



[Center for Machine Learning and Intelligent Systems](#)

Repository
 Web
 

[View ALL Data Sets](#)

Check out the beta version of the new UCI Machine Learning Repository we are currently testing! ✕
 Contact us if you have any issues, questions, or concerns. [Click here to try out the new site.](#)

Geographical Original of Music Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Instances in this dataset contain audio features extracted from 1059 wave files. The task associated with the data is to predict the geographical origin of music.

Data Set Characteristics:	Multivariate	Number of Instances:	1059	Area:	N/A
Attribute Characteristics:	Real	Number of Attributes:	68	Date Donated	2014-10-18
Associated Tasks:	Classification, Regression	Missing Values?	N/A	Number of Web Hits:	98493

Source:

Creators:

Fang Zhou (fang.zhou '@' nottingham.edu.cn)
 The University of Nottingham, Ningbo, China

Donors of the Dataset:

Fang Zhou (fang.zhou '@' nottingham.edu.cn)
 Claire Q (eskoala '@' gmail.com)
 Ross D. King (ross.king '@' manchester.ac.uk)

Data Set Information:

The dataset was built from a personal collection of 1059 tracks covering 33 countries/area. The music used is traditional, ethnic or 'world' only, as classified by the publishers of the product on which it appears. Any Western music is not included because its influence is global - what we seek are the aspects of music that most influence location. Thus, being able to specify a location with strong influence on the music is central.

The geographical location of origin was manually collected the information from the CD sleeve notes, and when this information was inadequate we searched other information sources. The location data is limited in precision to the country of origin.

The country of origin was determined by the artist's or artists' main country/area of residence. Any track that had ambiguous origin is not included. We have taken the position of each country's capital city (or the province of the area) by latitude and longitude as the absolute point of origin.

The program MARSYAS[1] was used to extract audio features from the wave files. We used the default MARSYAS settings in single vector format (68 features) to estimate the performance with basic timbral information covering the entire length of each track. No feature weighting or pre-filtering was applied. All features were transformed to have a mean of 0, and a standard deviation of 1. We also investigated the utility of adding chromatic attributes. These describe the notes of the scale being used. This is especially important as a distinguishing feature in geographical ethnomusicology. The chromatic features provided by MARSYAS are 12 per octave - Western tuning, but it may be possible to tell something from how similar to or different from Western tuning the music is.

[1] G. Tzanetakis and P. Cook, "MARSYAS: a framework for audio analysis," Organised Sound, vol. 4, pp. 169-175, 2000.

Attribute Information:

The dataset is present in two files, where each file corresponds to a different feature sets.

Both files contain the audio features of 1059 tracks.

In the 'default_features_1059_tracks.txt' file, the first 68 columns are audio features of the track, and the last two columns are the origin of the music, represented by latitude and longitude.

In the 'default_plus_chromatic_features_1059_tracks.txt' file, the first 116 columns are audio features of the track, and the last two columns are the origin of the music.

Relevant Papers:

The description of music collection and audio features can be found in:

Fang Zhou, Claire Q and Ross. D. King
Predicting the Geographical Origin of Music, ICDM, 2014

Citation Request:

The following citation is requested if you use the dataset:

Fang Zhou, Claire Q and Ross. D. King
Predicting the Geographical Origin of Music, ICDM, 2014



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)