Lab 5 – 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, The Front Bottoms, or All Get Out 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 audio, while The Front Bottoms likely contributed more to the lyrics.

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 "Allentown". 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.

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. Using this data, we were able to conclude which band contributed the most to "Allentown."

In the first week of this lab, we built a batch file (batfile.txt) to process all the song in the directory given to us. We then extracted data from the .JSON Essentia Output, Au Revoir (Adios) in the Talon of the Hawk album by the Front Bottoms, and conducted some musical analysis on the song.

In the second week of this lab, we continued to compile data for each song from their .JSON file, but we also compiled data from the EssentiaModelOutput.csv, providing us with audio info, and LIWCOutput.csv, providing us with lyrical info. We merge all of the data into a data frame, and used it to create a .csv file for testing and training.

Finally, we used tidyverse (Wickham et al., 2019) to write a function that identified key features that would help us conclude which band contributed the most to "Allentown" with the aid of plots and tables generated by tidyverse and xtable (Dahl et al., 2019).

2 Methods

2.1 Week 1

After obtaining all of the .wav files for each song, we first began by building a batch file for data processing. Using the stringr package (Wickham, 2023), we were able to create the command line prompt for each track, before using writeLines() to write all the command lines into the batch file. Our second task was to process the .JSON Output from a sample song, which we accomplished by using the jsonlite package (Ooms, 2014). Using this, along with the stringr package, we were able to extract key information from the song, like artist, album, and average loudness. We were then able to use this same skeleton to extract information from every song's .JSON output.

2.2 Week 2

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, except that we collected data from the Essentia Model Data (Alonso-Jiménez et al., 2020). This provided us with more auditory information on each song.

Likewise, we used the LIWC-22 program (Boyd et al., 2022) to conduct lyrical analysis on each song. We were able to gather data that describes the thoughts, feelings, and personality traits based on the language used. 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.

We consolidated the data collected in the previous steps using the merge() function. 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."

2.3 Week 3

During our final week, we picked out and summarized key features from each band to determine which one contributed the most to "Allentown." We first began by calculating the upper and lower fences, and the minimum and maximum for

each numerical feature. We compared these values to the same features found in the "Allentown" track, keeping note of all features within range for one band. Using this data, we were able to create graphs and tables using `ggplot()` and `xtable()` to help us better visualize each feature of interest.

3 Results

In order to identify which band contributed the most to the song, we first began by determining which features in "Allentown" to analyze. A feature was classified as "Within Range" if the numeric value was between the upper and lower fence for the same feature in each band. This was narrowed down even further to 21 features having only one band in range. A table of all key features (Table 1) can be found in the appendix. Overall, we can see that there were no key features with only All Get Out being the only band. However, there were 3 features within range for The Front Bottoms and 18 for Manchester Orchestra. It's also crucial to note here that 3 out of the 4 lyrical features are within range for The Front Bottoms, while all of the audio features were within range for Manchester Orchestra.

Below is a box plot (Figure 1) of the only key feature that is within range of The Front Bottoms. `positivewords` is one of the three lyrical features (out of four total), along with `OtherP` and `conj` within range for The Front Bottoms. This should still be taken with a grain of salt, as `positivewords` only overlaps with the upper fence, and there are still outliers from both the other bands with more positive words than The Front Bottoms. The `positivewords` for The Front Bottoms is skewed further right, and has a significantly larger IQR than the other two bands.
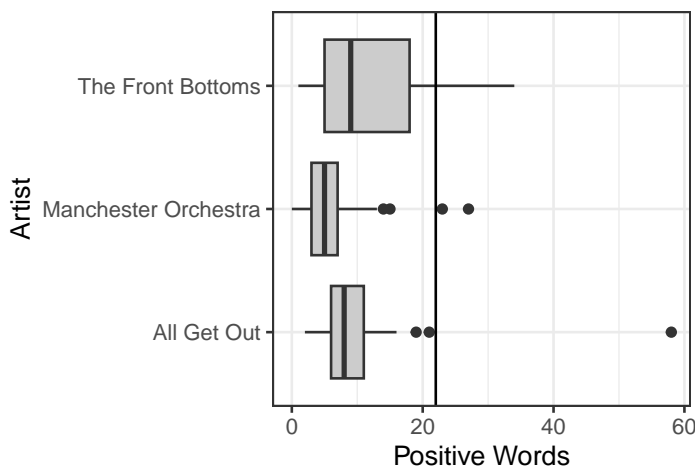

Figure 1: A boxplot of the `positivewords` feature for each band. The intercept represents the `positivewords` value for "Allentown."

The violin plot below (Figure 2) provides an example of a key feature that is clearly more characteristic of one band. The `average_loudness` of "Allentown" is within the IQR of Manchester Orchestra, while lying completely out of the

upper and lower fence the other bands. However, it is still key to note that the average loudness has a higher variability in Manchester Orchestra compared to the other two bands, which is characterized by its large IQR.
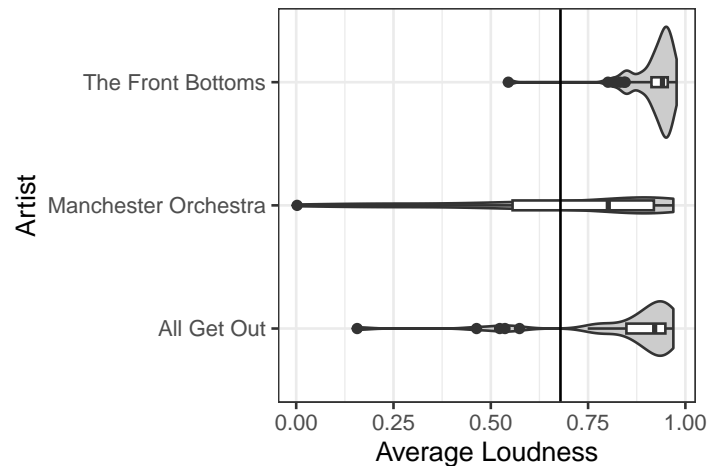

Figure 2: A violin plot of the average loudness for each band. The intercept represents the average loudness for "Allentown"

4 Discussion

Overall, the analysis of key features in "Allentown" suggest that Manchester Orchestra had the most significant contribution to "Allentown" compared to The Front Bottoms and All Get Out. Additionally, there is significant evidence that a lot of the audio features are similar to songs by Manchester Orchestra. To a lesser extent, The Front Bottoms seem to have a larger contribution to the lyrical features of the song, with All Get Out contributing very minimally to the song.

This matches up pretty closely to the description of how the song was created in the introduction. The first four lines of the track were created by All Get Out, while the rest of the melody and music was finished by Manchester Orchestra, matching up with Manchester Orchestra's contribution to the audio features. Meanwhile, The Front Bottoms developed the chorus, which matches up with the limited contribution to some lyrical features.

Throughout this lab, we've learned how to process and analyze audio and lyrical data, in order to determine the contribution of one artist in a collaborative song or album. This lab has showed us that it's key to make a distinction between audio and lyrical data during analysis. The analysis of data can be quite subjective, as there are numerous ways to filter out data and create tables or plots, which may yield different conclusions.

References

Alonso-Jiménez, P., Bogdanov, D., Pons, J., and Serra, X. (2020). Tensorflow audio models in essentia. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 266–270. IEEE.

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.

Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. (2022). The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10:1–47.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

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.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

5 Appendix

Key Features	All Get Out	Manchester Orchestra	The Front Bottoms
spectral_skewness	Outlying	Within Range	Out of Range
spectral_rolloff	Out of Range	Within Range	Out of Range
spectral_kurtosis	Outlying	Within Range	Out of Range
spectral_entropy	Outlying	Within Range	Out of Range
spectral_energyband_middle_high	Out of Range	Within Range	Out of Range
spectral_complexity	Out of Range	Within Range	Out of Range
spectral_centroid	Out of Range	Within Range	Out of Range
melbands_spread	Out of Range	Within Range	Out of Range
melbands_flatness_db	Out of Range	Within Range	Out of Range
erbbands_skewness	Out of Range	Within Range	Out of Range
erbbands_flatness_db	Outlying	Within Range	Out of Range
dissonance	Outlying	Within Range	Out of Range
barkbands_skewness	Out of Range	Within Range	Out of Range
barkbands_kurtosis	Out of Range	Within Range	Out of Range
barkbands_flatness_db	Outlying	Within Range	Out of Range
average_loudness	Outlying	Within Range	Outlying
chords_strength	Outlying	Within Range	Out of Range
conj	Out of Range	Outlying	Within Range
Perception	Out of Range	Within Range	Out of Range
OtherP	Outlying	Outlying	Within Range
positivewords	Outlying	Outlying	Within Range

Table 1: A summary of all key features (features only within range for one band)