An Inquiry into the Geographical Origin of Music

Armie Marie A. Aparri
Mike Christian B. Gragasin
Renzo Roel P. Tan
Ateneo de Manila University

1 Summary

In brief, the paper investigates the geographical origin of music through a clustering and classification approach. The *k*-means clustering method and the *k*-nearest neighbors and naive Bayes classifiers were used. Data containing audio features extracted from wave files was processed towards justifying a link between the songs and their source location. Apart from the notable insights, such an analysis resulted in the discovery of some deficiencies. The problems encountered are adequately treated in the discussion.

2 Introduction

While mostly qualitative, the relationship between music and its geographical origin has been the subject of many researches. The discussion in [3] makes mention of music being influenced by mostly geographical factors – politics, social order, and culture. Geography provides the particular space that helps shape music [4].

In line with the above, the study endeavors to present a computational perspective on the matter. The data set[1] by F. Zhou *et. al.* based on 1059 music tracks from 33 geographical areas is employed for the exploration. Each instance in the collection has 68 features representing different audio characteristics. These include the spectral centroid [7], spectral moments [1], pitch [6], harmonicity [**?**], the mel-frequency cepstral coefficients [2], and others, all of which are numerical. The audio features were extracted using MARSYAS [8], a known framework for audio analysis proposed by G. Tzanetakis and P. Cook.

As an aside, the data is accompanied by a paper [10] written by the same contributors. The paper utilized the *k*-nearest neighbors classification and the random forest regression, concluding that the latter performed better than the former. The accuracy of the models were not given; only a measure of distance error was indicated.

3 Objective

In accord with the earlier explication, the paper aims to:

1. Verify the results of the original paper through a similar implementation of the *k*-nearest neighbor classification; and

2. Propose a classification setup that yields an error falling approximately in the range of the determined errors.

[1] This is found in `http://archive.ics.uci.edu/ml/datasets/geographical+original+of+music`.

4 Methodology

The location of the track given in longitude and latitude is set as target. Upon further inspection, there exists only 33 unique combinations that stand for 33 different countries; hence, one relabels for convenience.

The conventional k-means clustering is first performed to check whether some similarity is present between the baseline and the resulting clusters. For this, $k = 33$ is optimistically set irrespective of the scores for k values corresponding to the 33 countries. In addition, a clustering with $k = 5$ is intuitively done as there are 5 continents of concern [10]. Both were done twice – with and without principal component analysis. For the one without, the visualization was based on the location. One plots the points using the given coordinates and simply inspects whether those who are geographically near form clusters.

Segueing into classification, the k-nearest neighbors classification was executed from $k = 1$ to 30. The accuracy and distance error figures were recorded. Lastly, the naive Bayes classifier was used for comparison.

5 Discussion

In the preliminary clustering experiment, all runs showed that there is little to no similarity between the resulting clusters and the actual geographical location of the tracks. For $k = 33$, music tracks belonging to a single country were put in different clusters; for $k = 5$, clusters were formed by tracks from countries not necessarily from the same continent. As the features were unidentified, justifying such a result is difficult. It is possible that the MARSYAS framework extracts audio features that are not strongly linked to geographical attributes.

Classification is then done through k-nearest neighbors. Among the scores for $k = 1$ to 30, $k = 5$ emerged the best. This gave an accuracy of 38.67% and an average distance error of 3788.98km. When set side by side with the basis paper results, the difference is not large. Results in [10] include an average distance error of 3,113.39km when $k = 5$. This is far from the acceptable threshold. The naive Bayes classifier was no different. The customary implementation had 35.22% accuracy and 3763.65km average distance error.

One proposes a hybrid technique in the hopes of reaching higher accuracy and smaller error. The k-nearest neighbors naive bayes classification is proposed. This involves:

- Identifying the k-nearest neighbors; and
- Computing for the probabilities using only the obtained k neighbors.

For $k = 5$, the accuracy was pulled up to 41.51% and the average distance error fell to 1652.50km, over-all better than the previous trials and [10].

6 Conclusion

To close, a larger corpus may be helpful for better accuracy. With 68 features and only 1059 instances, there is a high risk of model overfitting. Moreover, the identification of audio features and the pre-normalized data set may give valuable insights prior to classification.

References

[1] Fujinaga, I. (1998). Machine Recognition of Timbre Using Steady-State Tone of Acoustic Instruments. In *Proceedings of the ICMC*.

[2] Hunt, M. *et. al.* (1996). Experiments in Syllable-Based Recognition of Continuous Speech. In *Proceedings of the ICASSP*.

[3] Kearney, D. (2010). Listening for Geography: The Relationship Between Music and Geography. In *Chimera*.

[4] Leyshon, A. *et. al.* (1998). *The Place of Music*. Guilford Press, New York.

[5] Nash, P. (1968). Music Regions and Regional Music. In *Deccan Geography*.

[6] Rabiner, L. *et. al.* (1976). A Comparative Performance Study of Several Pitch Detection Algorithms. In *IEEE Transactions on Acoustics, Speech, and Signal Processing*.

[7] Scheirer, E. and M. Slaney (1997). Construction and Evaluation of a Robust Multifeature Speech or Music Discriminator. In *IEEE Transactions on Acoustics, Speech, and Signal Processing*.

[8] Tzanetakis, G. and P. Cook (2002). MARSYAS: A Framework for Audio Analysis. In *Organized Sound*.

[9] Wold, E. *et. al.* (1996). Content-Based Classification, Search, and Retrieval of Audio. In *IEEE Multimedia*.

[10] Zhou, F. *et. al.* (2014). Predicting the Geographical Origin of Music. In *IEEE International Conference on Data Mining*.