Identifying Musical Similarities Across Geographical Regions

Eric Su, Andy Cooper, Michael Valancius

April 2019

Abstract

The purpose of this project is to analyze musical similarities across geographical origins. Audio features were extracted from 1,059 songs originating from 73 countries around the world. We performed dimension reduction on these features using PCA, t-SNE, and an autoencoder. We then clustered on the reduced dimensions from the autoencoder using a 7-component Gaussian Mixture Model. Cluster analysis shows that smaller, island-like regions tend to have more isolated musical identities, while larger countries closer to trade routes tend to have less distinct musical identities.

1 Introduction

Music is a major aspect of most world cultures. Different areas of the world develop their own unique musical styles. However, as cultures mix, so do aspects of music composition. It is unclear how and to what extent aspects of music are shared across the world. This begs the question: can we use clustering techniques to identify musical similarities across geographical regions?

2 Related Work

The area of ethnomusicology relates to studying the cultural aspects of music. Previous researc h has been conducted on the nature of music across the world, its social functions, and it's cultural impact. Some research has attempted to capture musical differences across cultures quantitatively.

In particular, this project is based on the paper "Predicting the Geographical Origin of Music" (Zhao, Q, King, 2014). The paper examines whether audio features can be used to predict the geographical coordinates of a song's country of origin. The paper compares the performance of K-nearest neighbors and random forest regression with bagging. They found that there is indeed at least some predictive power in a song's audio features, indicating that music has qualities unique to geographical origins. In addition, they found that bagged random forest proved to be best of the tested methodologies at predicting the origin of a song.

3 Methods

3.1 Dataset

The dataset used for our analysis comes from the paper "Predicting the Geographical Origin of Music" (Zhao, Q, King, 2014). In their paper they chose to analyze tracks from available CD's. Using the CD sleeves they traced the origin of each track. The dataset has 1,059 songs in total originating from 73 unique countries.

Each song was processed through a software called "MARYSAS" ("Music Analysis, Retrieval and Synthesis for Audio Signals"). MARYSAS is an open-source software program that extracts several features from audio files. The extracted features are unitless numerical quantities. 68 different audio features were extracted from each song using the software. Additional chromatic features were also extracted from each song.

3.2 Dimension Reduction

The full dataset consists of both the audio and chromatic attributes, leading to 116 features in total. Furthermore, understanding the relationships amongst signal features poses enormous challenges, eliminating the possibility of individually evaluating and selecting features. That is, variable elimination strategies are difficult to assess when the meaning of the predictors is unclear. However, it is quite possible that much of the information, perhaps including information that is pertinent to understanding the geographical similarities in musical composition, resides in a lower dimensional space. This motivates our usage of dimension reduction: we hope to extract lower dimensional information from the data that best helps explain core musical attributes that might be shared across geographical regions.

Without specific knowledge about the individual features, there existed little reason to assume one dimensionality technique would outperform the others. Thus, we decided to pursue three different techniques that accomplish the task through different approaches: principal components analysis, t-distributed stochastic neighbor embedding, and an autoencoder.

3.2.1 Principal Components Analysis

In principal components analysis, we take an orthogonal projection of the data onto a lower dimensional linear subspace, simultaneously maximizing the variance of the projected data. It is the most commonly used linear method, taking a linear combination of the original data to define the new data.

3.2.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)

By using t-SNE, we hoped that the probabilistic approach that tries to find a mapping that represents similar data points closer together might better handle the complex relationships among the predictors. For visual purposes, the data was reduced to 2 dimensions and a perplexity parameter of 30 was chosen after tuning.

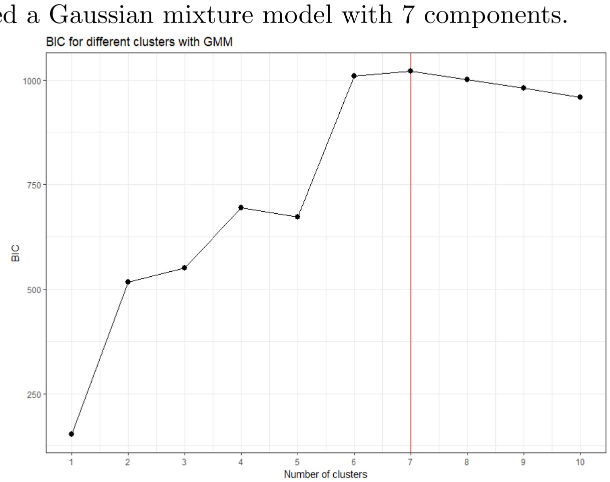
3.2.3 Autoencoder

Another frequently used method in dimension reduction takes advantage of a neural network architecture. In an autoencoder, the original data is encoded into a hidden layer and then decoded to its original dimension, with a cost penalty based on the distortion. When the activation functions are linear, this bears a close resemblance to PCA. Our symmetric architecture consisted of three hidden layers (projecting the data down from 116 features to 6 and then to 2) with hyperbolic tangent activation functions.

3.3 Gaussian Mixture Model

After identifying a lower dimensional structure from the autoencoder, we wanted to cluster data points together and identify regions with similar music. Since the size of clusters may vary, we decided to use a Gaussian mixture model to fit our data. A Gaussian mixture model uses a mixture of normal distributions to model data and estimates the mean and variance of each component.

We choose the number of clusters based on BIC. Below is a plot showing the BIC for differing numbers of clusters. Since BIC is maximized at 7 clusters, we used a Gaussian mixture model with 7 components.

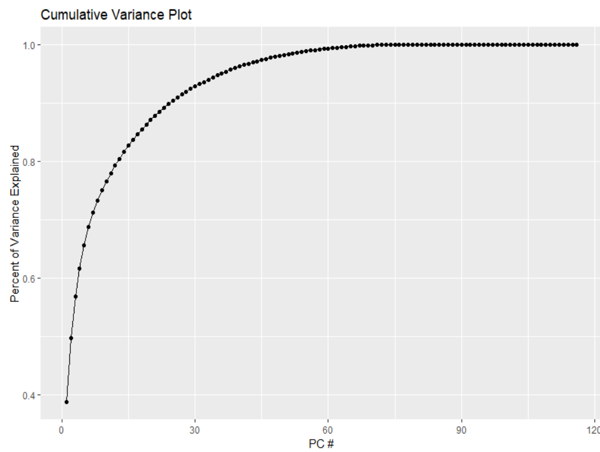


4 Results

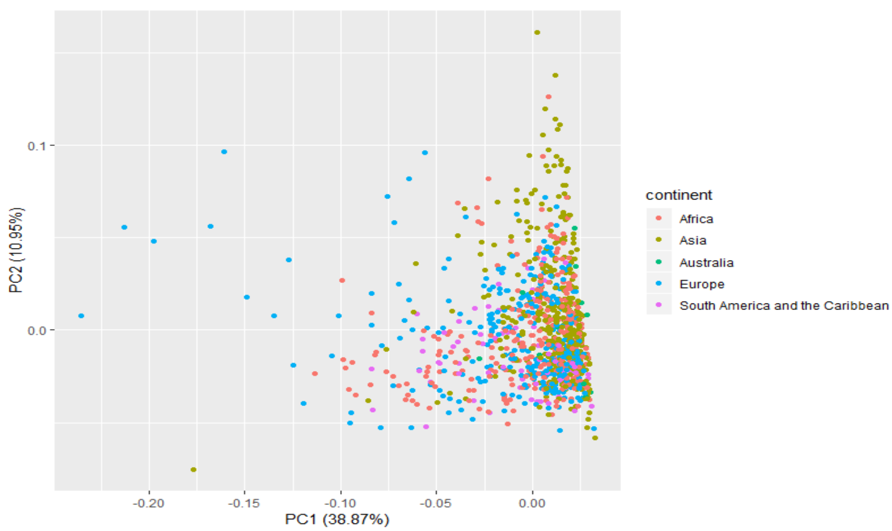
4.1 Dimension Reduction

4.1.1 PCA

Our PCA analysis revealed that the first 55 principal components capture 99% of the variation in the original dataset and the first 10 capture 76%.

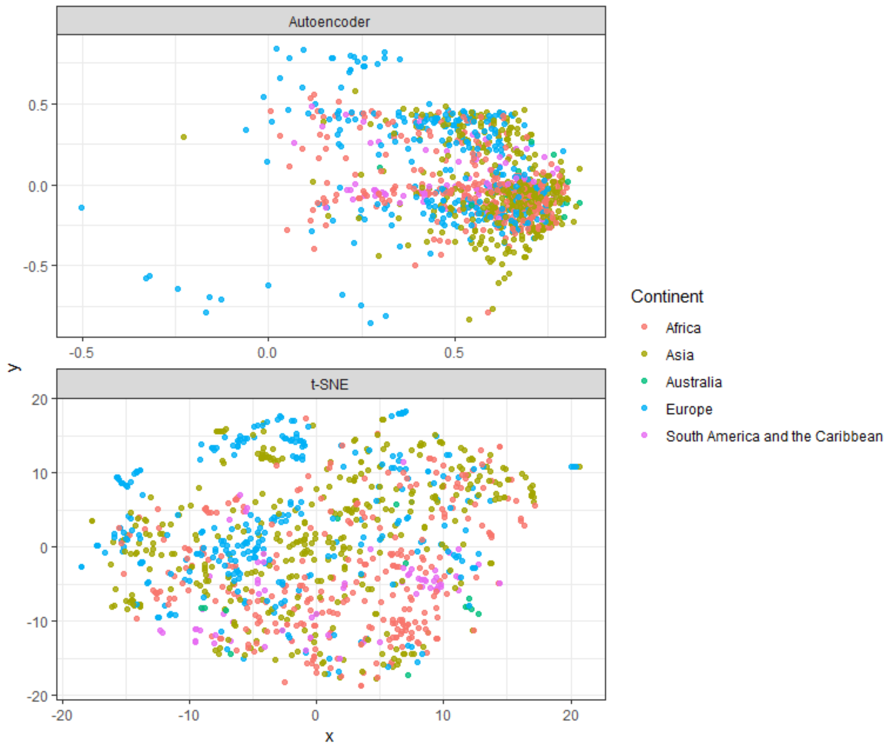


The following plot visually represents the relationship of geographic origin with the first two components, While there are some distinct patterns (particularly with Europe), there are not quite the distinct clusters in the data we hoped to find.



4.1.2 t-SNE

Visual representation unveils that some clusters are beginning to form, though they are somewhat heterogeneous in continent origin.

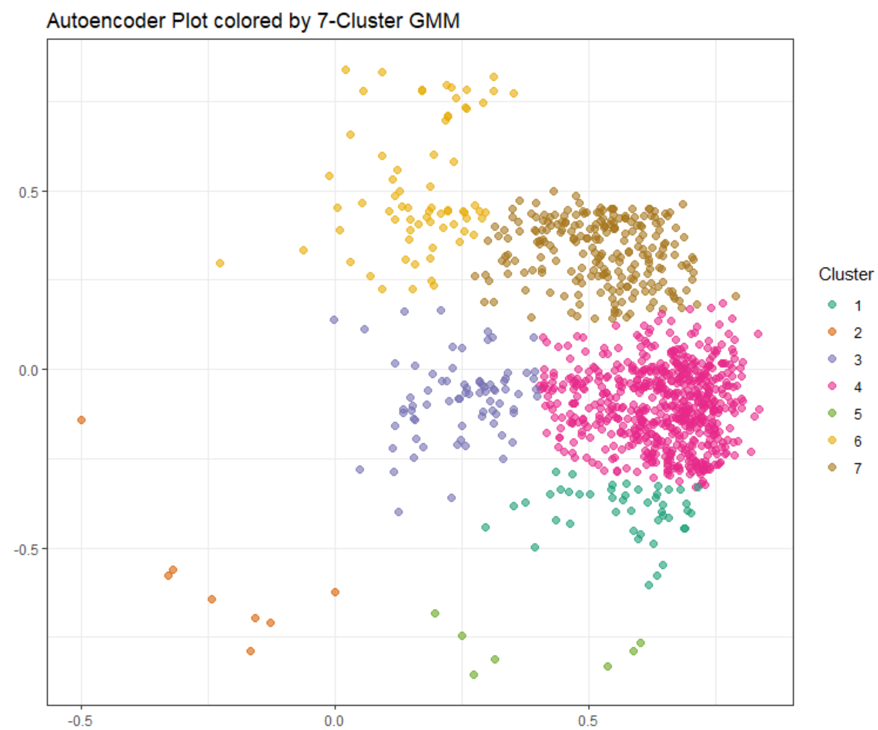


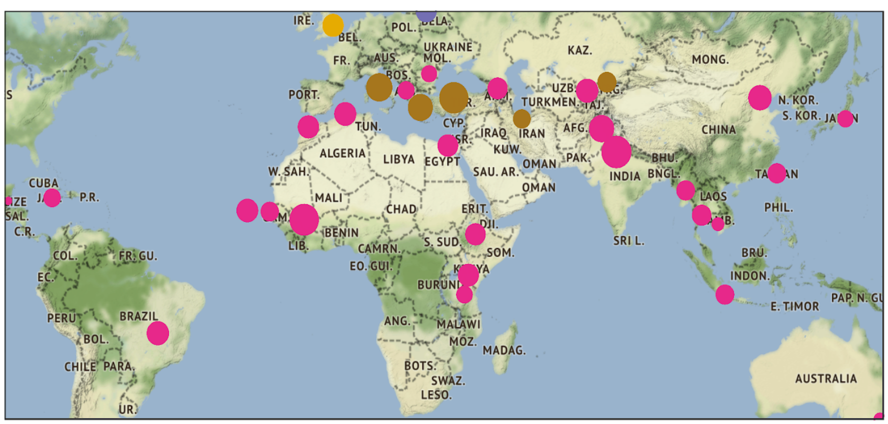
4.1.3 Autoencoder

The hidden layers of an autoencoder provided a reduced dimension upon which the data could be clustered. While the number of hidden layers were referred considered as a parameter to be tuned, given that clustering is oftentimes an ill-defined problem, we settled on a hidden layer with two units. As seen in the visualization, the most distinct clusters formulate through this dimension reduction method.

4.2 Clustering

We colored the data points by the clusters they were assigned to in the reduced dimensions from the autoencoder, as well as in the world map.





Points colored by most frequent cluster classification

One can see that clusters 4 and 7 represent more dominant music styles, clusters 3 and 6 represent music styles more common in specific countries and small clusters such as 2 and 5 may be less common musical styles which are not

Table 1: Countries with Highest Purity

Country	Purity
Japan	0.95
Australia	0.86
Cambodia	0.86
Algeria	0.83
Taiwan	0.80

Table 2: Countries with Lowest Purity

Country	Purity
Iran	0.48
Morocco	0.45
Italy	0.43
Turkey	0.41
Lithuania	0.28

country specific.

5 Discussion

5.1 Country "Purity"

We introduce a metric named "Purity", which is the percentage of data belonging to the dominant cluster in each country. Higher purity level indicates a more distinct, singular music style, while low purity suggests a more diverse music composition. Table 1 and 2 show the top five countries with the highest and lowest purity respectively.

The purity table shows countries like Japan and Cambodia have the most distinct clusters. Small, island-like countries tend to have the most distinct musical identities, perhaps because of the nature of their isolated geography. Conversely, countries close to European trade routes tend to have the least distinct music identities, perhaps because aspects like music can be shared more easily.

For future work, we consider the idea of audio features being used to classify the continent of origin for a piece of music. Can these audio features alone accurately predict where a song originates from? Further investigation would determine is a powerful function approximator, such as a neural network, can be used to predict country of origin based on musical features.

6 Conclusions

In this paper, we hoped to find unifying, perceptible themes in the musical data that would help characterize the relationship between musical composition and geographic origin. Given that this was an unsupervised problem with no clear objective to optimize, developing insights into this relationship required careful thought into the techniques we used as well as interpreting the generated results. Through combining dimensionality techniques with clustering methods, we aimed to extract fundamental information about the structure of the music and in order to then investigate ties between geography and musical similarities. When carrying out this analysis with an autoencoder and a Gaussian Mixture Model, we identified patterns that demonstrate a positive relationship between the homogeneity of music and the country's geographic and socioeconomic isolation, suggesting that there are learnable features within music that can help distinguish a song's origin.

7 Bibliography

F. Zhou, Q. Claire and R. D. King, "Predicting the Geographical Origin of Music," 2014 IEEE International Conference on Data Mining, Shenzhen, 2014, pp. 1115-1120. doi: 10.1109/ICDM.2014.73