Fusing Web and Audio Predictors to Localize the Origin of Music Pieces for Geospatial Retrieval

Markus Schedl[1] and Fang Zhou[2]

[1]Department of Computational Perception
Johannes Kepler University, Linz, Austria
markus.schedl@jku.at

[2]Center for Data Analytics and Biomedical Informatics
Temple University, Philadelphia, USA
fang.zhou@temple.edu

Abstract. Localizing the origin of a music piece around the world enables some interesting possibilities for geospatial music retrieval, for instance, location-aware music retrieval or recommendation for travelers or exploring non-Western music – a task neglected for a long time in music information retrieval (MIR). While previous approaches for the task of determining the origin of music either focused solely on exploiting the audio content or web resources, we propose a method that fuses features from both sources in a way that outperforms stand-alone approaches. To this end, we propose the use of block-level features inferred from the audio signal to model music content. We show that these features outperform timbral and chromatic features previously used for the task. On the other hand, we investigate a variety of strategies to construct web-based predictors from web pages related to music pieces. We assess different parameters for this kind of predictors (e.g., number of web pages considered) and define a confidence threshold for prediction. Fusing the proposed audio- and web-based methods by a weighted Borda rank aggregation technique, we show on a previously used dataset of music from 33 countries around the world that the median placing error can be reduced from 1,815 to 0 kilometers using K-nearest neighbor regression.

1 Introduction

Predicting the location of a person or item is an appealing task given today's omnipresence and abundance of information about any topic on the web and in social media, which are easy to access through corresponding APIs. While a majority of research focuses on automatically placing images [5], videos [22], or social media users [1, 7], we investigate the problem of placing music at its location of origin, focusing on the country of origin, which we define as the main country or area of residence of the artist(s). We approach the task by audio content-based and web-based strategies and eventually propose a hybrid method that fuses these two sources. We show that the fused method is capable of outperforming stand-alone approaches.

The availability of information about a music piece's or artist's origin opens interesting opportunities, not only for computational ethnomusicology [3], but also for location-aware music retrieval and recommendation systems. Examples

include browsing and exploration of music from different regions in the world. This task seems particularly important as the strong focus on Western music in music information retrieval (MIR) research has frequently been criticized [17, 13]. Other tasks that benefit from information about the origin of music are trend analysis and prediction. If we understand better where a particular music trend emerges – which is strongly related to the music's origin – and how it spreads (e.g., locally, regionally, or globally), we could use this information for personalized and location-aware music recommendation or for predicting the future popularity of a song, album, artist, or music video [10, 25]. Another use case is automatically selecting music suited for a given place of interest, a topic e-tourism is interested in [6].

The remainder of this paper is structured as follows. Section 2 presents related work and highlights the main contributions of the paper at hand. Section 3 presents the proposed audio- and web-based methods as well the hybrid strategy. Section 4 outlines the evaluation experiments we conducted, presents and discusses their results. Eventually, Section 5 rounds off the paper with a summary and pointers to future research directions.

2 Related Work

The research task of automatically position a given multimedia item, such as an image [24], video [22], or text message [7] has received considerable attention in recent years. Also approaches to localize social media users, for instance via deep neural networks have been proposed [11]. Predicting the position or origin of a music entity, such as a music piece, composer, or performer, has been studied to a smaller extent so far.

However, identifying an artist's or piece's origin provides valuable clues about their background and musical context. For instance, a performer's geographic, political, and cultural context or a songwriter's lyrics might be strongly related to their origin. Our problem is to predict the origin of music, relating data values with their spatial location, which is one of the spatial statistics [12]. In the literature, two strategies have been followed to approach this goal: exploiting musical features extracted from the audio content and building predictors based on information harvested from the web.

Audio-based Approaches. One kind of approach is to automatically learn connections between audio features and geographic information of music. Even though this strategy may be considered the most straightforward one, to the best of our knowledge, there exists only one paper exploiting audio signal-based features for the task. Zhou et al. [26] first analyze geographical distribution of music by extracting and analyzing audio descriptors through the MARSYAS [23] software. They use spectral, timbral and chroma features. The authors then apply K-nearest neighbor (KNN) and random forest regression methods for prediction.

Web-based Approaches. Another category of methods approach the problem via web mining. Govaerts and Duval [4] search for occurrences of country names in biographies on Wikipedia[1] and Last.fm,[2] as well as in properties such as *ori-*

[1] https://en.wikipedia.org
[2] http://www.last.fm

gin, *nationality*, *birthplace*, and *residence* on Freebase.[3] The authors then apply heuristics to predict the most probable country of origin for the target artist or band. For instance, one of their heuristics predict the country name that most frequently occurs in an artist's biography. Another one favors early occurrences of country names in the text. Govaerts and Duval show that combining the results of different data sources and heuristics yields superiors results. Schedl et al. propose three approaches that try to predict the country of origin from web pages identified by search engines [16, 14]. One approach is a heuristic that compares the page count estimates returned by Google for queries of the form `"artist/band" +country` and simply predicts the country with highest page count value for a given artist or band. Another approach takes into account the actual content of the web pages. For this purpose, up to 100 top-ranked web pages for each artist are downloaded and TF·IDF weights are computed. The country of origin for a given artist is then predicted as the country with highest TF·IDF score using the artist name as query. Their third approach computes text distances between country names and key terms such as "born" or "founded" in the set of web pages retrieved for the target artist. The country whose name appears closest to any of the key terms is eventually predicted. Schedl et al. show that the approach based on TF·IDF weighting outperforms the other two methods. A shortcoming of all of these web-based approaches is that they only operate on the level of artists, not on pieces. In this paper, by contrast, we build a web-based approach using as input the name of the music piece under consideration.

A related problem is to predict countries in which a music item or artist is particularly popular. This might correspond to their country of origin. Koenigstein et al. propose an approach based on localizing IP addresses of queries issued in peer-to-peer networks [10]. For the same task, they also look into the content of folders users share on peer-to-peer networks [9].

The main contributions of the paper at hand are (i) the investigation of the state-of-the-art audio similarity measure based on the block-level feature extraction framework [21, 20] for the task of predicting the country of origin of individual music pieces, (ii) an extension and comprehensive evaluation of the state-of-the-art web-based method to predict the country of origin of artists, and (iii) a novel method that fuses audio- and web-based predictors. All of these methods are evaluated in classification and regression experiments.

3 Localizing the Origin of Music Pieces

3.1 Music Features

For content-based description of the music pieces, we used a set of six features defined in the block-level framework (BLF) [21]. This choice is motivated by the fact that these features already proved to perform very well for music similarity and retrieval tasks [18], for music autotagging [19], and for content-based modeling in location-aware music recommender systems [6]. They have, however, never been exploited for the task at hand.

[3] http://www.freebase.com

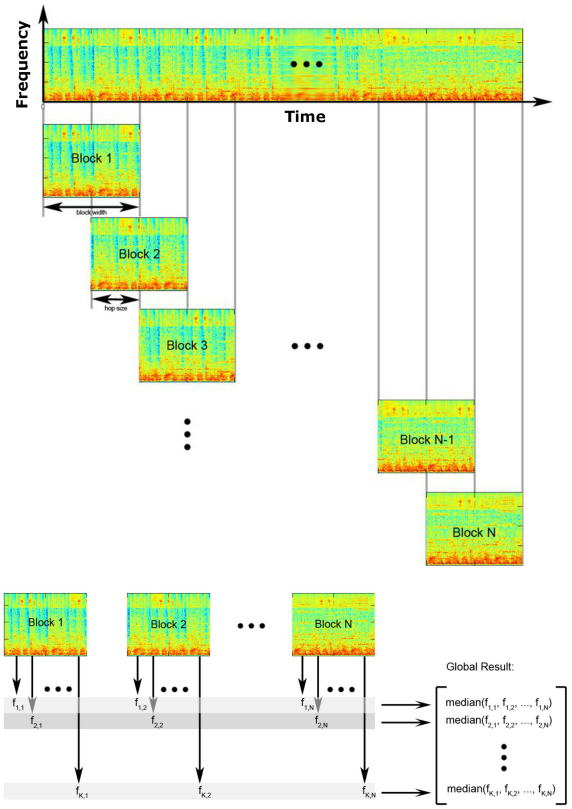


Fig. 1. Overview of the feature extraction (top) and summarization process (bottom) in the block-level framework.

The BLF describes a music piece by defining overlapping blocks over the spectrogram of the audio signal, in which frequency is modeled on the Cent scale, as illustrated in Figure 1 (top). Concretely, a window size of 2,048 samples per frame and a hop size of 512 samples are used to compute the short time Fourier transform on the Hanning-windowed frames of the audio signal. The resulting magnitude spectrum exhibits linear frequency resolution, it is mapped onto the logarithmic Cent scale to account for the human perception of music.

From the resulting Cent spectrogram representation, several features computed on blocks of frames are inferred. Within the BLF, we use the following features: *Spectral Pattern* (SP) characterizes the frequency content, *Delta Spectral Pattern* (DSP) emphasizes note onsets, *Variance Delta Spectral Pattern* (VDSP) captures variations of onsets over time, *Logarithmic Fluctuation Pattern* (LFP) describes the periodicity of beats, *Correlation Pattern* (CP) models the correlation between different frequency bands, and *Spectral Contrast Pattern* (SCP) uses the difference between spectral peaks and valleys to identify tonal and percussive elements. Since the features for a given music piece are computed on the level of blocks, all features of the same kind are eventually aggregated using

a statistical summarization function (typically, percentiles or variance), which is illustrated in Figure 1 (bottom). After this aggregation, each music piece is described by six feature vectors of different dimensionality, totalling to 9,948 individual feature values.

For comparison to the previous audio-based state-of-the-art method [26], we also considered two other groups of audio features, NMdef and NMdefchrom, which were extracted by the program MARSYAS [23]. NMdef contains basic spectral and timbral features, which are *Time Zero Crossings*, *Spectral Centroid*, *Flux* and *Rolloff*, and *Mel-Frequency Cepstral Coefficients* (MFCC), whereas NMdefchrom includes additional *chromatic* attributes to describe the notes of the scale being used. No feature weighting or pre-filtering was applied. All numerical features (i.e. all features) were transformed to have a mean of 0 and a standard deviation of 1.

3.2 Audio-based Prediction of Origin

As proposed in [21], similarities between music pieces are computed as inverse Manhattan distance, considering each of the six BLF features separately. The corresponding six similarity matrices are then Gauss-normalized and eventually linearly combined. For the MARSYAS feature sets, Euclidean distance is used to construct the similarity matrix. Using the similarity matrix, we apply the standard K-nearest neighbor (KNN) [26] as a regression model for the prediction of origin.

For each music piece in the test set, KNN computes the distance between its audio features and the audio features of each music item in the training set, and then sorts the training data items according to the feature distance to the test music in an ascending order. The predicted position of the target music piece is then the midpoint of the K closest training items' spatial position.

To calculate the geodesic midpoint, both latitude and longitude (ϕ, λ) in the top K training data instances are converted to Cartesian coordinates (x, y, z).

$$x = \cos(\phi)\cos(\lambda) \tag{1}$$

$$y = \cos(\phi)\sin(\lambda) \tag{2}$$

$$z = \sin(\phi) \tag{3}$$

The average coordinates $(\bar{x}, \bar{y}, \bar{z})$ are converted into the latitude and longitude (ϕ^p, λ^p) for the midpoint.

$$\phi^p = \arctan 2(\bar{z}, \sqrt{\bar{x}^2 + \bar{y}^2}) \tag{4}$$

$$\lambda^p = \arctan 2(\bar{y}, \bar{x}) \tag{5}$$

The quality of prediction is then measured by calculating the great circle distance from the true position $L^{Te} = (\phi^{Te}, \lambda^{Te})$ to the predicted position $L^p = (\phi^p, \lambda^p)$. The great circle distance $d(L^{Te}, L^p)$ between two points $(\phi^{Te}, \lambda^{Te})$ and (ϕ^p, λ^p) on the surface of the earth is defined as

$$\begin{aligned} &d(L^{Te}, L^p) = 2 \cdot R \cdot \arctan 2(\sqrt{a}, \sqrt{1-a}), \\ &a = \sin^2(\frac{\phi^p - \phi^{Te}}{2}) + \cos\phi^p \cos\phi^{Te} \sin^2(\frac{\lambda^p - \lambda^{Te}}{2}), \end{aligned} \tag{6}$$

where $R = 6{,}373$ kilometers.

3.3 Web-based Country of Origin Prediction

To make predictions for a given music piece p, we first fetch the top-ranked web pages returned by the Bing Search API[4] for several queries: `"piece" music`, `"piece" music biography`, and `"piece" music origin`, in which `"piece"` refers to the exact search for the music piece's name.[5,6] In the following, we abbreviate these query settings by M, MB, and MO, respectively. We subsequently concatenate the content of the retrieved web pages for each p to yield a single document for p. Previous web-based approaches for the task at hand [16, 14] only considered the problem at the artist level and only employed the query scheme `"artist" music`.

Given a list of country names, we compute the term frequency (TF) of all countries in the document of p, and we predict the K countries with highest scores. We do not perform any kind of normalization, nor account for different overall frequencies of country names. This choice was made in accordance with previous research on the topic of country of origin detection, as [16] shows that TF outperforms TF·IDF weighting, and also outperforms more complex rule-based approaches.

In addition to different query settings (M, MB, and MO), we also consider fetching either 20 or 50 web pages per music piece. Knees et al. investigate the influence of different numbers of fetched web pages for the task of music similarity and genre classification [8]. According to the authors, considering more than 50 web pages per music item does not significantly improve results, in some cases even worsens them. Since overall best results were achieved when considering between 20 and 50 pages, we investigate these two numbers.

In order to control for uncertainly in made predictions, we further introduce a confidence parameter α. For each of the top K countries predicted for p, we relate its TF value to the sum of TF values of all predicted countries. We only keep a country c predicted for p if its resulting relative TF value is at least α, C_p being the set of top K countries predicted for p:

$$\frac{TF(p,c)}{\sum_{c \in C_p} TF(p,c)} \geq \alpha \tag{7}$$

3.4 Fusing Audio- and Web-based Predictions

In order to fuse the predictions of our audio-based and web-based algorithms, we propose a variant of the Borda rank aggregation technique [2] with a mixture parameter ξ for linear combination of scores. Variants of this aggregation technique have already been proven useful for music recommendation tasks [6]. Our approach first ranks separately the predictions made by the audio- and the web-based method for a given piece p and then converts ranks to scores, i.e., the top-ranked country among K receives a score of K, the second ranked a score of

[4] https://datamarket.azure.com/dataset/bing/search

[5] Please note that the obvious query scheme `"piece" (music) country` does not perform well as it results in too many irrelevant pages about country music.

[6] Please further note that investigating queries in languages other than English is out of the scope of the work at hand, but will be addressed as part of future work.

$K-1$, and so on. The individual scores for each country are then added up over the approaches to fuse. Since previous research on hybrid music similarity has shown that a simple linear weighting of individual similarities performs well [15], we use a parameter ξ that controls the weight of the audio-based scores. The whole scoring function is thus

$$s(p,c) = \xi \cdot s_{audio}(p,c) + (1-\xi) \cdot s_{web}(p,c). \tag{8}$$

4 Evaluation

We used the dataset presented by Zhou et al. [26], containing 1,059 pieces of music originating from 33 countries. Music was selected based on the following two criteria: First, no "Western" music is included, as its influence is global. We only consider the music that is strongly influenced by a specific location, namely only traditional, ethnic, or "World" music was included in this study, Second, any music that has ambiguous origin was removed from the dataset. The geographical origin was collected from compact disc covers. Since most location information is country names, we used the country's capital city (or the province of the area) to represent the absolute point of origin (represented as latitude and longitude), assuming that the political capital is also the cultural capital of the country or area. The country of origin is determined by the artist's or artists' main country or area of residence. If an artist has lived in several different places, we only consider the place that presumably had major influence. For example, if a Chinese artist is living in New York but composes a piece of traditional Chinese music, we take it as Chinese music.

For evaluation, all music from the same country is equally distributed among 10 groups. We then apply 10-fold cross-validation and report the mean and median error distance (in kilometers) from the true positions to their corresponding predicted positions. We also measure prediction accuracy, i.e. the percentage of music pieces assigned to the correct class, treating countries as classes.

4.1 Prediction Performance of Audio-based Approach

We first compare the prediction performance of using different audio-based features, which is accessed by two criteria, mean distance (solid lines) between the true and the predicted geographical points and median distance (dashed lines). In Figure 2, the black and blue lines are the results of using the baseline features from [26], extracted by MARSYAS, and the orange line represents the performance of the block-level features (cf. Section 3.1). The smallest mean distance error achieved using the NMdef and NMdefchrom features (cf. Section 3) is 3,410 km, and the smallest median distance error (1,815 km) is achieved using the NMdef features. In contrast, using BLF features the smallest mean distance error is 2,191 km, and median distance error is 0 km. All these results obtained considering only one nearest neighbor. Furthermore, Figure 2 clearly shows that the BLF features yield the best results over the whole range of investigated K values when evaluated by both mean and median distance.

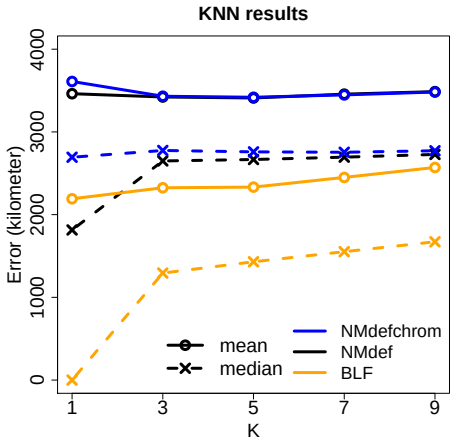


Fig. 2. Mean and median distance errors (kilometers) as a function of K values.

4.2 Prediction Performance of Web-based Approach

Table 1 shows the predictive accuracies for the different parameter settings (α, query scheme, and number of retrieved web pages). Obviously, the `"piece" music` (M) query setting yields the best results, compared to the other two query settings. This is in line with similar findings that strongly restricting the search may yield too specific results and in turn deteriorate performance [8]. As expected, for the same query setting, a higher confidence threshold improves the predictive accuracy.

Regarding the number of web pages, in general, using more web pages increases the amount of information considered. However, only the general query setting M seems to benefit from this. For a threshold of $\alpha = 0.5$, accuracy increases by 4.3 percentage points when comparing the M20 to the M50 setting. In contrast, for the other query settings MB and MO, no substantial increase (MB) or even a decrease (MO) can be observed, when using a large number of pages (and a high α). Taking a closer look at the fetched pages, we identified as a reason for this an increase of irrelevant or noisy pages using the more specific query settings. This might, however, also be influenced by the fact that we are dealing with "World" music. Therefore, many pieces in the collection are not very prominently represented on the web, meaning a rather small number of relevant pages is available.

4.3 Prediction Performance of Hybrid Approach

Figure 3 shows the predictive accuracy for 1-NN for the different parameters (confidence threshold α and mixture coefficient ξ) of the hybrid approach. When ξ equals 0, it means there is no audio-based prediction input. With increasing ξ values, the weight of the audio-based predictions is increasing; when ξ reaches 1, solely audio-based predictions are made. We can clearly observe from Figure 3 a strong improvement of the web-based results when adding audio-based predictions, irrespective of the confidence threshold α. For large confidence thresholds,

Table 1. Accuracies for different variants of the web-based approach (query settings and number of web pages) and various confidence thresholds α. Settings yielding the highest performance are printed in boldface.

α	M20	M50	MB20	MB50	MO20	MO50
0.0	0.439	**0.445**	0.387	0.371	0.349	0.359
0.1	0.439	**0.450**	0.388	0.372	0.349	0.362
0.2	**0.461**	0.460	0.421	0.420	0.365	0.372
0.3	0.503	**0.551**	0.487	0.464	0.442	0.428
0.4	0.573	**0.636**	0.510	0.519	0.532	0.538
0.5	0.641	**0.684**	0.565	0.572	0.612	0.559

including only a small fraction of audio-based predictions actually increases accuracy the most. This means the more confident we are in the web-based predictions, the less audio-based predictions we need to include. Nevertheless, even when $\alpha = 0$, i.e., we consider all web-based predictions, including audio improves performance. We can also observe that different mixture coefficients ξ are required for different levels of α in order to reach peak performance.

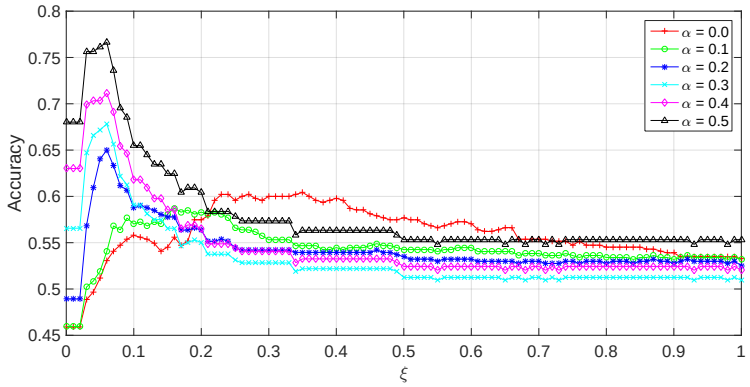


Fig. 3. Accuracies achieved by the proposed hybrid approach, for different values of parameters α (confidence threshold for web-based predictions) and ξ (mixture coefficient for Borda rank aggregation).

Based on the results in Figure 3, we chose the corresponding ξ for different α levels, and then applied KNN regression on the dataset. The respective mean distance errors are shown in Table 2. Please note that the median distance error is 0 km for all settings since the accuracy is always $> 50\%$. The best result that the hybrid approach could reach is 1,824 km, whereas the minimum error of the web-based approach is 2,748 km.

4.4 Comparison of Approaches with Respect to Country Confusions

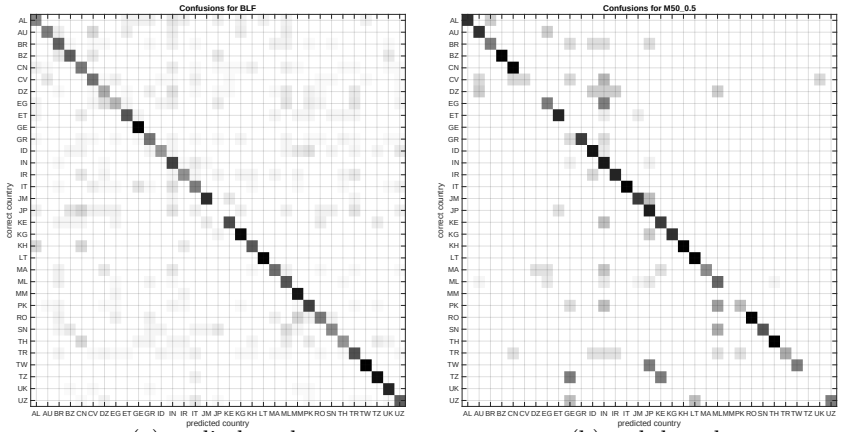
To further assess the types of mistakes made by the different approaches, we show in Figure 4 the country-wise confusion matrices. Rows correspond to the

Table 2. Mean distance error (kilometers) for 1-NN predictions for the audio-based prediction and the hybrid approach with different confidence thresholds α and ξ.

α	ξ	BLF+M50	BLF
0.0	0.35	2,656	2,621
0.1	0.16	2,791	2,616
0.2	0.06	2,540	2,576
0.3	0.06	2,221	2,769
0.4	0.06	2,077	2,833
0.5	0.06	1,825	2,749

true countries, columns correspond to the predicted countries. From the figure we can observe that the predicted locations of origin are spread across the whole matrix when using the audio-based approach (Figure 4(a)), whereas the web-based approach tends to frequently make the same kinds of errors (Figure 4(b)). The audio-based predictor obviously has particular problems correctly localizing Japanese (JP) music. The web-based method frequently misclassifies music as originating in Georgia (GE), India (IN), or Japan (JP), but on the other hand correctly classifies almost all music truly being from Japan. In contrast, Algerian (DZ) and Tanzanian (TZ) music is always misclassified by the web-based predictor.

Due to the different behavior of the audio- and web-based predictors in terms of errors made, fusing the results of both in the way we proposed in Section 3.4 yields better results than the stand-alone approaches.



(a) audio-based (b) web-based

Fig. 4. Confusion matrix for the audio-based BLF approach and the web-based approach (M, $p = 50$, $\alpha = 0.5$). Country names are encoded according to ISO 3166-1 alpha-2 codes.

5 Conclusion and Outlook

We have proposed a novel approach that fuses audio content-based and web-based strategies to predict the geographical origin of pieces of music. We further investigated for this task the use of block-level audio features (BLF), which are already known to perform well for music classification and autotagging. The proposed web-based predictor extends a previous approach, which we modified to (i) make predictions on the level of pieces, not only artists, (ii) consider different query schemes and (iii) numbers of fetched web pages, and (iv) include a confidence threshold for predictions based on relative TF weights.

We conducted KNN experiments on a standardized dataset consisting of 1,059 pieces of music originating from 33 countries. From the experimental results, we conclude that: (i) the audio-based approach that uses block-level features outperforms other standard audio descriptors, such as spectral, timbral, and chromatic features, for the task, (ii) for music from most countries, web-based results are superior to audio-based results, and (iii) the hybrid method produces substantially better results than the single audio-based and web-based approaches.

Given the large amount of non-Western music in the collection, we will look into multilingual extensions to our web-based approach. Furthermore, based on the finding that, for the used dataset, more specific query schemes deteriorate performance, rather then boost it, which is because of the small amount of relevant web pages, we will investigate whether this finding also holds for mainstream Western music. To this end, we will additionally investigate larger datasets, as ours is relatively small in comparison to the ones used in geolocalizing other kinds of multimedia material. We also plan to create more precise annotations for the origin of the pieces since the current granularity, i.e. the capital of the country or area, may introduce a distortion of results. Finally, we plan to look into data sources other than web pages, for instance social media, and to investigate aggregation techniques other than Borda rank aggregation.

6 Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655. The authors would further like to thank Klaus Seyerlehner for his implementation of the block-level feature extraction framework and Ross D. King and the reviewers for their valuable comments on the manuscript.

References

1. Z. Cheng, J. Caverlee, and K. Lee. You Are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users. In *Proc. CIKM*, Oct 2010.
2. J.-C. de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 1781.
3. E. Gómez, P. Herrera, and F. Gómez-Martin. Computational Ethnomusicology: Perspectives and Challenges. *Journal of New Music Research*, 42(2):111–112, 2013.
4. S. Govaerts and E. Duval. A Web-based Approach to Determine the Origin of an Artist. In *Proc. ISMIR*, Oct 2009.

5. C. Hauff and G.-J. Houben. Placing Images on the World Map: A Microblog-based Enrichment Approach. In *Proc. SIGIR*, Aug 2012.
6. M. Kaminskas, F. Ricci, and M. Schedl. Location-aware Music Recommendation Using Auto-Tagging and Hybrid Matching. In *Proc. RecSys*, Oct 2013.
7. S. Kinsella, V. Murdock, and N. O'Hare. "I'm Eating a Sandwich in Glasgow": Modeling Locations with Tweets. In *Proc. SMUC*, Oct 2011.
8. P. Knees, M. Schedl, and T. Pohle. A Deeper Look into Web-based Classification of Music Artists. In *Proc. LSAS*, Jun 2008.
9. N. Koenigstein and Y. Shavitt. Song Ranking Based on Piracy in Peer-to-Peer Networks. In *Proc. ISMIR*, Oct 2009.
10. N. Koenigstein, Y. Shavitt, and T. Tankel. Spotting Out Emerging Artists Using Geo-Aware Analysis of P2P Query Strings. In *Proc. KDD*, Aug 2008.
11. J. Liu and D. Inkpen. Estimating User Location in Social Media with Stacked Denoising Auto-encoders. In *Proc. Vector Space Modeling for NLP*, Jun 2015.
12. B. D. Ripley. *Spatial Statistics*. Wiley, 2004.
13. M. Schedl, A. Flexer, and J. Urbano. The Neglected User in Music Information Retrieval Research. *J. Intell. Inf. Syst.*, 41:523–539, Dec 2013.
14. M. Schedl, C. Schiketanz, and K. Seyerlehner. Country of Origin Determination via Web Mining Techniques. In *Proc. AdMIRe*, Jul 2010.
15. M. Schedl and D. Schnitzer. Hybrid Retrieval Approaches to Geospatial Music Recommendation. In *Proc. SIGIR*, Jul–Aug 2013.
16. M. Schedl, K. Seyerlehner, D. Schnitzer, G. Widmer, and C. Schiketanz. Three Web-based Heuristics to Determine a Person's or Institution's Country of Origin. In *Proc. SIGIR*, Jul 2010.
17. X. Serra. Data Gathering for a Culture Specific Approach in MIR. In *Proc. AdMIRe*, Apr 2012.
18. K. Seyerlehner, M. Schedl, P. Knees, and R. Sonnleitner. A Refined Block-Level Feature Set for Classification, Similarity and Tag Prediction. In *Extended Abstract MIREX*, Oct 2009.
19. K. Seyerlehner, R. Sonnleitner, M. Schedl, D. Hauger, and B. Ionescu. From Improved Auto-taggers to Improved Music Similarity Measures. In *Proc. AMR*, Oct 2012.
20. K. Seyerlehner, G. Widmer, and T. Pohle. Fusing Block-Level Features for Music Similarity Estimation. In *Proc. DAFx*, Sep 2010.
21. K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees. Automatic Music Tag Classification based on Block-Level Features. In *Proc. SMC*, Jul 2010.
22. M. Trevisiol, H. Jégou, J. Delhumeau, and G. Gravier. Retrieving Geo-location of Videos with a Divide & Conquer Hierarchical Multimodal Approach. In *Proc. ICMR*, Apr 2013.
23. G. Tzanetakis and P. Cook. MARSYAS: A Framework for Audio Analysis. *Organised Sound*, 4:169–175, 2000.
24. S. Workman, R. Souvenir, and N. Jacobs. Wide-Area Image Geolocalization with Aerial Reference Imagery. In *Proc. ICCV*, Dec 2015.
25. H. Yu, L. Xie, and S. Sanner. Twitter-driven YouTube Views: Beyond Individual Influencers. In *Proc. ACM Multimedia*, Nov 2014.
26. F. Zhou, Q. Claire, and R. D. King. Predicting the Geographical Origin of Music. In *Proc. ICDM*, Dec 2014.