Lab 3 – MATH 240 – Computational Statistics

Henry Sun
Colgate University
Department of Mathematics
`hlsun@colgate.edu`

Abstract

This lab aims to answer the question of determining which band: Manchester Orchestra or The Front Bottoms, contributed more to the song "Allentown." We used Essentia and LIWC to conduct audio and lyrical analysis of song. Manchester Orchestra likely contributed more to the song than The Front Bottoms.

Keywords: Installing, loading, and learning to use libraries; working with characters objects; coding `for()` loops; accessing elements of vectors and lists.

1 Introduction

In 2018, two bands – The Front Bottoms and Manchester Orchestra – released a song they collaborated on called Allen Town. In a statement to Noisey (Ross, 2018) – the music arm of Vice – Andy Hull of Manchester Orchestra recalled that the creation of this track started when Nate Hussey of All Get Out sent him the first four lines of the track. Andy Hull worked out the melody and music and shared it with Brian Sella of The Front Bottoms, who then helped develop the chorus.

This brings us to an interesting question: which band contributed most to the song?

1.1 Week 1

In the first week of this lab, we built a batch file (`batfile.txt`) for data processing, containing all of the songs in the directory given to us.

We then processed a sample .json song, Au Revoir (Adios) in the Talon of the Hawk album by the Front Bottoms, and conducted some musical analysis on the song.

1.2 Week 2

In the second week of this lab, we continued to compile data for each song from their .JSON file, but we also pulled data from each song from the `EssentiaModelOutput.csv` and `LIWCOutput.csv` files. We merge all of the data into a data frame, and used it to create a `.csv` file for testing and training. We also used the Shiny app to make plots with our data.

2 Methods

Throughout this lab, we used Essentia (Bogdanov et al., 2013) to conduct musical analysis on each song, giving us a rough idea of what each band's songs "sound" like.

2.1 Week 1: Building a Batch File for Data Processing

For our first task, we first began by downloading the directory of songs onto the computer and downloading the `stringr` package (Wickham, 2023). By using the functions in the `stringr` package, we were able to identify which directories were albums by using the `str_count()` function. For each album, we used a similar process to count the occurrences of `.wav` in each file to identify which files were songs. By using a for loop through each album, we were able to create a vector called `code.to.process`, with each song being stored in the format as a `.wav` file, and the corresponding song output in a .JSON file. We used other `stringr` functions to extract `artist`, `album`, and `track`, from each directory or file and used `paste()` to create the .JSON file name. Finally, we used `writeLines()` to turn our vector of songs into a batch file.

2.2 Week 1: Processing JSON Output

Our second task was to process .JSON Output from a sample song, Au Revoir (Adios) on the Talon Of The Hawk album by The Front Bottoms. We first begin by installing the `jsonlite` package (Ooms, 2014), and repeating a similar process in our first task, by extracting the `artist`, `album`, and `track` from our sample file. We used `fromJSON()` function from the `jsonlite` to load the file in as a list, where we could then extract the average loudness, beats per minute (bpm), musical key, song length, and tempo from each song.

2.3 Week 2: Compile Data from Essentia

Our first task for the second week of this lab was nearly identical to the procedure followed in the second task from the first week. Using our previous example song (Au Revoir (Adios)), extracting additional data from each song, before repeating this process for each song in the EssentiaOutput folder, saving all the data into a data frame.

After loading into more data from each song from the `EssentiaModelOutput.csv` file, we extracted additional data by averaging values from certain data points. For example, we can get data on valence and arousal by average the data

from the DEAM, emo Music, and MuSe datasets. By running all the lyrics of each song through a text analysis tool called LIWC, we were able to gather data that describes the thoughts, feelings, and personality traits based on the language used. We passed all this data into yet another data frame. Finally, we took all the data collected in the previous steps, and used the `merge()` function to merge the data frames into one. The resulting dimensions of the data were by 181 by 140. We used this data set to create a testing `.csv` file, containing only the track "Allentown" and a training `.csv` file, containing all tracks except "Allentown."

3   Results

Using the Shiny app, I created a violin plot and a numerical summary of the data. The summary (Table 1) and the violin plot (Figure **??**) can be found in the appendix.

The mean of `relaxed` values for Manchester Orchestra and The Front Bottoms is 0.14 and 0.34, respectively. Standard deviation between all three bands was rather similar as well. Manchester Orchestra had overall lower `relaxed` values, having a smaller min and max values than The Front Bottoms. The max valye for Manchester Orchestra corresponds to the third quartile of The Front Bottoms.

4   Discussion

Overall, on average, this means that a song by Manchester Orchestra is much more relaxing than a song by The Front Bottoms. The `relaxed` value for "Allentown" is 0.16, which suggests that from evaluting each artist purely by the `relaxed` value, Manchester Orchestra contributed more to the song than the Front Bottoms.

The `relaxed` value was the most visibly distinguishing value between Manchester Orchestra and The Front Bottoms when using the violin plot. I found the inclusion of each song as a point on the chart to be helpful in visualizing the general trend of each band. The Front Bottoms are generally more relaxed than Manchester Orchestra, except for a few outliers, forming a small bulge in the violin plot left of most of the rest of the songs.

References

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., and Serra, X. (2013). Essentia: an open-source library for sound and music analysis. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, page 855–858, New York, NY, USA. Association for Computing Machinery.

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [stat.CO]*.

Ross, A. R. (2018). Manchester orchestra and the front bottoms are finally together on "allentown". https://www.vice.com/en/article/manchester-orchestra-and-the-front-bottoms-are-finally-together-on-allentown.

Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.

artist	Observations	Mean	Standard Deviation	Min	Q1	Median	Q3	Max	IQR
All Get Out	42	0.25	0.11	0.03	0.16	0.27	0.33	0.44	0.16
Manchester Orchestra	77	0.14	0.11	0.01	0.05	0.12	0.23	0.41	0.18
The Front Bottoms	61	0.34	0.09	0.10	0.31	0.35	0.41	0.47	0.10

Table 1: Numerical summary of the `relaxed` values for each song, sorted by artist