Location-Aware Music Artist Recommendation

Markus Schedl[1] and Dominik Schnitzer[2]

[1] Johannes Kepler University Linz, Austria
`http://www.cp.jku.at`

[2] Austrian Research Institute for Artificial Intelligence, Vienna, Austria
`http://www.ofai.at`

Abstract. Current advances in music recommendation underline the importance of multimodal and user-centric approaches in order to transcend limits imposed by methods that solely use audio, web, or collaborative filtering data. We propose several hybrid music recommendation algorithms that combine information on the *music content*, the *music context*, and the *user context*, in particular integrating geospatial notions of similarity. To this end, we use a novel standardized data set of music listening activities inferred from microblogs (`MusicMicro`) and state-of-the-art techniques to extract audio features and contextual web features. The multimodal recommendation approaches are evaluated for the task of music artist recommendation. We show that traditional approaches (in particular, collaborative filtering) benefit from adding a user context component, geolocation in this case.

Keywords: Music Information Retrieval, Hybrid Music Recommendation, Personalization, Evaluation

1 Introduction

Music Information Retrieval (MIR) is currently seeing a paradigm shift, away from system-centric perspectives towards user-centric approaches [4]. Incorporating user models and addressing user-specific demands in music retrieval and music recommendation is hence becoming more and more important.

Given the importance of user-centric and hybrid methods to MIR, we propose here several approaches that combine *music content*, *music context*, and *user context* aspects to build a music retrieval system [14]. Music content and music context are incorporated using state-of-the-art feature extractors and corresponding similarity estimators. The user context is addressed by taking into account *musical preference* and *geospatial data*, using a standardized collection of listening behavior mined from microblog data [13].
We make use of the best feature extraction and similarity computation algorithms currently available to model *music content* and *music context*. We then integrate these similarity models as well as a *user context* model into several novel user-aware music recommendation approaches that encompass all three modalities important to human music perception [14].

The remainder of the paper is organized as follows. Section 3 details the acquisition of the raw music (meta-)data, which serves as input to the feature extraction and data representation techniques presented in Section 4. Section 5 proposes several methods to incorporate geospatial information into music recommendation models. We further provide experimental evidence that adding a geospatial, user-aware component to a single-modality recommendation strategy is capable of improving recommendation results. Section 2 briefly reviews related literature. Finally, Section 6 draws conclusions and points to further research directions.

2 Related Work

Specific related work on geospatial music retrieval is very sparse, probably due to the fact that geospatially annotated music listening data is hardly available. Among the few works, Park et al. [7] use geospatial positions and suggest music that matches a selected environment, based on aspects such as ambient noise, surrounding, or traffic. Raimond et al. [10] combine information from different sources to derive geospatial information on artists, aiming at locating them on a map. Another possibility to link music to geographical information is presented by Byklum [2], who searches lyrics for geographical content like names of cities or countries. Zangerle et al. [17] use a co-occurrence-based approach to map tweets to artists and songs and eventually construct a music recommendation system. However, they do not take location into account.

On a more general level, this work relates to context-based and hybrid recommendation systems, a detailed review of which is unfortunately beyond the scope of the paper. A comprehensive elaboration, including a decent literature overview, can be found in [11].

3 Data Acquisition

The only standardized public data set of general microblogs, as far as we are aware of, is the one used in the TREC 2011 and 2012 Microblog tracks[3] [5]. Although this set contains approximately 16 million tweets, it is not suited for our task as it is not tailored to music-related activities, i.e. the amount of music-related posts is marginal.

We hence have to acquire multimodal data sets of *music items* and *listeners*, reflecting the three broad aspects of human music perception (*music content*, *music context*, and *user context*) [14]. Whereas the *music content* refers to all information that is derived from the audio signal itself (such as ryhthm, timbre, or melody), the *music context* covers contextual information that cannot be derived from the actual audio with current technology (e.g., meaning of song lyrics, background of a performer, or co-listening relationships between artists).

[3] http://trec.nist.gov/data/tweets

The *user context* encompasses all information that describe the listener. Examples range from musical education to spatiotemporal properties to physiological measures to current activities.

User Context Only very recently a data set of music listening activities inferred from microblogs has been released [13]. It is entitled `MusicMicro` and is freely available[4], fostering reproducibility of social media-related MIR research. This data set contains about 600,000 listening events posted on `Twitter`[5]. Each event is represented by a tuple *¡twitter-id, user-id, month, weekday, longitude, latitude, country-id, city-id, artist-id, track-id¿*, which allows for spatiotemporal identification of listening behavior.

Music Content Based on the lists of artist and song names in the `MusicMicro` collection, we gather snippets of the songs from `7digital`[6]. These serve as input to the music content feature extractors.

Music Context To capture aspects of human music perception which are not encoded in the audio signal, we extract music-related web pages. Following the approach suggested in [15], we retrieve the top 50 web pages returned by the `Bing`[7] search engine for queries comprising the artist name[8] and the additional keyword "music", to disambiguate the query for artists such as "Bush", "Kiss", or "Hole".

In summary, we gathered raw data covering each of the three categories of perceptual music aspects [14]: *music content* (audio snippets), *music context* (related web pages), and *user context* (user-specific music listening events with spatiotemporal labels).

4 Data Representation

To represent the *music content*, we use state-of-the-art audio music feature extractors proposed in [8], which constitute a reference in music feature extraction for similarity-based retrieval. In particular, we extract auditory music features that combine various rhythmic information derived from the audio signal, e.g., "onset patterns" and "onset coefficients" (note onsets), with timbral features, e.g., "Mel Frequency Cepstral Coefficients" and the two handcrafted specialized descriptors for "attackness" and "harmonicness". The eventual output is pairwise similarity estimates between songs, which are later aggregated to the artist level.

We again employ a state-of-the-art technique to obtain features reflecting the *music context*. To describe the music items at the artist level, we follow the approach proposed in [15]. In particular, we model each artist by creating a "virtual

[4] `http://www.cp.jku.at/musicmicro`

[5] `http://www.twitter.com`

[6] `http://www.7digital.com`

[7] `http://www.bing.com`

[8] Please note that issuing queries at the song level is not reasonable, as doing so typically yields only very few results.

artist documents", i.e. we concatenate all web pages retrieved for the artist. In accordance with findings of [12], we then use a dictionary of music-related terms (genres, styles, instruments, and moods) to index the resulting documents. From the index, we compute term weights according to the best feature combination found in the large-scale experiments of [15]: `TF_C3.IDF_I.SIM_COS`, i.e. computing term weight vectors and artist similarity estimates according to Equations 1, 2, and 3, respectively for *tf*, *idf*, and *cosine similarity*; $f_{d,t}$ represents the number of occurrences of term t in document d, N is the total number of documents, $\mathcal{D}_t$ is the set of documents containing term t, F_t is the total number of occurrences of term t in the document collection, $\mathcal{T}_d$ is the set of distinct terms in document d, and W_d is the length of document d.

$$tf_{d,t} = 1 + \log_2 f_{d,t} \tag{1}$$

$$w_t = 1 - \frac{n_t}{\log_2 N}, \quad n_t = \sum_{d \in \mathcal{D}_t} \left(- \frac{f_{d,t}}{F_t} \log_2 \frac{f_{d,t}}{F_t} \right) \tag{2}$$

$$S_{d_1,d_2} = \frac{\sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} \cdot w_{d_2,t})}{W_{d_1} \cdot W_{d_2}} \tag{3}$$

Rectifying the Similarity Space

Recent work has shown that "hubs" can be a problem in similarity spaces [9]. Hubs are data items that are frequently found among the nearest neighbors of many other data items, but cannot have all of these data items as nearest neighbors themselves. In recommendation systems, such hubs are usually undesired, because they are unjustifiably recommended much more frequently than any other data items, which strongly hinders serendipitous encounters, hence harms user satisfaction. To alleviate this problem, [16] suggests an approach called "mutually proximity" (MP), which rectifies high-dimensional similarity spaces in which the data set itself has low intrinsic dimensionality. This MP approach proved particularly beneficial for text features and music audio features, as shown in [16]. In the case of the audio features used here, this normalization is already included in the employed feature extraction algorithm. For the web-based music context features, we apply MP on the similarity matrix to suppress the formation of hubs in the ultimate recommendation approach.

Availability of the Data Sets

All components of the data set used in this paper are publicly available to allow researchers reproduce the results reported. The sole exception is the actual audio content of the songs under consideration. We cannot share them due to copyright restrictions. However, we provide identifiers by means of which corresponding 30-second-clips can be downloaded from `7digital`. The `MusicMicro` set of geolocalized music listening events from microblogs [13] can be downloaded as well[9]. All other data (audio feature vectors and artist term weight vectors) can be shared upon request to the first author.

5 Music Recommendation Models

Hybrid music retrieval and recommendation approaches, which base their similarity computation on more than one modality, are frequently suggested in literature, e.g. [3, 1, 4, 14, 6]. A systematic evaluation of approaches that integrate state-of-the-art music content (audio) and music context (web) similarity measures was only conducted very recently, though `[omitted-due-to-review]`. One finding of this study is that including a small amount of audio features in an otherwise solely web-based similarity measure (or vice versa) considerably improves retrieval performance. Given audio similarities $asim(i, j)$ and web similarities $wsim(i, j)$ between two artists i and j, Equation 4 shows the hybrid similarity model that performed best for artist retrieval according to `[omitted-due-to-review]`[10]. It is hence used in the following as music content/music context-based model (CB).

$$sim(i, j) = 0.15 \cdot asim(i, j) + 0.85 \cdot wsim(i, j) \tag{4}$$

Building user-aware recommendation systems obviously requires a user model. In our case, each user u is modeled by the set of artists $UM(u)$ he or she listened to. Based on this simple model, we implement the following recommendation strategies:

- CB: the hybrid (music content and music context) music retrieval model according to Equation 4
- CF: a standard user-based collaborative filtering model
- GEO: a model solely based on geospatial proximities
- GEO-CF: a model that combines GEO and CF by taking the union of the two recommended artist sets
- CF-GEO-LIN and CF-GEO-GAUSS: CF-based models that weight users according to their geospatial distance to the seed user, using either a linear or exponential geospatial distance measure
- RB: a random baseline model

In the CB model, the hybrid music similarity function (Equation 4) is used to determine the artists closest to $UM(u)$, which are then recommended. In the CF model, the K users closest to u are determined (using the Jaccard index between the user models), and the artists listened to by these nearest users are recommended. The GEO model defines user distance solely via the geospatial distance between users. To this end, we first compute a centroid of each user

[9] `http://www.cp.jku.at/musicmicro`

[10] Audio similarities are aggregated on the artist level by computing the minimum of the distances between all pairs of tracks by i and j.

u's geospatial listening distribution $\mu_u[\lambda, \varphi]$[11]. For recommendation, the artists of the seed user's closest neighbors, measured via geodesic distance between the centroids, are suggested. The GEO-CF model simply recommends the union of the GEO and CF model's output.

To integrate geospatial information into the CF model (CF-GEO-LIN and CF-GEO-GAUSS), we use the normalized geodesic distance $gdist(u, v)$ (Equation 5) between the seed user u and each other user v to weight the distance based on the user models. To this end, we propose two different weighting schemes: linear weighting and weighting according to a Gaussian kernel around $\mu_u[\lambda, \varphi]$. We eventually obtain a geospatially modified user similarity $sim(u, v)$ by adapting the Jaccard index between $UM(u)$ and $UM(v)$ via geospatial linear or Gauss weighting, according to Equation 6 (CF-GEO-LIN) or Equation 7 (CF-GEO-GAUSS), respectively. We recommend the artists listened to by u's nearest users v.

$$\begin{aligned} gdist(u,v) = &\arccos\big(\,\sin(\mu_u[\varphi]) \cdot \sin(\mu_v[\varphi]) + \cos(\mu_u[\varphi]) \cdot \\ &\qquad \cos(\mu_v[\varphi]) \cdot \cos(\mu_u[\lambda] - \mu_v[\lambda])\,\big) \cdot \\ &\max(gdist)^{-1} \end{aligned} \tag{5}$$

$$sim(u,v) = J(UM(u), UM(v)) \cdot gdist(u,v)^{-1} \tag{6}$$

$$sim(u,v) = J(UM(u), UM(v)) \cdot \exp(-gdist(u,v)) \tag{7}$$

For comparison, we further implemented a random baseline model (RB) that randomly picks K users from the filtered user set (filtering with respect to the parameter τ, see below) and recommends the artists they listened to. In addition, we ensure that all algorithms recommend approximately the same number of artists on average, to make results comparable. To this end, we use the number of artists recommended by the CF approach as baseline and adapt the parameters of the other approaches in such a way that they output a similar number of artists.

5.1 Experimental Setup

In order to ensure sufficient artist coverage of users, we evaluate our models using different thresholds τ for the minimum number of unique artists a user must have listened to in order to include him or her in the experiments. We vary τ between 30 and 200 using a step size of 10. Denoting as U_τ the number of users in the `MusicMicro` data set with equal or more than τ unique artists, $U_{30} = 881$, $U_{100} = 32$, and $U_{200} = 5$. We perform U_τ-fold leave-one-out cross-validation for each value of τ.

[11] It is common to denote longitude by λ and latitude by φ.

5.2 Results

Figure 1 shows accuracies for $K = 5$ nearest neighbors and $\tau = [30 \ldots 200]$. We can see that all approaches significantly outperform the random baseline. Results for the CB approach and the CF approaches show an inverse characteristics over τ, which suggests a combination of both. As this is not trivial, it will be part of future work. The reason for CB outperforming CF for large numbers of τ is obviously the limited diversity among the K nearest neighbors (for $\tau > 150$), which seriously hampers CF-based approaches. Hence, "power users" benefit more from CB approaches than from CF approaches. The GEO approach performs rather poorly, being quite close to the baseline in most settings. Creating a recommender solely on location information hence seems not beneficial.

The hybrid approaches that use geospatial weighting to adapt CF-based similarities do not outperform the CF only approach. Possible explanations are (i) that using the centroid of a user's listening positions as summary of his or her overall location is too coarse a description, in particular for users who travel a lot, and (ii) that geodesic distance alone frequently does not reflect cultural distance, which seems more important for the recommendation task at hand. For instance, same geodesic distances between two users can have very different meaning in regions with different population density (e.g., Hong Kong versus Russia) or at cultural borders (North Korea versus South Korea, Spain versus Morocco, etc.). Future work will take a closer look at these aspects and investigate whether incorporating political and cultural information will improve results. In contrast, the hybrid approach GEO-CF that takes the set union of GEO- and CF-based recommendations performs superiorly. Considering both similar users and similar locations as equally important (GEO-CF) thus outperforms geospatial weighting of user similarities performed in CF-GEO-LIN and CF-GEO-GAUSS.

6 Conclusions and Outlook

We presented different hybrid music recommendation approaches that use state-of-the-art music content (audio) and music context (web) features, as well as contextual user information, more precisely a data set of geolocated music listening activities mined from microblogs. Experimental results indicate that hybrid (music content/music context/user context) strategies, in general, are capable of outperforming approaches using only one data source. However, the question of how to combine the different modalities is crucial. Among the hybrid approaches, we found that recommending the set union of a user-based collaborative filtering recommender and of a recommender based on geospatial proximity of users performed superior.

Future work includes investigating other data sources related to the user context, for instance, listening time or demographics. Furthermore, we presume that refining the notion of spatial proximity by taking into account political and cultural borders, for instance, defined by language or religion, will lead to better

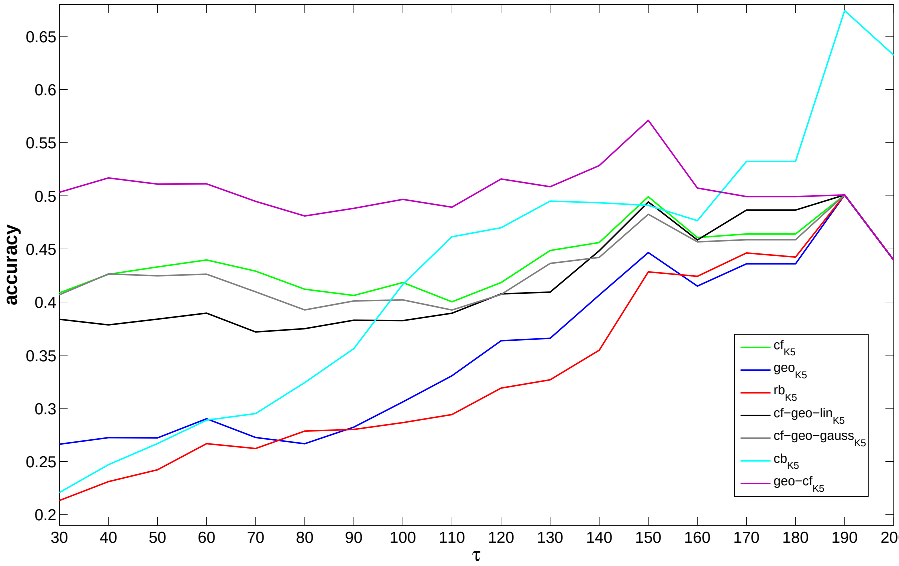

Fig. 1. Accuracy plots for different values of τ and $K = 5$.

prediction accuracy, thus in turn to better user-aware music recommendation systems.

Acknowledgments

This research is supported by the Austrian Science Funds (FWF): P22856 and P24095, and by the EU FP7 project PHENICX: 601166.

References

1. D. Bogdanov, J. Serrà, N. Wack, P. Herrera, and X. Serra. Unifying Low-Level and High-Level Music Similarity Measures. *IEEE Transactions on Multimedia*, 13(4):687–701, Aug 2011.
2. D. Byklum. Geography and Music: Making the Connection. *Journal of Geography*, 93(6):274–278, 1994.
3. E. Coviello, A. B. Chan, and G. Lanckriet. Time Series Models for Semantic Music Annotation. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(5):1343–1359, Jul 2011.
4. C. Liem, M. Müller, D. Eck, G. Tzanetakis, and A. Hanjalic. The Need for Music Information Retrieval with User-centered and Multimodal Strategies. In *Proc. MIRUM*, Scottsdale, AZ, USA, 2011.
5. R. McCreadie, I. Soboroff, J. Lin, C. Macdonald, I. Ounis, and D. McCullough. On Building a Reusable Twitter Corpus. In *Proc. SIGIR*, Portland, OR, USA, 2012.

6. B. McFee and G. Lanckriet. Heterogeneous Embedding for Subjective Artist Similarity. In *Proc. ISMIR*, Kobe, Japan, 2009.
7. S. Park, S. Kim, S. Lee, and W. S. Yeo. Online Map Interface for Creative and Interactive MusicMaking. In *Proc. NIME*, Sydney, Australia, 2010.
8. T. Pohle, D. Schnitzer, M. Schedl, P. Knees, and G. Widmer. On Rhythm and General Music Similarity. In *Proc. ISMIR*, Kobe, Japan, 2009.
9. M. Radovanović, A. Nanopoulos, and M. Ivanović. Hubs in Space: Popular Nearest Neighbors in High-dimensional Data. *The Journal of Machine Learning Research*, pages 2487–2531, 2010.
10. Y. Raimond, C. Sutton, and M. Sandler. Automatic Interlinking of Music Datasets on the Semantic Web. In *Proc. WWW: LDOW Workshop*, Beijing, China, 2008.
11. F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors. *Recommender Systems Handbook*. Springer, 2011.
12. M. Schedl. #nowplaying Madonna: A Large-Scale Evaluation on Estimating Similarities Between Music Artists and Between Movies from Microblogs. *Information Retrieval*, 15:183–217, June 2012.
13. M. Schedl. Leveraging Microblogs for Spatiotemporal Music Information Retrieval. In *Proc. ECIR*, Moscow, Russia, 2013.
14. M. Schedl and A. Flexer. Putting the User in the Center of Music Information Retrieval. In *Proc. ISMIR*, Porto, Portugal, 2012.
15. M. Schedl, T. Pohle, P. Knees, and G. Widmer. Exploring the Music Similarity Space on the Web. *ACM Transactions on Information Systems*, 29(3), Jul 2011.
16. D. Schnitzer, A. Flexer, M. Schedl, and G. Widmer. Local and Global Scaling Reduce Hubs in Space. *Journal of Machine Learning Research*, 13:2871–2902, October 2012.
17. E. Zangerle, W. Gassler, and G. Specht. Exploiting Twitter's Collective Knowledge for Music Recommendations. In *Proc. WWW: #MSM Workshop*, Lyon, France, 2012.