and J. SundaraKanchana. 2020. Music genre classification using lyric mining based ontf-idf.