Acoustic Coherence: Similarity of Musical Genre Through Time

The question that I’m interested in answering is inspired by Ted Underwood’s *The Life Cycle of Genres*. In the article, Underwood explores the life cycle of the genre in the literary domain. More specifically, he is looking at how a few different genres—detective fiction, the gothic, and science fiction—to determine the lifespan of genres. To do this, underwood trains a model to predict a certain genre of literature and traces how well his model can predict certain novel’s probability of being from that genre over time and forming some sort of textual coherence between generations (Underwood, 2016). Instead of looking for textual coherence, I will be exploring the auditory coherence of audio characteristics through time to understand whether music conforms to genre in a similar way as it does in literature. That is, does musical genre exhibit coherency over time or do they go through life cycles of death and rebirth? I will be considering the genre of “rock” music because, as Johan Fornäs suggests of rock music in *The Future of Rock,* “the musical generic system is spun like a web of aesthetic rules undissolvably tied to social and psychic factors” (Fornäs, 1995). It’s this elegant web that I believe will derive the most interesting insights.

My first step in conducting the analysis was collecting my data. I chose to use the Spotify API to gather data. I opted for this method over other premade datasets for a couple of reasons. First, and most importantly, I wanted to analyze and compare the actual sound of songs. By sound, I’m referring to audio terms like tempo, pitch, loudness, and timbre. These were made easiest to find using the Spotify API. Most other datasets found online contained only song meta-data which only included things like the name of the artist or song. The next reason for choosing the Spotify data was the recency that it allowed me to analyze songs from. Since the API can access songs and data from Spotify that is are most up to date, it saves me from only being able to analyze data up to a certain year that a premade dataset was created. For example, the Million Song Dataset presented itself early on as an enticing data option. That is until I found that the latest music in the dataset reached until 2012. The Spotify data also had labels for genre already given, meaning I wouldn’t have to manually label the dataset.

The way that I chose values for this dataset was by randomly sampling the songs from the top 50 artists of a particular genre based on monthly listening. To run a classification algorithm, I need more genres than just rock. Since data collection takes some time, I randomly sampled three other genres (polish, post-punk, and gospel) and conducted the same sampling techniques as I did for rock to sample songs from the top 50 artists in those genres, respectively.

One caveat I must make, however, is that these genres have been assigned using the Echo Nest API. The Echo Nest is a subsidiary company of Spotify that specializes in music intelligence. Their methods for determining genre are not described, but likely calculated using similar or the same data that I have access to. While the chances that I uncover the method and algorithm which Spotify uses to classify genre are low, it’s worth mentioning, since there is a non-zero chance that I have the data and methods available to me to produce the perfect prediction.

As I mention briefly above, I will be working with rock music as my main genre of interest. Not only do I choose this genre because of *The Future of Rock* by Fornäs, but also because I think there’s some truth in the idea that the average user’s conception of rock music has changed. When bands like The Beatles and Green Day are garnering the same genre label, there’s an evident case to be made that broad genres like rock inevitably go through changes.

The next important aspect of answering my research question is the method itself. Since I was doing classification, my pool of statistical learning techniques needed to be limited as such. I next wanted a method which would give me predicted probabilities, rather than just outcomes. This so that I could create a plot much like Figure 2 in Underwood’s paper (Underwood, 2016). This meant the use of logistic regression which can give predicted outcomes in terms of probabilities. This does mean I will need to satisfy two main parametric conditions: no multicollinearity between my predictors and a linear relationship between my predictors and the log odds of my outcome. The variables that I used were the loudness, tempo, and release year of the song. Since tempo and loudness were both recorded as continuous variables over a particular song, the predictor that I used for any song was the weighted average either of these values weighted by their duration in the song. For example, if a song had measured a loudness of -60 decibels for one quarter of the song and -30 decibels for the remainder of the song, the value for loudness of that song would be decibels.

As mentioned above, the method that I have chosen to implement is a multiple logistic regression with predictors tempo and loudness and the outcome of a particular song being rock music. The results of this model are in Table 1. The VIFs for these three variables are 1.03, 1.01, and 1.03, respectively. Overall, the model found that loudness and tempo were both predictive of rock music. To interpret these coefficients, one must exponentiate that value ( to see what it

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | z-value | p-value |
| Intercept | 54.37 | 13.27 | 4.10 | 0.000042 \* |
| loudness | 0.034 | 0.15 | 2.37 | 0.018 \* |
| tempo | -0.027 | 0.0066 | -4.01 | 0.000057 \* |
| year | -0.002 | 0.0026 | -0.76 | 0.45 |

Table 1: Results from Logistic Regression Model

it does to the relationship with the odds ratio of the outcome. When doing this, we find that a one decibel increase in loudness increases the odds of being rock music by a factor of 1.03.

Chart, scatter chart

Description automatically generated

Figure 1: Predicted probability of being rock music by year of release of a song. The predicted probabilities are fit using the logistic regression model with year, tempo, and loudness as predictors.

Similarly, a one unit increase in beats per minute in a song decreases the odds of being rock music by a factor of 0.97. The year variable was found to be insignificant in the model. The model was optimized on accuracy, with a maximum accuracy of 59.45%. The cut point in the logistic model for this output was 0.497.

Figure 1 and Figure 2 display some results from the model. In Figure 1, we’re able to see that the model predicts that songs that are older are more likely to be rock music. It does a good job at distinguishing these rock songs from other genres like gospel and polish, but less well at distinguishing rock from post-punk. By Figure 2, we can see that as loudness increases, certain songs are more likely to be classified as rock. Again, out of the points that have been classified with prediction of being rock over 0.7, there seems be a mix of rock and post-punk. Both of these observations in either figure could hint at some audio coherence between different genres, namely rock and post-punk. Post-punk is a genre which includes artists such as The Smiths and The Cure. Depending on who you ask, these artists may be classified by listeners as being themselves rock artists. While my method was able to give me a plot much like the one that I had hoped to produce, I clearly didn’t achieve the decision boundary that I had hoped for. Figure 1

Chart, scatter chart

Description automatically generated

Figure 2: Predicted probability of being rock music by loudness of a song. The predicted probabilities are fit using the logistic regression model with year, tempo, and loudness as predictors.

offers some clarity in terms of what is predictive of rock music, but there are still many songs that are misclassified.

While I was able to run a model I deem to be viable, there is still room for improvement. The Spotify API and my data analysis can be adjusted to include variables that I think will be more important like pitch and timbre but will need even more time working with data cleaning before implementation. The method for including these will be like what it was for loudness and tempo, a weighted average over the durations that the pitch and tempo vectors are observed over. Including just loudness and tempo in the model couldn’t produce a figure which gives me much information on how I can trace genre through time. I believe that using these two new variables as well as sampling from a larger pool of genres will improve the model and be able to answer the question of more acutely of whether musical genre can be traced through time through acoustic coherence between generations.

References

<https://github.com/r-coburn/DA-401>

Underwood, T. (2016). The life cycles of genres.

Fornäs, J. (1995). The future of rock: discourses that struggle to define a genre. *Popular music*, *14*(1), 111-125.