STA 325 Final Project

Genre Classification using Spotify Data

Connie Wu, Jason McEachin, Joe Choo, Scott Heng

10/23/2020

1 Introduction

Cataloging and organizing music is an essential aspect of collecting and storing music. It allows us not only to identify and differentiate music, but also allows us to better understand the evolution of different types of music over periods of time. Genre has been the main and most efficient mode to categorize different kinds of music, and genre classification has always been a consistent challenge due to the complex nature of songs and the difficulty in differentiating the unique features specific to each genre. In the past, genre classification was largely performed by using pattern recognition after breaking the songs down frame by frame, and understanding the elements of chord progressions and stylistic features of the song [1]. However, the rise of big data has allowed us to more efficiently and accurately extract audio features and consequently automate the arduous task of classifying songs in particular genres. Genre classification is not only important in increasing the efficiency of cataloging music which is relevant to music companies and artists in organizing elements of their craft, but also for academics who wish to better understand the evolution of music and particular genres in their research.

Our study aims to build on modern statistical techniques that perform genre classification, by predicting the genre of songs based on multiple audio features each song possesses. Using well-known classification techniques such as logistic regression and support vector machines (SVM), we leverage the substantial capacity of modern computing to perform such statistical modeling on a large dataset of songs and various audio features taken from Spotify, and compare the performance of both classification techniques so as to evaluate their predictive accuracies, strengths and weaknesses of using either approach for genre classification. Using both methods to perform binary classification and also multi-group classification, this study aims to comprehensively and comparatively evaluate the effectiveness of both statistical methods with respect to music analytics.

Our paper is organized holistically and with simplicity to provide a comprehensive report on genre classification, starting with introducing research goals and providing the background motivations for the study. In Section 2, we provide a a description of the data as well as extensive exploratory data analysis which will motivate some design decisions during our statistical modeling. Section 3 describes our methodology with respect to implementing logistic regression and support vector machines for predicting genres of songs, structured with 2-group binary classification and multi-group classification. In Section 4, we discuss our models' results as well as model diagnostics to evaluate the accuracy of each of the models' fit, and relate our results to relevant parties that will use genre classification. Finally in section 5, we consolidate our findings and conclude our study with its strengths and its limitations.

2 Data

Our data was taken from Spotify's API, that draws from its large database of songs and their respective audio features. The data set was consolidated and released on kaggle, containing 232,725 tracks across 26 genres [2]. Each data point represents a song along with its tagged genre and various audio attributes such as tempo, key, danceability, valence, acousticness etc. For a detailed description of each feature, please refer to Table 2.1.

Table 1: Table 2.1 Descriptions of data set variables

Col	Column.Name	Description	Value
1	acousticness	A measure of degree of how acousitc a track is. 1.0 is most acoustic	float (0.0-1.0)
2	danceability	A measure based on a combination of music elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 1.0 is most danceable	float (0.0-1.0)
3	duration_ms	Duration of the track in milliseconds	integer
4	energy	A perceptual measure of intensity and activity attributing dynamic range, perceived loudness, timbre, onset rate, and general entropy.	float (0.0-1.0)
5	instrumentalness	A measure of the presence of vocals, rap or spoken word. The higher the value, the greater likelihood the track contains no vocal content.	float (0.0-1.0)
6	key	Integers that map to pitches using the standard pitch class (C=1, C#=2 . . .)	integer
7	liveness	A measure that detects the presence of an audience in the recording. A value above 0.8 provides strong likelihood that the track is live	float (0.0-1.0)
8	loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track. Values typical range between -60 and 0 db	float (0.0-1.0)
9	speechiness	A measure that detects the presence of spoken words in a track	float (0.0-1.0)
10	popularity	A measure of the song's popularity, calculated by considering the track's total number of plays and how recent those plays are.	integer (0-100)
11	tempo	The overall estimated tempo of a track in beats per minute (BPM)-the speed or pace of a given piece and derives directly from the average beat duration.	float (0.0-1.0)
12	valence	A measure describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric)	float (0.0-1.0)
13	year	Year that the track was released	integer
14	mode	Modality of the track (1=Major, 0=Minor)	integer
15	name	Name of track	string
16	genres	Genres associated with the track's artist(s)	array
17	artist	Name of the track's main artist	string

Before proceeding further, we performed some data cleaning. Firstly, we eliminated any observations that had the genre 'acapella', because it was a duplicate of a full song already included in the dataset and it had only 119 observations. We also removed any observations with 'Comedy' as they are lengthy tracks of spoken word by comedians, and we considered them not to be actual music.

[Got rid of children's music bc count was too high when combined (twice others)] [Got rid of Reggae because Reggae and Ska are considered very similar, Raggaeton is the preferred genre nowadays]

2.1 Exploratory Data Analysis

We began our exploratory data analysis to better understand the composition of our dataset. Starting with understanding the breakdown of genres, we see from Fig 1 that there is a good amount of observations for each group (>6000), which tells us that there will be enough observations to classify each of our 22 genres, and techniques like bootstrapping or strategies to deal with limited data do not need to be considered.

We then looked to understand the compositions of audio features based on genres, by generating density plots for each audio feature. Fig 1.2 shows the density plots of 6 audio features, and we can observe that

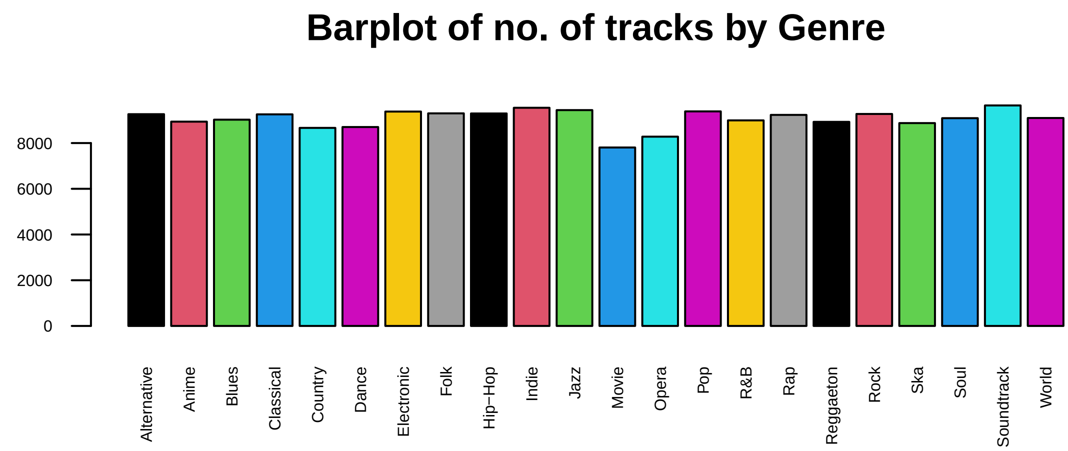


Figure 1: Count of tracks per Genre

although for features such as danceability, energy and tempo, each genre has a relatively distinctive density plot while for features such as liveness, loudness and valence, the densities are much hard to differentiate. Please refer to the appendix for density plots for the remaining audio features.

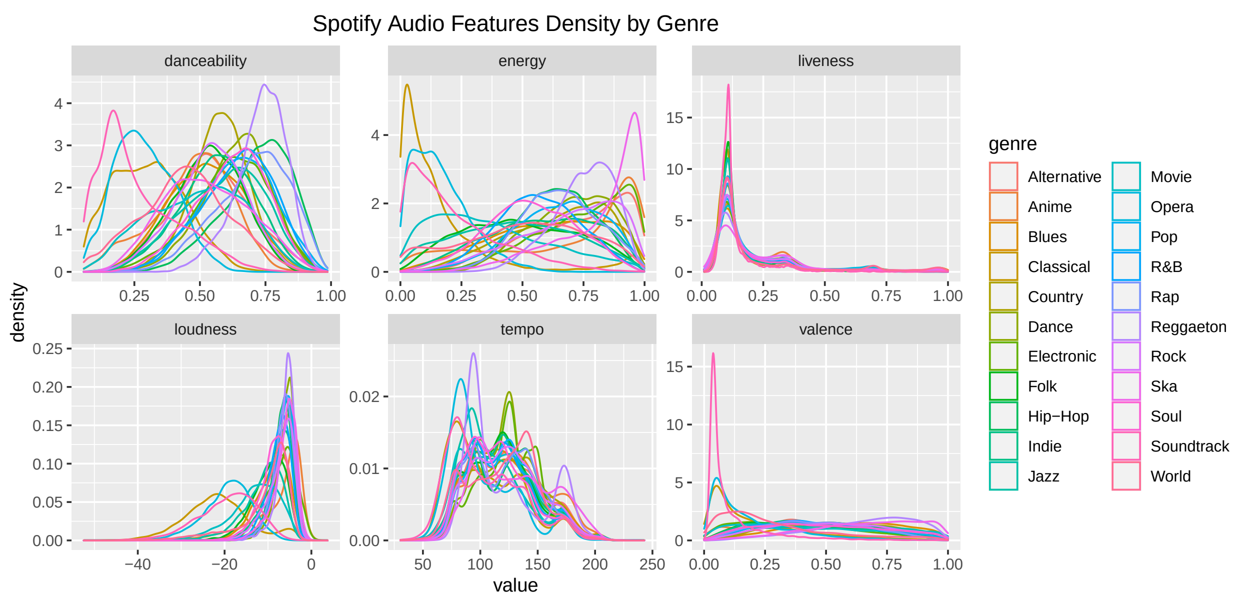


Fig 1.2 Density plots by Genre for 6 audio features

[I was thinking we could add Joe's tableau plots here]

3 Methodology

3.1 Binary and Multinomial Logistic Regression

The first of our two classification techniques we chose to model to predict genres is logistic regression. Logistic regression is appropriately used in cases when the dependent variable is categorical, which in our study's dependent variable being genre. It is not only computationally efficient but is produces results that are easy to interpret. As our EDA has shown a substantial amount of observations (>6000 for each genre) that are much more than the number of features (~13), there is a high level of confidence that overfitting will not occur, however more comprehensive model diagnostics will be performed and described in the later sections. Logistic Regression is usually performed when the dependent variable is dichotomous, meaning that the model can perform predictive classification over 2 genres. Also known as the log-odds model, logistic regression can be written mathematically as:

$$l = \log(\frac{p}{1-p}) = \beta_0 + \sum_{i=1}^{13} \beta_i x_i$$

where l is the log-odds and $p = P(Y = 1)$ is the probability of the observation being classified as one group labelled $Y = 1$, β_0 is the intercept and β_i are the coefficients of the 13 predictors, in our study being audio features, represented as x_i.

While logistic regression can be used for binary classification between two genres, this approach can be easily expanded to perform multi-group classification. The extension appropriate for our study is called multinomial logistic regression, which is similar to binary logistic regression, with the exception of having J-1 equations instead of one, J being the number of categories encompassed in the model. This can be written in mathematical notation as:

$$l = log(\frac{\pi_{ij}}{\pi_{iJ}}) = \beta_0 + \mathbf{x_i}\boldsymbol{\beta_j}$$

where $\boldsymbol{\beta_j}$ is a vector of regression coefficients, similar to $\mathbf{x_i}$ being a vector of predictors. This produces $J - 1$ multinomial logit equations that contrast each of the categories, compared to binary logistic regression that contrasts between successes $Y = 1$ and failures $Y = 0$.

3.2 Support Vector Machine

The second classification technique is a supervised machine learning model, chosen appropriately as our data is completely labeled. SVM performs classification by generating one or multiple hyperplanes with p dimensions, p being the number of predictors included in the model. A function is intuitively set to divide the points between two classes, forming what is known as a separating hyperplane. Among the separating hyperplanes created, the one making the largest margin between the two classes is chosen as the optimal model and is used for predictions.

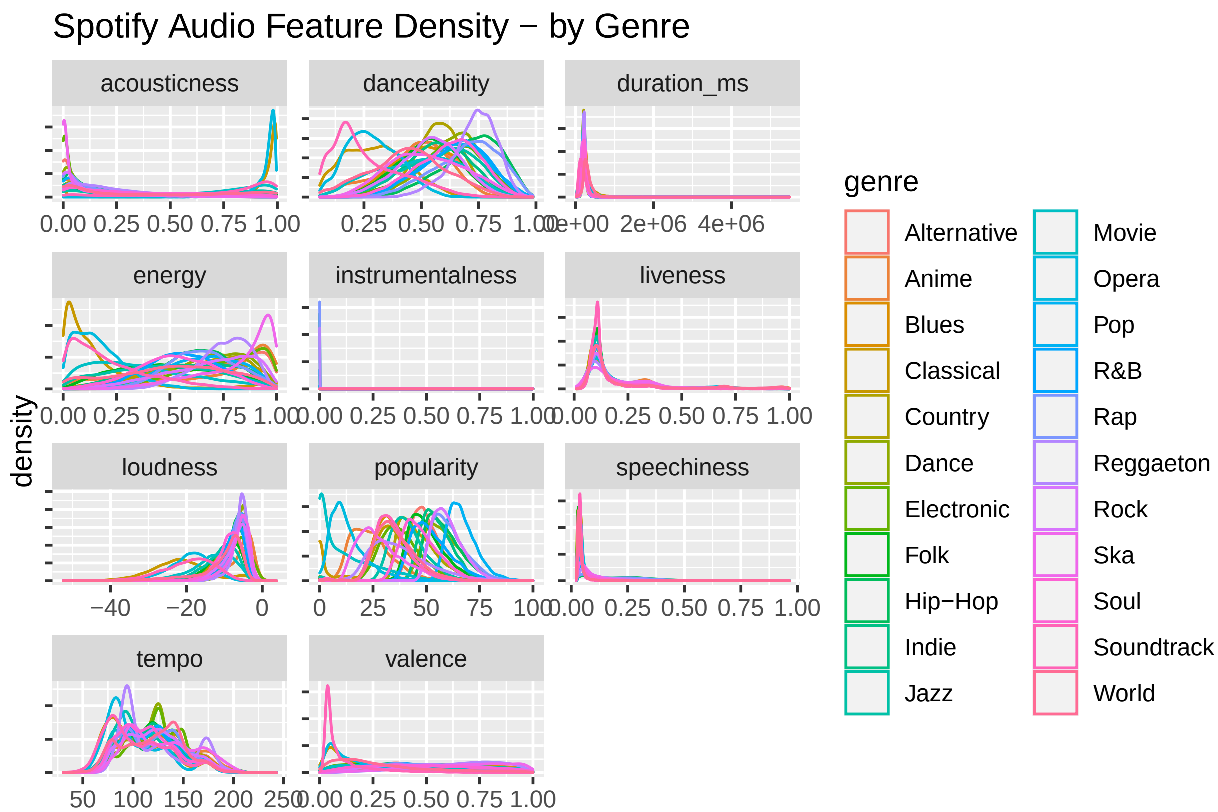
4 Results

– Results: Statistical analyses of the fitted model(s), and a translation of these findings into meaningful & understandable conclusions for the target audience (e.g., engineers, business managers, policy-makers, etc). See project rubric for details.

5 Conclusions

– Conclusion: A summary of key findings and potential impacts of your project

6 Appendix



6.1 Works Cited

[1] C.N. Silla Jr., A.L. Koerich, C.A.A. Kaestner A machine learning approach to automatic music genre classification J Braz Comput Soc, 14 (3) (2008)

[2]https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db