Automatic detection of outliers in world music collections

Maria Panteli, Emmanouil Benetos, Simon Dixon
{m.panteli, emmanouil.benetos, s.e.dixon}@qmul.ac.uk
Centre For Digital Music, Queen Mary University of London, United Kingdom

Abstract

In big data collections it is often important to identify outlier behaviour that should be filtered out or treated differently. In music this could help us identify sound recordings that stand out in a recorded music collection. We call these recordings 'outliers' and perform a computational analysis to detect them. We focus on world and traditional music collected from available sound archives. Using signal processing tools we extract audio features that denote rhythmic, melodic, harmonic, and timbral aspects of each recording in our collection. Outliers in the dataset are detected with data mining techniques. A sound recording is an outlier depending on how distinct its musical characteristics are compared to other recordings in the collection. However, outlier recordings can also be detected in cases where the descriptors have failed to capture the correct attributes. To evaluate our findings we perform a listening test with music experts. From preliminary results we are able to capture, amongst other, songs with distinct patterns of timbre and rhythm but also speech samples as outliers. The proposed methodology can help identify sound recordings that have a unique musical character or filter irrelevant audio from music collections.

Keywords: computational ethnomusicology, music information retrieval, data mining

1 Introduction

Outlier detection is a common pre-processing step in the analysis of big data collections [1]. We work with large recorded music collections and we are interested to identify sound recordings that are significantly different from the rest. There are two main reasons for this. First, we would like to have an automatic way to go through thousands of recordings and 'clear' our collection from irrelevant audio. For example, a speech or heavily noisy sample could be regarded extreme and undesirable in a collection of sung melodies. Second, given a representative sample of recorded world music, outliers can reveal recordings with outstanding musical characteristics. Tracking the geographic origin of these recordings could help identify areas of the world that have possibly developed a unique musical character.

2 Method

Using music information retrieval techniques we are able to automatically process audio recordings and extract content-based descriptors. We focus on state of the art descriptors (and adaptations of them) that aim at capturing relevant rhythmic, melodic, harmonic, and timbral content. In particular, we extract onset patterns with the scale transform [2], pitch bigrams [3] and chromagrams [4] with 20-cent pitch resolution, and Mel frequency cepstrum coefficients [5]. These features are extracted on a frame base and averaged across all frames in time. Following this process, each recording in our collection is represented by an n-dimensional feature vector, standardised and PCA transformed.

The musical meaning of the selected features could be summarised as follows. Mel frequency cepstrum coefficients provide an estimate of which frequency bands are used in each time frame. This represents an estimate of the quality of sound which includes the timbre of an instrument or instruments, and any background noise. Onset patterns with scale transform describe the loudness and spectral change envelope of different frequency bands over time. They can provide an overview of the accent patterns in the music. Chromagrams define a representation of which pitch values are used in each time frame, wraped within an octave. Averaging a chromagram over time provides an overview of the pitch values most frequently used throughout the recording which could often give rise to the scale or mode. Pitch bigrams are also derived from chromagrams and provide an estimate of which pitch intervals occur more frequently in successive time frames.

To detect outliers we use a method based on squared Mahalanobis distances, a common approach for multivariate data [1, 6]. Using Mahalanobis, an n-dimensional feature vector is expressed as the distance to the mean of the distribution in standard deviation units. Data points that lie beyond a threshold, in this case set to the 97.5% quantile of the chi-square distribution with n degrees of freedom [7], are considered outliers.

In an ideal scenario, the feature extraction process is error free and outliers refer to recordings which have truly distinct musical characteristics. In our case, the feature extraction process can sometimes lead to erroneous representations which could consequently be captured as outliers. To evaluate our findings we will perform a listening experiment. We use the 'spot the odd song out' experiment design [8] where subjects listen to three songs and select the 'odd' candidate as the one that is most different from the other two. Results from this experiment will give more insights on the outlier detection accuracy.

3 Dataset

From available metadata and short (30-second) audio excerpts of the Smithsonian Folkways Recordings collection[1] we sample recordings from different countries of the world. In particular we focus on countries for which there are at least 10 recordings recorded before 1960. In our collection this results in a total of 820 recordings, 10 recordings (chosen at random) from each of 82 countries.

4 Results

Preliminary analysis shows that 5% of the total of 820 recordings are detected as outliers. As shown in Figure 1, the countries in our collection with the most outliers (in our case 3-4 outliers from the total of 10 recordings) are Philippines, Pakistan, and Hungary. An outlier example from Philippines is a sample of Jew's Harp performance from the Hanunoo culture group[2]. Listening to some examples that have been detected as outliers we note that timbral roughness and rhythmic irregularity are, amongst others, distinct characteristics of these outliers. We expect the listening experiment to provide more insights on this evaluation.

5 Discussion & Future Work

A recording is described as the unweighted combination of the features described in Section 2. An outlier means that this combination is different from the other samples in the collection. For example, a recording is identified as outlier not because it has a unique rhythm, but rather because the combination of this particular accent pattern, with the given instruments' timbre, the scale and melodic intervals is unique in

[1] http://www.folkways.si.edu/folkways-recordings/smithsonian

[2] An audio preview and metadata for this outlier example can be found at http://www.folkways.si.edu/hanunoo-kinaban-player/kinaban/world/music/track/smithsonian.

the collection. The features selected for this study capture low-level characteristics of the music signal and they could be used as a basis to infer high-level features, e.g. for detecting the presence of ornaments, estimating complex rhythms, identifying instruments and instrument families, detecting the scale or mode. Depending on the application the concept of an outlier might vary and one should choose features to describe audio recordings accordingly.

While this analysis can give insights on how to find outliers in recorded music collections, there are several issues that need to be addressed before coming to musicological conclusions. First, musical characteristics are captured in the feature vector and if the feature extraction is not reliable then we cannot conclude anything about the music itself. Second, to be able to identify 'world' music outliers one needs to start with a representative world music collection. We have set simple criteria to create a sample collection for this study but one could incorporate more metadata to identify a set of recordings that are representative for a country, culture or region. Lastly, deciding on what is outlier and what is inlier depends on many musical parameters, the collection itself, and the subjective judgement of the listener. In future work we aim to evaluate more our findings and improve the corresponding features and dataset. With the above issues resolved the proposed method can reveal interesting insights on how unique music recordings might be.

6 Acknowledgements

EB is supported by a RAEng Research Fellowship (RF/128). MP is supported by a Queen Mary Principal's research studentship.

References

[1] C. C. Aggarwal and P. S. Yu, "Outlier detection for high dimensional data," *International Conference on Management of Data (ACM SIGMOD)*, pp. 37–46, 2001.

[2] A. Holzapfel and Y. Stylianou, "Scale Transform in Rhythmic Similarity of Music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 176–185, 2011.

[3] J. Van Balen, D. Bountouridis, F. Wiering, and R. Veltkamp, "Cognition-inspired Descriptors for Scalable Cover Song Retrieval," in *International Society for Music Information Retrieval Conference*, pp. 379–384, 2014.

[4] M. Bartsch and G. Wakefield, "Audio thumbnailing of popular music using chroma-based representations," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 96–104, 2005.

[5] J. J. Aucouturier, F. Pachet, and M. Sandler, ""The way it sounds": Timbre models for analysis and retrieval of music signals," *IEEE Transactions on Multimedia*, vol. 7, no. 6, pp. 1028–1035, 2005.

[6] V. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.

[7] P. Filzmoser, "A Multivariate Outlier Detection Method," in *International Conference on Computer Data Analysis and Modeling*, no. 1, pp. 18–22, 2004.

[8] D. Wolff and T. Weyde, "Adapting Metrics for Music Similarity Using Comparative Ratings," in *International Society for Music Information Retrieval Conference*, pp. 73–78, 2011.

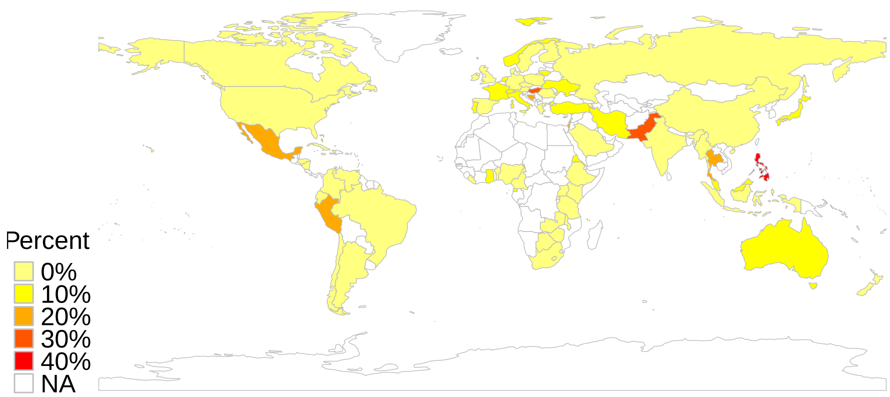

Figure 1: Outliers per country for a total of 820 recordings (10 from each of 82 countries) in our sample collection.