INT15 Final Project

Spotify Genre Analysis

Qirong He, Zilong Liu

University of California, Santa Barbara

Abstract

People's preferences for music genres are decided by multiple factors, such as instrumentation, music style, lyrics, artists, and emotions conveyed by music. Therefore, understanding what determines the genre of a song and analysis of music reviews are crucial in finding the reasons behind it. In our project, we explored the task of music genre classification and sentiment analysis based on album reviews. Our data comes from Spotify Web API, including 13 acoustic features of more than 300 thousand songs and their related albums, and Pitchfork reviews on albums. First, two measurements based on average ranks and average scores were implemented respectively to study the popularity of different genres. Second, based on acoustic features such as 'danceability', we used PCA to achieve dimension reduction and classified every two genres of the seven selected unique music genres. At least nine principal components were needed to explain 90% of the total variance. Logistic regression was then used to distinguish one genre from another. The logistic regression prediction achieved the highest accuracy for 'Metal' and 'Rap' according to cross-validation results. Also, keywords in Pitchfork reviews were extracted after tokenization, lemmatization, and part-of-speech tagging. The most common keywords, including adjectives, used to describe different genres were found by calculating their frequency. Finally, a TF-IDF music recommendation model was created to find Top 10 similar albums to a given album based on music reviews.

Introduction

Understanding people's preferences for the music of different genres is crucial in music analysis and recommendation these days. A wide variety of studies uses machine learning algorithms, such as support vector machines (SVM), to classify music genres and build music genre classification systems [1]. Based on spectro-temporal features, these systems achieve feature selection and classifier construction. Feature-selection and dimension-reduction algorithms can be used such as linear discriminant analysis (LDA)[2], non-negative Tensor factorization (NTF)[4] and SVM ranker methods[1]. In the meantime, the polarity of customer reviews on music was explored by doing sentiment analysis[3], helping understanding evolving sentiments in music reviews. Features extracted from harmonic elements can also satisfactorily predict music genre[5].

Our goal is to understand the features of music genres and build a recommendation system based on music reviews. The Spotify datasets include acoustic features of mil ions of songs released from 1963 to 2018 and the albums they belong to. The Pitchfork dataset contains reviews from 1999 to 2019. By implementing exploratory data analysis, we aim at studying these research questions:

1. What genre do people prefer compared with others?
2. Can we use principal component analysis (PCA) to classify songs of different genres?
3. How can we predict the genre of a song based on its audio features?
4. Can we use music reviews to achieve music recommendation?

Data and Methods

The Spotify datasets come from Spotify Web API [6], including acoustic_features.csv and albums.csv. "acoustic_features.csv" contains 339855 songs and their acoustic features, names, albums that they belong to and release dates and "albums.csv" includes 573947 albums on their ranks, lengths, track lengths and artists that they belong. The Pitchfork review dataset reviews.csv comes from Pitchfork [7], including reviews, scores, dates, authors and their roles. The two datasets fit our purpose of music analysis and recommendation and have enough precision to achieve our goals.First, by merging these three datasets and taking a look at the average scores or ranks of all songs related to different genres, we can get a glimpse at the popularity of different genres. Besides, the audio features of the audio features of each song include 'acousticness', 'danceability', 'duration_ms', 'energy', 'instrumentalness', 'key', 'liveness', 'loudness', 'mode', 'speechiness', 'tempo','time_signature' and 'valence'. By implementing dimension reduction techniques, such as PCA, after data normalization, we can classify every two genres based on these 13 audio features. We selected seven genres, which are only ‘unique’ genres such as Rap and Rock because combined genres such as ‘Rap, Rock’ may have the similar audio features to ‘Rap’ or ‘Rock’. After examining the correlation matrix indicating the correlation between every two features, further analysis like logistic regression can be used to help make predictions for every two genres. Cross-validation will help determine how well our trained models will be in prediction.

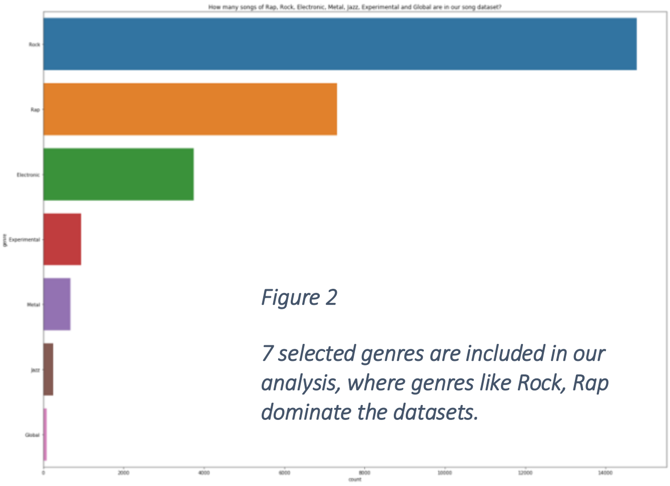

Figure 2

7 selected genres are included in our analysis, where genres like Rock, Rap dominate the datasets.

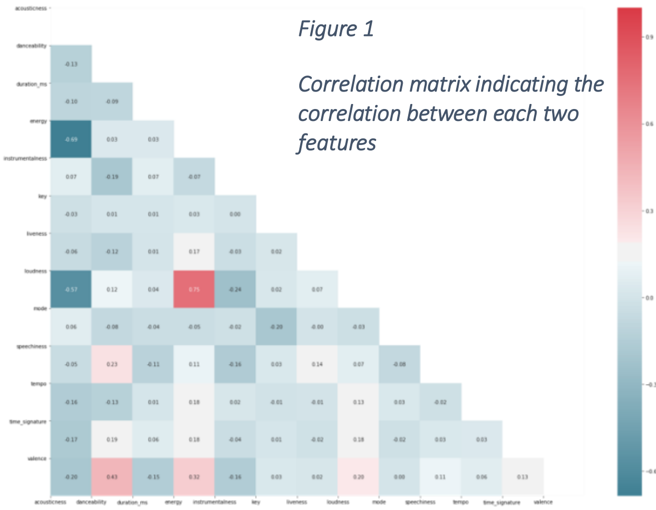

Figure 1

Correlation matrix indicating the correlation between each two features

Furthermore, using tokenization and lemmatization to extract keywords from reviews, we can see what the most common words used to describe songs of different genres are. And the use of part-of-speech tags can help us find out what adjectives are often used to describe specific genres. Finally, by creating a tf-idf model, we can achieve music recommendation based on the seven selected genres are included in our similarity between reviews on albums.

However, some records in the datasets have NaN values of certain features, meaning these records should be removed afterward. Besides that, audio features have different units, which means data normalization is necessary. Also, songs of different genres have different sample sizes (see Figure 2), which might lead to false negatives and positives when using regression analysis due to an unbalanced distribution of data. Therefore, we need to balance the influence of sample sizes before putting them in regression analysis.

▪ Results

I. Users' Preferences for Different Genres

genre	rank
Jazz, Metal	29.50
Electronic, Experimental, Global, Rock	42.00
...	...
Experimental, Rap	161.50
Electronic, Experimental, Folk/Country	179.00

Table 1 Popularity Based on Average Ranks

genre	score
Rock, Experimental	8.45
Experimental, Rap	8.40
...	...
Jazz, Experimental	4.40
Electronic, Jazz, Pop/R&B	2.60

Table 2 Popularity Based on Average Scores

By grouping by ‘genre’ and ‘album’, we can get the rank and score of each album. If measuring the popularity of a genre based on the average ranks of all songs related to it, we can see 'Jazz, Metal' ranks first, while 'Electronic, Experimental, Folk/Country' has the lowest popularity. And if measuring the popularity of a genre based on the average scores of all songs related to it, we can see ‘Electronic, Jazz, Pop/R&B' has the lowest score, while 'Rock, Experimental' has the highest score.

II. Using PCA to classify Genres

Why are these genres have different popularity? It is interesting to see that every song has its unique audio features, which could be the reason that users have preferences for specific genres. PCA will help us classify different genres by drawing clusters, and the best result we wanted is to reduce dimensions into the 2D projection. However, the results (see Figure 3) seem to be unsatisfactory because there is no clear cutoff point to distinguish one genre from another.

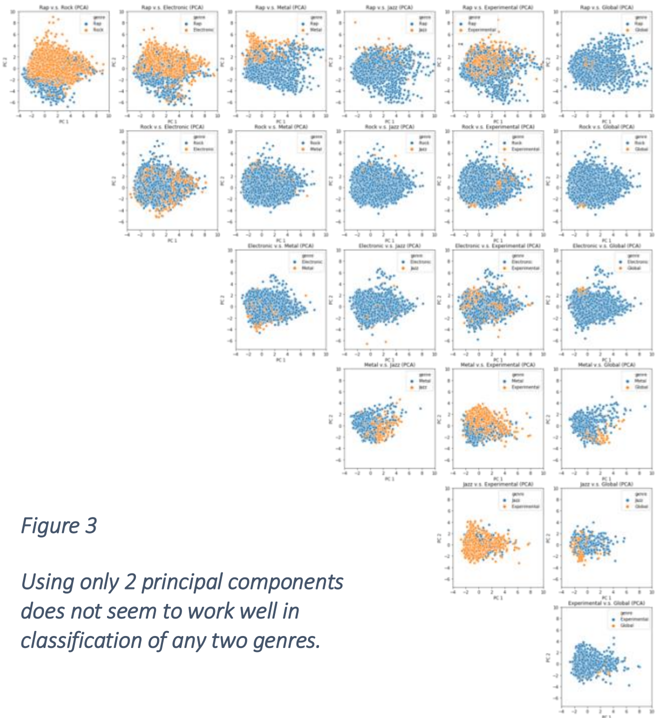

Figure 3

Using only 2 principal components does not seem to work well in classification of any two genres.

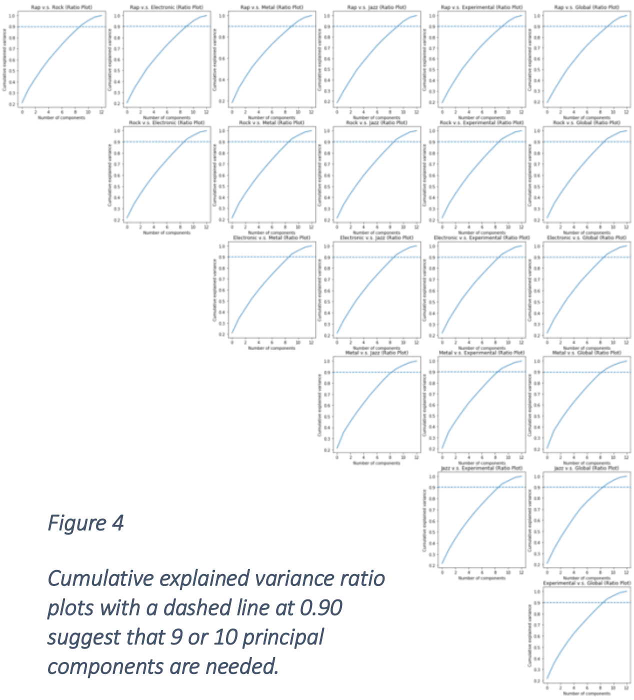

Figure 4

Cumulative explained variance ratio plots with a dashed line at 0.90 suggest that 9 or 10 principal components are needed.

By further looking at the cumulative explained variance ratio plots (see Figure 4), we found that for all the classifications, 9 to 10 principal components are needed to explain 90% of the total variance.

III. Using Logistic Regression to Make Prediction for Genres

After examining the correlation matrix (see Figure 1), we can find that there is no correlation coefficient larger than 0.8. Therefore, al 13 audio features were selected to be explanatory variables in the logistic regression analysis on music genre prediction. Also, we reduce the bias by balancing the sample sizes of each two genres before being input in logistic regression models.

Maximum score is : 0.83				
genre	precision	recall	f1-score	support
Metal	0.85	0.85	0.85	171
Rap	0.85	0.85	0.85	168

Table 3 Prediction model for ‘Metal’ and ‘Rap’ has the highest score.

Minimum score is : 0.16				
genre	precision	recall	f1-score	support
Electronic	0.55	0.37	0.44	243
Experimental	0.5	0.67	0.57	226

Table 4 Prediction model for 'Electronic' and 'Experimental' has the lowest score.

Furthermore, cross validation helped us evaluate our models. The maximum score is 0.83, which occurs when predicting 'Metal' and 'Rap' (see Table 3), while the minimum score is 0.16, which occurs when predicting 'Electronic' and 'Experimental' (see Table 4). The mean score is 0.61, and the median score is 0.67.

IV. Music Recommendation Based on Reviews

By extracting keywords after tokenization, lemmatization, and part-of-speech tagging, we can see the most common adjectives used to describe genres, such as 'Pop/R&B'. According to the frequency plot (see Figure 5), 'Pop/R&B' has characteristics like 'new', 'vocal' and 'sound', which matches our understanding of this genre.

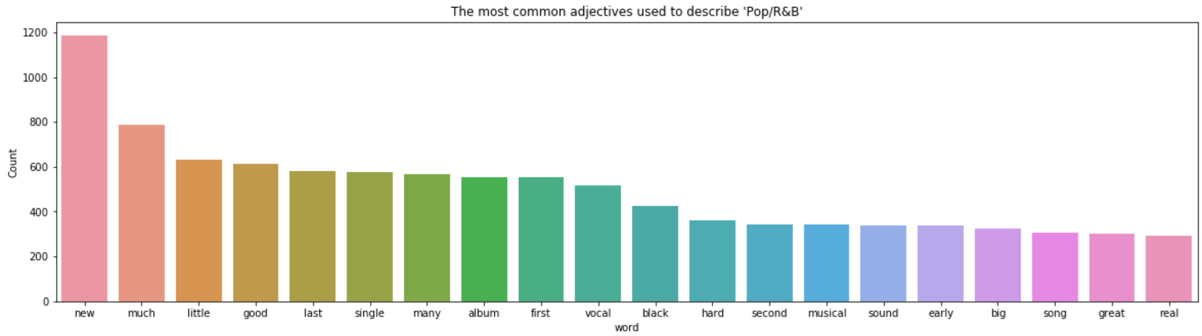

Figure 5 Frequency plot showing the most common adjectives used to describe 'Pop/R&B' in a descending order

Besides common words like 'like', 'album', 'music', 'song', 'band', 'track', 'even', 'new' or 'sound', the most common words used to describe seven selected genres are listed as follows. It is reasonable to see that the most common word used to describe a genre is its name.

The word 'time' is often used in describing Rock.
The word 'way' is often used in describing Electronic.
The word 'rap' is often used in describing Rap.
The word 'work' is often used in describing Experimental.
The word 'metal' is often used in describing Metal.
The word 'jazz' is often used in describing Jazz.
The word 'guitar' is often used in describing Global.

At last, we created the tf-idf model to find similar albums with respect to the same genre shows good results in recommending related albums, providing Top 10 words associated with the input album and Top 10 most similar albums (see Figure 6). As follows, we give an example using Reputation produced by **Taylor Swift** with the results.

Top 10 words associated with this music by tf-idf are:

'broken' (tf-idf score = 0.243)
'west' (tf-idf score = 0.174)
'heaven' (tf-idf score = 0.150)
'blitheness' (tf-idf score = 0.109)
'elm' (tf-idf score = 0.109)
'encouragingly' (tf-idf score = 0.109)
'highhatriding' (tf-idf score = 0.109)
'hurtles' (tf-idf score = 0.109)
'incongruently' (tf-idf score = 0.109)
'peevishness' (tf-idf score = 0.109)

Top 10 most similar musics for album Reputation are:

1 Gone Now produced by Bleachers (similarity score = 0.081)
2 Witness produced by Katy Perry (similarity score = 0.065)
3 Love Remains produced by How to Dress Well (similarity score = 0.064)
4 The Rite of May EP produced by Oklou (similarity score = 0.060)
5 Daydream produced by Mariah Carey (similarity score = 0.059)
6 25 produced by Adele (similarity score = 0.057)
7 Devotion produced by Laura Jean (similarity score = 0.057)
8 Know It All produced by Alessia Cara (similarity score = 0.056)
9 Kindred produced by Passion Pit (similarity score = 0.055)
10 HEAVN produced by Jamila Woods (similarity score = 0.054)]

Top 10 associated words and most similar albums to Reputation produced by Taylor Swift

▪ Discussion

The conclusions are as follows:

1) The popularity of different genres was determined based on average ranks and scores respectively;
2) PCA successfully helps classification of different genres, and at least 9 principal components are needed to explain 90% of total variance;
3) Logistic regression used for prediction for genres shows an overall high precision and cross-validation shows the prediction model to distinguish 'Metal' from 'Rap has the highest score while the model to distinguish 'Electronic' from 'Experimental' shows the lowest score;
4) We extracted the most common keywords (especially adjectives) used to describe different genres, and construct a music recommendation system based on reviews.

Though we achieve our goals, there are certain aspects that can be improved. PCA did not show good classification results while projecting the data matrix into the 2D and logistic regression model to distinguish specific genres show relatively low precision, indicating audio features are not the only factors that determine the genre that a song belongs.

In further studies, we can combine the results of review-based similarity analysis with the results from genre classification to enhance our music recommendation system. We also hope to involve more in-depth analysis of the feature space, combining different types of features, like the music video of songs, the lyrics, etc. To use interpretable features to predict music genres, assessing more information about songs, and constructing better classifiers using other methods such as decision trees can bring us new insights.

▪ References

[1] Shin-Cheol Lim, Jong-Seol Lee, Sei-Jin Jang, Soek-Pil Lee, and Moo Young Kim, Senior Member, IEEE "Music-Genre Classification System based on Spectro-Temporal Features and Feature Selection", IEEE Transactions on Consumer Electronics, Vol. 58, No. 4, November 2012

[2] Ming-Ju Wu and Jyh-Shing R. Jang. 2015. Combining acoustic and multilevel visual features for music genre classification. ACM Trans. Multimedia Comput. Commun. Appl. 12, 1, Article 10 (August 2015), 17 pages. DOI: http://dx.doi.org/10.1145/2801127

[3]Sergio Oramas, Luis Espinosa-Anke, Aonghus Lawlor, Xavier Serra, Horacio Saggion"Exploring Customer Reviews For Music Genre Classification and Evolutionary Studies"

[4] E. Benetos and C. Kotropoulos, "Non-negative tensor factorization applied to music genre classification," IEEE Audio, Speech, and Language Process., vol. 18, no. 8, pp. 1955-1967, 2010.

[5] Bruna D. Wundervald, Bruna D. Wundervald "MACHINE LEARNING AND CHORD BASED FEATURE ENGINEERING FOR GENRE PREDICTION IN POPULAR BRAZILIAN MUSIC"

[6] Spotify Web API. https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

[7] Pitchfork Reviews. https://pitchfork.com/reviews/albums/