2014 IEEE International Conference on Data Mining

Predicting the Geographical Origin of Music

Fang Zhou
School of Computer Science and IDIC
The University of Nottingham Ningbo, China
fang.zhou@nottingham.edu.cn

Claire Q
Department of Computer Science
Aberystwyth University, UK
eskoala@gmail.com

Ross D. King
Manchester Interdisciplinary Biocentre
The University of Manchester, UK
ross.king@manchester.ac.uk

***Abstract*—Traditional research into the arts has almost always been based around the subjective judgment of human critics. The use of data mining tools to understand art has great promise as it is objective and operational. We investigate the distribution of music from around the world: geographical ethnomusicology. We cast the problem as training a machine learning program to predict the geographical origin of pieces of music. This is a technically interesting problem as it has features of both classification and regression, and because of the spherical geometry of the surface of the Earth. Because of these characteristics of the representation of geographical positions, most standard classification/regression methods cannot be directly used. Two applicable methods are K-Nearest Neighbors and Random forest regression, which are robust to the non-standard structure of data. We also investigated improving performance through use of bagging. We collected 1,142 pieces of music from 73 countries/areas, and described them using 2 different sets of standard audio descriptors using MARSYAS. 10-fold cross validation was used in all experiments. The experimental results indicate that Random forest regression produces significantly better results than KNN, and the use of bagging improves the performance of KNN. The best performing algorithm achieved a mean great circle distance error of 3,113 km.**

***Keywords*-geographical ethnomusicology; regression; random forest regression;**

I. Introduction

A. An objective approach to understanding art

We hold the philosophical position that we do not fully understand a phenomenon unless we can make a machine that reproduces it. The advantage of this approach to understanding a subject area is that it is objective and operational. This approach contrasts strikingly with that of the traditional approach, which has almost always been based around the subjective judgment of human critics. Often great insight is gained by this subjective approach, but it also has to be granted that there are limitations to the results relying upon the peculiarities of the listener.

We propose to extend the objective approach to understanding phenomena to art - in this case, music. Specifically we will use the success of predictive machine learning programs as a measure of objective success in understanding a phenomenon. This removes personal opinions and expectations in the listener because all decisions are made by machine.

B. Geographical Ethnomusicology

The world contains a vast variety of types of music. This music arose as the result of complex geographical, historical, and prehistorical processes. One way to better understand these processes is to analyse the current geographical distribution of music. The study of this distribution is termed Geographical Ethnomusicology. The problem of determining the geographical origin of a piece of music is complicated. Musical forms are rarely pure. Over time they have influenced each other, and many forms of music have travelled far from their point of origin. *The question we wish to answer is given these complications, how well can a computer predict the geographical origin of a piece of music?* It could be argued that unsupervised spatial clustering methods such as Kohonen nets [1] would be best suited to such a task. However, the problem with such clustering methods is that there is generally no objective measure of success. We could find groups of similar music in terms of the audio, but it would not extract those features most suited to predicting a location. This contrasts with supervised methods, where the labels on known examples (classes or numbers) enable the objective measuring of whether a method is working or not - does it predict well or badly? As we know the geographical location of origin of the music (to some degree) in our corpus we should exploit this information. We therefore cast the problem as that of training a machine learning program to be able to predict the geographical origin of pieces of music, i.e. the computer learns a functional relationship between the audio content and its geographic origin on the globe. This predictive task is possible to some extent by human musicologists.

C. The interesting structure of the data

We decided to cast the problem as a regression problem (predicting a spherical coordinate, latitude and longitude) rather than a classification problem (predicting a country/area). The reasons for this are 1) There are a large number of countries/areas and low number of examples per country/area. 2) There is a natural error metric - geographical distance from true position. In a classification setting this would have to be encoded separately in a cost matrix. This is analogous to the case of a problem with three classes, when they can unordered or ordered (class 2 is intermediate between classes 1 and 3), and in the ordered case you wish to make mistaking an example of class 1 for class 2 less expensive than class 3.

The geographical position of countries/areas on the globe makes the regression problem technically interesting, as it is special two-dimensional manifold. The most straightforward way of representing geographical position is via latitude and

1550-4786/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDM.2014.73

longitude. This however introduces complications: longitude is discontinuous, and latitude/longitude grids of equal degree size have different areas - larger as they approach the equator. Therefore, most standard regression methods, such as linear regression, cannot be directly used as they either predict one real number, or they assume the predicted multiple real numbers are independent.

In order to side-step these difficulties, we apply two methods that are robust to the non-standard structure of the data: K-Nearest Neighbors and Random forest regression. KNN estimates label by computing the distance in the feature space. Whereas, Random forest regression partitions the cases based on the case label. Both methods implicitly utilize the connections between the audio content and its geographical position.

II. Related work

A large amount of research has been done on the development of audio features (attributes) for the computational analysis of music, e.g. [2]. These attributes relate closely to how humans perceive music. One example is the spectral centroid, which is mathematically simple and yet is strongly correlated with human perception of 'brightness' in sound [3].

Audio attributes have typically been used to perform automatic classification and clustering to identify similar pieces of music (for recommendation systems), e.g. to identify mood, genre, emotive content, and various other purposes [4]. In addition to audio attributes other meta-data have been utilized, for example, web searches and social tags, but also MIDI, score reading and lyric mining [5]–[7]. Various machine learning and statistical methods have been applied using these attributes: Support Vector Machines, k-Nearest-Neighbor, Neural Nets etc. with good success for certain applications [8]. However, surprisingly little computational work has been done on computational ethnomusicology. Liu *et al.* which demonstrated the applicability of music analysis techniques to non-western music [9]. Gomez *et al.* applied machine learning to discriminate between western and non-western music, and also found some important features relating to the latitude and longitude of origin of a piece [10]. Tzanetakis' work on computational ethnomusicology [11] sought to investigate the potential application of music information retrieval (MIR) to ethnomusicology.

Our problem is one of Spatial statistics [12]. This is the branch of statistics that relates data values with their spatial location. The most common applications are in statistical geography but its use in epidemiology is also well-known. One famous example is the early work of John Snow who in 1855 proved that cholera was waterborne through demonstrating the occurrence of cholera was clustered around a particular water pump. On a larger scale, geographic information systems including global positioning systems in recent years have created the discipline geospatial information studies wherein large databases of geographic information are analysed using geospatial relationships such as adjacency, containment and distance. Our work can be considered part of the latter category as a distance measure is the eventual output that is used to measure success.

III. Methods

A. The special structure of the data

The problem of predicting spherical coordinates is complicated because of the special characteristics of the coordinate–label. This label refers to a position on the surface of the earth, which is represented by two values, latitude and longitude. Latitude, denoted by ϕ, specifies the point is on the north or south part of the earth. The value is from $-90°$ to $90°$. Longitude, denoted by λ, specifies the point is on the east or west part of the earth. The value is from $0°$ at Prime Meridian to $+180°$ eastward, and to $-180°$ westward.

The latitude/longitude representation has two characteristics. The first characteristic is the discontinuity in longitude values. The line which is opposite the Prime Meridian has two longitude values, $\pm 180°$. Therefore, the longitude of two positions geographically near each other, may be significantly different, one being positive and the other negative. The second characteristic is the latitude/longitude grid is non–linear. The area of grid unit near to the equator is much larger than the one near the pole. This is illustrated in the standard Mercator projection of the globe where countries near the poles are unnaturally large. In general there is no perfect flat projection. In considering the sparsity of our data, we choose a sphere as an approximate representation for the globe, though with more precision still it is an oblate spheroid with certain peaks and troughs across the surface.

In order to deal with these difficulties we applied K-Nearest Neighbor method and Random forest regression as regression methods for the prediction of points on the Earth. These were selected for their ability to do regression predicting data points on the Earth using a latitude/longitude representation.

B. (Standard) K-Nearest Neighbor

A song is a two-tuple $S_i = (F_i, L_i)$, where $F_i = (f_1^i, f_2^i, \cdots, f_n^i)$ is a feature vector and $L_i = (\phi_i, \lambda_i)$ is the corresponding label. Let S^{Tr} be the set of training data, and S^{Te} be the set of test data.

For each test example, S_j^{Te}, the (standard) K-Nearest Neighbor (KNN) method [13] computes the Euclidean distance between the test data S_j^{Te} and each of training data, S_i^{Tr}, in the full feature space, that is, $D(S_j^{Te}, S_i^{Tr}) = \sqrt{\sum_{k=1}^{n}(f_k^j - f_k^i)^2}$. The predicted position of the test data S_j^{Te} is the midpoint of K training data which are nearest to the test data.

To calculate the geodesic midpoint, both latitude and longitude (ϕ_i, λ_i) in the top-K training data are converted to Cartesian coordinates (x_i, y_i, z_i). The average coordinates $(\bar{x}, \bar{y}, \bar{z})$ are converted into the latitude and longitude (ϕ_j^p, λ_j^p) for the midpoint.

$$\phi_j^p = \arctan 2(\bar{z}, \sqrt{\bar{x}^2 + \bar{y}^2}) \quad (1)$$

$$\lambda_j^p = \arctan 2(\bar{y}, \bar{x}) \quad (2)$$

The results are measured by calculating the great circle distance from the true position, L_j^{Te}, to the predicted position, L_j^p . The great circle distance $d(L_j^{Te}, L_j^p)$ is

$$\begin{aligned} &d(L_j^{Te}, L_j^p) = 2 * R * \arctan 2(\sqrt{a}, \sqrt{1-a}), \\ &a = \sin^2(\frac{\phi_j^p - \phi_j^{Te}}{2}) + \cos\phi_j^p \cos\phi_j^{Te} \sin^2(\frac{\lambda_j^p - \lambda_j^{Te}}{2}), \end{aligned} \quad (3)$$

where $R = 6373$.

C. Random KNN

The accuracy of (standard) K-Nearest Neighbor is degraded by the presence of noisy or irrelevant or redundant features, and could be affected by the skewed distribution of samples from different countries/areas. We investigated two modifications of (standard) KNN [14]: (1) Computing the feature distance for a subset of features, instead of all features. The features are randomly and uniformly selected from the feature space without replacement, in a way similar to that used in Random Forests. (2) Use of bagging to select subsets of examples. The training data is randomly and uniformly selected from the original training set with replacement. These approaches reduce the variance of the predictions.

Given a training set S^{Tr} of size m, bagging generates q new training sets $\{S^{Tr'_1}, \cdots, S^{Tr'_q}\}$, each of size m, by sampling from S^{Tr} uniformly with replacement. For each new generated training data, suppose $\mathcal{R}$ features are randomly selected. The Euclidean distance between the test data and each of training data in the selected feature space is computed, and the positions of K training data that are nearest to the test data are selected. For all new generated training data, the previous procedure is repeated, and then total $q * K$ positions are gathered together. From $q * K$ positions, the K most frequent positions are selected, and the predicted position of the test data is computed using Eq. 1 and Eq. 2.

D. Random forest regression

The Random Forest Regression (RFR) method is a process of binary recursive partitioning [15]. A good tree will partition data well, that is, the labels of data in a leaf node are similar to each other. Let $H(U)$ be the function that measures how close the labels of data in the node U are. Assume node U is split into U^L and U^R. The split can be evaluated by computing how pure it will make the data in the child nodes, that is,

$$I(U, U^L, U^R) = H(U) - \sum_{i \in L,R} H(U^i). \quad (4)$$

A larger $I(U, U^L, U^R)$ value indicates that the split makes the labels in the child nodes much purer. In constructing random trees, on each "parent" node, a certain number of features are randomly selected, we choose the one which yields the highest $I(U, U^L, U^R)$ value.

There are several ways to compute how close the data labels are in one group. One option is to calculate the variance of the labels of cases in the node U [16], that is,

$$H_{var}(U) = \sum_{i \in U} d(L_i, \bar{L}_U)^2, \quad (5)$$

where $\bar{L}_U$ is the position of midpoint in the node U. A small variance indicates that the positions tend to be very close.

Another option is to calculate the standard deviation of the labels in the node U [17], that is,

$$H_{sd}(U) = |U| \sqrt{\frac{\sum_{i \in U} d(L_i, \bar{L}_U)^2}{|U| - 1}}. \quad (6)$$

The third option is to compute the absolute deviation of the labels in the node U [18], that is,

$$H_{abs}(U) = \sum_{i \in U} |d(L_i, \bar{L}_U)|. \quad (7)$$

Computing the absolute deviation gives less weight to the extreme cases. Thus it is more robust with respect to the presence of outliers and skewed distributions [19].

IV. Data preparation

A. Music collection

Our corpus was built from a personal collection of 1,142 tracks covering 73 countries/areas. The music used is traditional, ethnic or 'world' only, as classified by the publishers of the product on which it appears. We have not included any Western music 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.

To determine the geographical location of origin we manually collected the information from the CD sleeve notes, and when this information was inadequate we searched other information sources. There are most certainly other options as demonstrated by Govaerts *et. al.* but these have varying levels of accuracy and indeed their ground truth for the experiment was 'personal knowledge' or 'by looking up the origin' [20]. We did not wish to confound the ability of the predictor with incorrect location information. The location data is limited in precision to the country/area of origin - we did not have the time to try to find out more about each track.

The country/area of origin was determined by the artist's or artists' main country/area of residence. Any track that had ambiguous origin was removed from the dataset. We have taken the position of the capital of the country (or the province of the area) by latitude and longitude as the absolute point of origin. The assumption here is that the political capital (or province) is also the cultural capital (or province).

Fig. 1 shows the distribution of songs per country/area for the 33 countries/areas with the most songs.

B. Audio features

The program MARSYAS [21] 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 timbal information covering the entire length of each track. No feature weighting or pre-filtering was applied. All numerical features (that is, all features) were transformed to have a mean of 0, and a standard deviation of 1. We also investigated the utility of adding chromatic

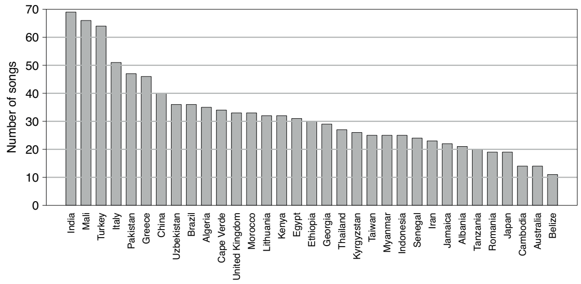
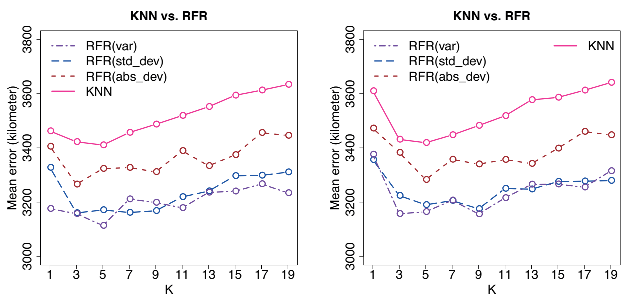


Fig. 1. Partical sample of music distribution by country (or area).



(a) Default features (b) Default + chromatic features

Fig. 2. Mean distance between the true and the predicted positions as a function of K value with the default features (68) and the default with chromatic features (116) respectively.

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.

V. Results

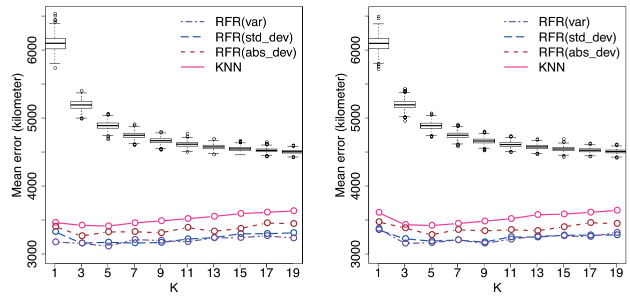
A. Experimental setup

We used data from all countries/areas with more than 10 tracks. In total 1059 tracks were used in the experiments. The tracks from the same country/area are equally distributed in 10 groups. We utilized 10-fold cross-validation in all experiments. Overall prediction accuracy was estimated by calculating the mean error distance from true positions of 1059 tracks to their corresponding predicted positions. The data is available free online at https://sites.google.com/site/icdm2014music/.

Using the Random forest regression model 200 trees were constructed. At each "parent" node, 13 features are randomly selected, and 8 thresholds are set among the range of the feature value. If the node contains less than 8 cases the node was not split.

B. Prediction performance

We first assessed how large the mean distance was between the true and the predicted positions. The average distance error (Fig. 2) is from 3,100 km to 3,600 km. Using the default MARSYAS features (68), the best predictive performance we achieved was a 3,113.392 km mean distance when K is 5. Using the default with the additional chromatic features (116),



(a) Default features (b) Default + chromatic features

Fig. 3. Mean distance error as a function of K. The boxplot shows the distribution of 1000 permutation test results.

the best predictive performance we achieved was 3,157 km with K equal to 3.

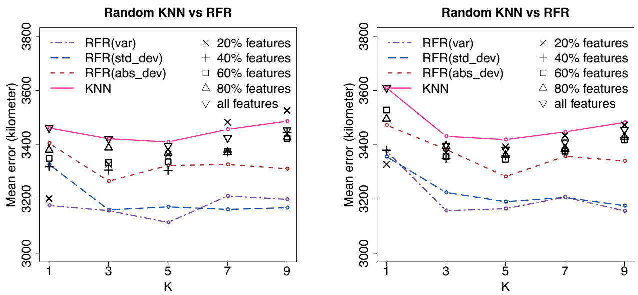
We then investigated whether there is a predictive relationship between the audio features and geographical locations by comparing the results of permutation tests with results of (standard) KNN and Random forest regression. In the permutation test: for each track we randomly selected K cases and computed the midpoint of K selected labels as the predicted position, and then computed the mean error for the whole tracks. The distribution of 1000 fold permutation test results are shown (by using boxplot) in Fig. 3 in both feature spaces. The empirical results (Fig. 3) indicate that the results of both (standard) KNN and Random forest regression fall outside the entire distribution. We can therefore conclude that both types of methods produce significantly (P-value < 0.001) better results than permutation tests. Thus we can infer that there do exist a connection between the audio features and geographical locations.

C. KNN vs RFR

We next compared the performance of (standard) KNN with Random forest regression (Fig. 2). In both feature sets, the (standard) KNN gave the worst results over the whole range of K values. This may be because it takes all features into account, and its performance can be degraded by the presence of noisy or irrelevant or redundant features. The differences between KNN results and Random forest regression results are examined by using paired t-test. Their P-values related to different K using 68 features are in Table I, which show that three approaches in Random forest regression produce significantly better results than KNN especially when K is bigger than 5. The performance is similar when using 116 features.

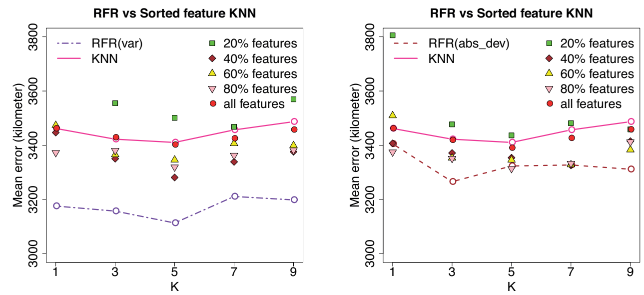
The second worse results (Fig. 2) were produced by Random forest regression with absolute deviation split criterion $H_{abs}(U)$. This is probably because it gives less weights to the outliers. Table I indicates that other two approaches in Random forest regression produce significantly better results than Random forest regression with absolute deviation when K is larger than 1.

Random forest regression with variance split criterion $H_{var}(U)$ and with standard deviation split criterion $H_{std}(U)$ produced the best performance in both feature settings (Fig. 2).



(a) Default features (b) Default + chromatic features

Fig. 4. Mean distance error as a function of K. The points are the results of Random KNN. The number of random features selected are 20%, 40%, 60%, 80%, and 100%.



(a) (b)

Fig. 5. Mean distance error as a function of K. The features are the default 68 features. The points are the results of KNN using a certain percentage of sorted features.

(a) Taiwan (b) Greece

Fig. 6. Distribution of predicted positions of songs from two regions.

Again the reason for this is probably because these two approaches give more weights to the outliers. The differences between the results produced by these two approaches are not significant (see. Table I).

Comparing Fig. 2a with Fig. 2b, we noticed that all approaches produce slightly better results when they use the 68 default features.

D. Random KNN vs RFR

We then compared the performance of Random KNN with (standard) KNN and Random forest regression methods (Fig. 4). 20%, 40%, 60% and 80% features were randomly selected respectively from the feature space without replacement, and their corresponding results are plotted using points with different shape.

In both feature settings Random KNN gives better results than the (standard) KNN (Fig. 4), but the results are worse than the ones produced by Random forest regression approaches, especially Random forest regression with the variance split criterion. We also noticed that when Random KNN uses all features, it still gives better results than the ones produced by (standard) KNN.

E. Sorted feature KNN vs RFR

We compared the performance of Random forest regression methods with KNN using selected features (Fig. 5). The selected features were the ones that are used to split nodes in constructing random forest, and are sorted based on the selected frequency (this was done avoiding training on the test set).

Since Random forest regression with variance criterion and the one with standard deviation criterion produced similar results, we here only show results of Random forest regression with variance criterion and with absolute deviation criterion. Overall, Random forest regression with variance criterion produced better results than the ones produced by KNN using sorted features. KNN using only the first 20% highly selected features produces the worst results. However, when the percentage of selected features increases the performance is better than the (standard) KNN. The results on the default and the chromatic features are analogous to the ones in Fig. 5.

F. Performance by country/area

The algorithm performed much better on some countries/areas than others - even with the same number of tracks available, suggesting that the music from some countries/areas is more distinct than that from others. An additional problem is the relative size of countries/areas.

We applied (Standard) KNN using the default features and gave two examples to illustrate the distributions of predicted positions of songs (Fig. 6). Fig. 6a shows the predicted locations of music from Taiwan. Among the 25 tracks which are from Taiwan, 18 tracks are all correctly positioned precisely in Taiwan area. The maximal distance from the predicted position to the true position is 7,539.853 km, and the mean value is 1,019.651 km.

Fig. 6b shows the distribution of predicted locations of Greek music. Among 47 Greek music, one fourth tracks (the points in Cyan color) are predicted 200 km away from the Greek capital (the blue point). From this it is clear that the prediction range for a country/area can be quite tightly distributed - the furthest estimate is 3,661.478 km but there is a close cluster central to the image that reflects the more general skew in the distribution of estimates for all countries/areas.

VI. Discussion and Conclusions

We have proposed an objective machine learning approach to understanding music. We analyzed the geographical distribution of music - geographical ethnomusicology. The problem is of technical interest to data analysis task because of the special characteristics of the representation of global geographical positions. To deal with this we applied K-Nearest Neighbor and Random forest regression methods for prediction.

We conclude the following from the experiments: (1) All methods performed significantly better than the random null model, which demonstrates there is a predictive relationship between audio features and geographical locations; (2) The

TABLE I

P VALUES OF PAIRED T-TEST. THE BOLD VALUES ARE THE ONES THAT ARE SMALLER THAN THE SIGNIFICANCE LEVEL (0.05).

	K									
	1	3	5	7	9	11	13	15	17	19
RFR(var) $\leq$ RFR(std_dev)	**0.0388**	0.48	0.103	0.888	0.8	0.13	0.445	**0.031**	0.157	**0.005**
RFR(var) $\leq$ RFR(abs_dev)	**0.004**	**0.0489**	**2.19e-05**	**0.009**	**0.009**	**1.99e-06**	**0.01**	**0.001**	**1.82e-06**	**4.21e-08**
RFR(std_dev) $\leq$ RFR(abs_dev)	0.208	**0.044**	**0.002**	**0.0007**	**0.0009**	**8.44e-05**	**0.0136**	**0.03**	**6.68e-05**	**0.0009**
RFR(var) $\leq$ KNN	**0.0045**	**0.0001**	**3.19e-06**	**2.37e-05**	**2.72e-07**	**3.49e-10**	**1.32e-09**	**1.5e-12**	**6.52e-13**	**4.2e-17**
RFR(std_dev) $\leq$ KNN	0.114	**0.0002**	**8.46e-05**	**1.21e-07**	**2.03e-09**	**3.17e-09**	**1.96e-10**	**2.19e-10**	**3.53e-12**	**2.2e-12**
RFR(abs_dev) $\leq$ KNN	0.313	**0.0186**	0.0897	**0.017**	**0.0007**	**0.007**	**1.19e-05**	**2.83e-06**	**0.00055**	**4.78e-05**

three approaches of Random forest regression produced significantly better results than KNN, especially with the variance and standard deviation split criteria; (3) The use of bagging improves the performance of KNN; (4) The best predictive performance achieved has a mean great circle distance of 3,113 km.

There is much scope for further research and improvement in prediction performance. With a larger corpus with both more tracks from each country/area, and more countries/areas represented, the prediction results will inevitably improve. More geographical information could also be utilized. It would be better to have access to the exact location of the origin of the music, rather than just the capital or province, as most countries, like China, have strong regional variations in style. Some cultures change drastically over small areas, while some are unchanged over large expanses, and this needs to integrated in the prediction method.

The music could be better represented for computational analysis. It is a truism within machine learning that the hard part is getting the features correct, and with the correct features almost any learning algorithm will work.

It is difficult to know how good our prediction results are as there are no previously published related comparisons. It would therefore be very interesting to compare the results of the machine learning programs with that of human performance in predicting musical origin. We suspect that the machine learning methods are already quite competitive with humans.

ACKNOWLEDGMENT

The authors would like to thank Chao Zhang for providing RFR code, and also like to thank Guoping Qiu and the anonymous reviewers for their useful comments. This work has been supported by Ningbo Education Bureau, Ningbo Science and Technology Bureau, China's MOST, and by The University of Nottingham (EPSRC grant EP/L015463/1).

REFERENCES

[1] R. O. Duda and P. Hart, "Pattern Classification and Scene Analysis," 1989.

[2] M. Grachten, M. Schedl, T. Pohle, and G. Widmer, "The ISMIR Cloud: A Decade of ISMIR Conferences at Your Fingertips," in *Proceedings of the 10th International Society for Music Information Retrieval Conference*, 2009, pp. 63–68.

[3] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *Journal of The Acoustical Society of America*, vol. 61, pp. 1270–1277, 1977.

[4] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal Music Mood Classification using Audio and Lyrics," in *Proceedings of the 7th International Conference on Machine Learning and Applications*, ser. ICMLA'08. IEEE, 2008, pp. 688–693.

[5] G. Widmer, S. Dixon, P. Knees, E. Pampalk, and T. Pohle, "From Sound to Sense via Feature Extraction and Machine Learning: Deriving High-Level Descriptors for Characterizing Music," *Sound to Sense: Sense to Sound: A State-of-the-Art*, 2005.

[6] D. R. Turnbull, L. Barrington, G. Lanckriet, and M. Yazdani, "Combining Audio Content and Social Context for Semantic Music Discovery," in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '09. ACM, 2009, pp. 387–394.

[7] P. Knees, T. Pohle, M. Schedl, and G. Widmer, "A Music Search Engine Built Upon Audio-based and Web-based Similarity Measures," in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '07. ACM, 2007, pp. 447–454.

[8] M. Mckinney and J. Breebaart, "Features for Audio and Music Classification," in *Proceedings of the International Symposium on Music Information Retrieval*, 2003, pp. 151–158.

[9] Y. Liu, Q. Xiang, Y. Wang, and L. Cai, "Cultural Style Based Music Classification of Audio Signals," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, ser. ICASSP '09, 2009, pp. 57–60.

[10] E. Gómez, M. Haro, and P. Herrera, "Music and Geography: Content Description of Musical Audio from Different Parts of the World," in *Proceedings of the 10th International Society for Music Information Retrieval Conference*, ser. ISMIR'09, 2009, pp. 753–758.

[11] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright, "Computational Ethnomusicology," *Journal of Interdisciplinary Music Studies*, vol. 1, no. 2, pp. 1–24, 2007.

[12] B. D. Ripley, *Spatial Statistics*. Wiley, 2004.

[13] Y. Sun, G. P. Moss, M. Prapopoulou, R. Adams, M. B. Brown, and N. Davey, "Prediction of Skin Penetration Using Machine Learning Methods," in *Proceedings of the 8th IEEE International Conference on Data Mining*, ser. ICDM'08. IEEE, 2008, pp. 1049–1054.

[14] S. Li, "Random KNN Modeling and Variable Selection for High Dimensional Data," Ph.D. dissertation, West Virginia University, 2009.

[15] R. J. Lewis, "An Introduction to Classification and Regression Tree (CART) Analysis," in *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California*, 2000, pp. 1–14.

[16] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] J. Quinlan, "Learning with continuous classes," in *Proceedings of the 5th Australian joint Conference on Artificial Intelligence*, vol. 92. Singapore, 1992, pp. 343–348.

[18] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Chapman & Hall, New York, NY, 1984.

[19] L. Torgo, "Inductive Learning of Tree-based Regression Models," Ph.D. dissertation, Department of Computer Science of the Faculty of Sciences. University of Porto, 1999.

[20] S. Govaerts and E. Duval, "A web-based approach to determine the origin of an artist," in *Proceedings of the 10th International Society for Music Information Retrieval Conference*, ser. ISMIR'09, 2009, pp. 261–266.

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