Predicting the Geographical Origin of Music

Arjun Kumar
Department of Computer Science and Electrical Engineering, UMBC
arkumar1@umbc.edu

Abstract

Contemporary to the human judgment in classifying music, a few machine learning approaches have been presented in this paper. Geographical ethnomusicology: dispersion of music around the globe, is the problem which has been addressed. The dataset along with the spherical symmetry of the globe, makes it interesting to approach the problem using the geometric operations on the dataset. ε-KNN and K-Means clustering have been applied to correctly predict the geographical region from where a given music track originates.

Introduction

The study of music in its cultural context has based on the subjective judgment of the critics. Modeling the areas of ethnomusicology into a problem that can be solved using data mining principles holds a great promise. There are many different types of music spread across the globe. They originated as a result of cultural, geographical and philosophical factors, and often relate to or reflect onto the society of the region of their origin. Analyzing the geographic origin of music is a good way to understand the process of studying the cultural aspects of it.

Geographical ethnomusicology is the study of how music is distributed across the world. Over the time, with globalization, various forms of music have influenced each other. Many, if not all, music forms have travelled long distances across the globe from their origin and thus regional music is rarely pure. This makes the problem of geographical ethnomusicology both a complex and an interesting one.

Problem Definition

This project intends to train a machine to correctly predict the geographic origin of a given music track. The dataset used consists of 1059 instances each with 68 auditory features extracted using the MARSYAS program from the wave file of the corresponding sound tracks originating from 33 different regions.

The aim of this project is to establish a relationship between these data inputs to predict the geographic origin of a track.

Special Structure of the Data

Label

The labels of the instances consist of longitude and the latitude of the political capital of the geographic region from which the track originated.

- **Latitude:** Latitude values range from -90◦ to +90◦ representing the southern and the northern hemisphere of the globe respectively.
- **Longitude:** Longitude values range from -180◦ westward to +180◦ eastward of the prime meridian (which is represented by 0◦).

Instances

The instances are classified as 'traditional', 'ethnic' or 'world only'. These classifications are taken as marked on the album by the publishers. If the classification is not provided, the track is classified based on the place from where the artist of the track originates.

Western music tracks are not included in the dataset as their effect is global and hence would make the training of the learning agent regarding the geographic origin of the track based on the features a difficult task.

Features

The dataset consists of two files

- **default_features_1059_tracks.txt:** This file consists of default auditory features extracted from the tracks by MARSYAS.
- **default_plus_chromatic_features_1059_tracks.txt** This file consists of the default auditory plus 12 Octave-Western tuning chromatic features that describe the scale of the notes used in the tracks.

Proposed Methods

The classification task was handled using ε-K Nearest Neighbors to predict the originating region of the track provided.

Clustering was also implemented using K Means Clustering algorithm to form clusters of tracks originating from the same region based on the auditory features provided for the instances.

Intuition

Since the label consists of latitude and longitude, the dataset follows a natural ordering of clustering the instances into various clusters represented by the latitude and longitude of the cluster centroid. It can be hoped that, based on the auditory features, K-Means clustering will form clusters of tracks originating from the same geographic region in the training phase and will also be able to correctly assign a cluster to a test instance.

As the corpus includes the geographic location of the origin of the tracks, it seems that implementing the geometric methods to classify the data using KNN (or some variation of it), one would be able to train the machine to classify the music tracks based in their origin correctly to a large extent.

Methods

In this project, sphere has been chosen as approximate representation of the globe [1]. Predicting the spherical coordinates of a point on the Earth is complex because of the following reasons:

1. The area of the grid represented by latitude and longitude lines is non-linear. The grids near the equator have more area compared to those near the poles. This makes it difficult to calculate the error in the distance between the predicted and the marked label of an instance.
2. The line opposite to the Prime Meridian on the globe has two values ±180◦. Hence, two geographical locations which might be very near to each other may have a large difference in their longitude values, one negative and the other positive.

Result Calculations

The performance of an algorithm is calculated in terms of the mean distance error between the predicted and the actual label among all the instances (test or train). The distance is calculated in terms of Kilometers using the following method presented in [1].

A music track is represented as a two-tuple $S_i = (F_i, L_i)$ where F_i represents the feature vector $F_i = (f_i^1, f_i^2, ..., f_i^n)$ and $L_i = (\varphi_i, \lambda_i)$ represent the corresponding label, where φ_i represent the latitude and λ_i represent the longitude of the class label.

The great circle distance from the true position L_i^{Te} and the predicted position L_i^{P} is calculated as follows:

$$d(L_i^{Te}, L_i^{P}) = 2 * R * \arctan2(\sqrt{a}, \sqrt{1-a}), \text{ where}$$

$$a = \sin^2\left(\frac{\varphi_j^P - \varphi_j^{Te}}{2}\right) + \cos\varphi_j^P \cos\varphi_j^P \sin^2\left(\frac{\lambda_j^P - \lambda_j^{Te}}{2}\right)$$

and R = 6373.

This great circle distance represents the error in prediction.

ε-K Nearest Neighbors

The K Nearest Neighbors algorithm is based on the inductive bias that in the n-dimensional feature space (where n represents the number of attributes in an instance) the label of an instance should be similar to that of the nearby points.

The training phase of the algorithm consists of just storing all the training examples in the memory of the classifying agent. The test phase for the primitive algorithm is implemented in the following steps:

1. Calculate the n-dimensional Euclidean distance between the test point $\hat{z}$ and all the training points z_i.
2. Poll the K closest points z_i for their labels.
3. The majority label among these K points is predicted as the label of the test point $\hat{z}$.

In the ε-KNN extension of this algorithm, step 2 of the test phase is modified wherein now only the training points z_i that are at a Euclidean distance ε from the test point $\hat{z}$ are poled for their class labels and a majority among these is predicted as the class label for the test instance.

In this project, ε-KNN has been implemented in the following manner:

1. Store all the training instances in the memory.
2. For all test points repeat steps 3-5:
4. Calculate the n-dimensional Euclidean distance between the test point $\hat{z}$ and all the training points z_i.
3. Poll all the training instances within a distance of ε of the test point.
4. Find the majority label among these points.
5. Calculate the great circle distance between this label and the true label of the class.
6. Find the mean of the great circle distance calculated for all the test instances.
7. Repeat step 2-6 for values 6, 6.5, 7, 7.5, 8, 8.5, 9 of ε.

K Means Clustering

K Means Clustering is an unsupervised learning algorithm and has been implemented in this project in the following way.

Training Phase

1. Initialize the clusters centroids to some random training instances.
2. For each training instance, calculate its n-dimensional Euclidean distance with each of the cluster centroid.
3. Assign the training instance to the cluster whose centroid is closest to it according to the distances calculated in step 2.
4. When all the training samples have been assigned to a cluster, recalculate the cluster centroid by assigning the centroid to the mean of all the members of the cluster.
5. Repeat step 2-4 till the cluster centroids stop changing.

Test Phase

1. For each test instance, calculate the n-dimensional Euclidean distance between it and the centroid of each of the cluster.
2. Predict the label of the test instance as that of the centroid of the cluster which is closest to this test point based on the distance calculated in the precious step.
3. Calculate the great circle distance between the predicted label and the true class label provided in the dataset.
4. After the label for all the test points have been predicted, calculate the mean error in the great circle distance obtained for each of them.

The test and the training phase is repeated using only 20%, 40%, 60%, 80% and 100% of the features of the instances for comparison and the results are reported.

Conclusion

- Some audio features have more effect than others in deciding which region the song belongs to. As evident from the results of K Means Clustering where the least error distance was obtained at 60% and 80% of feature usage for default and chromatic plus default features respectively.
- Different error distance is observed for different values of ε which suggests that a music track originating from one region does influence, or bears similarity to, a music track that originates from another region. How close these regions need to be for the affect is not clear from the results.
- K Means Clustering performs better than ε-K Nearest Neighbor for the given dataset for both default features as well as default plus chromatic features.
- The label of the instances is taken as the latitude and longitude of the political capital of a country. The political capital might not always be the cultural hub of a nation. Moreover, two music tracks originating from the same country might be different in styles based in two different cultural regions present in the same political boundary.
- This along with the fact that when no classification is explicitly provided by the producers of a music track, the location of the region from where the artist originates is taken as the label, might have added noise in the data and can make the prediction inaccurate.
- More powerful clustering methods such as density clustering hold a promise for the prediction task with higher accuracy.

References

[1]Zhou, F., Claire, Q., & King, R. D. (2014, December). Predicting the Geographical Origin of Music. In Data Mining (ICDM), 2014 IEEE International Conference on (pp. 1115-1120). IEEE.

[2]Daume, H. (n.d.). A Course in Machine Learning(1st ed.) Baltimore, Maryland.

[3]G. Tzanetakis and P. Cook, “MARSYAS: a framework for audio analysis,” Organised Sound, vol. 4, pp. 169–175, 2000.