A Model for Serendipitous Music Retrieval

Markus Schedl, David Hauger
Johannes Kepler University
Department of Computational Perception
{markus.schedl,david.hauger}@jku.at

Dominik Schnitzer
Austrian Research Institute for
Artificial Intellgence
dominik.schnitzer@ofai.at

ABSTRACT

Music retrieval systems that take into account the user's taste and information or entertainment need when building the results set to a query are of vital interest for academia, industry, and the passionate music listener. Unfortunately, preliminary attempts to incorporate such aspects have been rather sparse so far. Focusing on the problem of music recommendation, we therefore present a new model that combines several factors we deem to be important for personalizing retrieval results: similarity, diversity, popularity, hotness, recentness, novelty, and serendipity. We further propose different ways to measure the corresponding aspects and, where available, point to literature for a more detailed elaboration of the corresponding measures. In addition, we propose the use of social media mining techniques to address the problem of estimating popularity and hotness in a geo-aware manner.

MOTIVATION AND BACKGROUND

Most music retrieval approaches focus on the concept of *musical similarity* to create the results set for a query, which may be an excerpt of a music piece or the name of an artist of a song. This musical similarity may be computed on some kind of acoustic features extracted from the audio signal via signal processing techniques (*content-based*); alternatively, it might be derived from listening co-occurrences among users (*context-based* or *collaborative filtering*). Accordingly, the performance of such a classical retrieval system is judged the better the more similar the returned pieces are to a given seed. Also the most important evaluation forum for music retrieval methods, the annual `Music Information Retrieval Evaluation eXchange` (MIREX) [4] focuses strongly on similarity as relevance criterion. Although this is a very intuitive manner of assessment, it does not take into account that the information need of the user might not be centered around the concept of similarity alone. Indeed, for many popular music retrieval tasks, such as automated playlist generation [6, 10] and music recommendation [9, 7], the listener does not necessarily seek for a list of closest matches in terms of similarity. In a user study we performed to assess the quality of an automated, content-based playlist generation approach [15], we were in fact often told that our playlists were too perfect or homogeneous, thus boring.

We therefore believe that a new generation of music retrieval systems should not only rely on similarity scores derived from the audio signal or from contextual data, but should instead take the following factors into account:

Similarity:
Similarity should be taken into account in various dimensions. One set of dimensions might be based on music properties such as rhythm, harmony, or timbre, inferred from the audio signal [8, 19]. Another might model resemblance according to other data sources, such as collaborative tags, playlist co-occurrences, or even images of album covers or promotional photographs [12, 11]. A third set of dimensions might be learned from a user's listening preferences, for example, by relating certain properties of the user context to particular categories of music [13]. To give an example, similarity could be defined as pieces that are usually listened together while a user is jogging or while being together with friends.

Diversity:
Although the results of a music retrieval request should contain music items similar to the query, they should also show a certain degree of diversity. For example, there is the well-known "album effect" [20], i.e., due to same recording settings, tracks on the same album usually show a higher level of audio similarity than other tracks (even by the same artist). To alleviate this issue, some retrieval systems filter results from the same album or even by the same artist as the seed. Others offer a parameter N to avoid repetitions of the same artist within N consecutive songs, for example, `YouJuke` [5].

Familiarity/Popularity vs. Hotness/Trendiness:
Familiarity or popularity describes how well-known an artist is, whereas hotness or trendiness relates to the amount of buzz or attention an artist is currently getting [1]. Popularity has a more positive connotation than the neutral expression of familiarity. However, we will use the terms interchangeably in the remainder of the paper, likewise the terms hotness and trendiness. According to the temporal dimension, popularity can be seen as a longer lasting property, whereas hotness usually relates to recent appreciation of typically shorter duration, although hot artists might also be very familiar to many people. To give an example, "The Beatles" are certainly popular, whereas "Lady Gaga" currently tends to rank higher on the hotness dimension.

Recentness:
This aspect distinguishes recently released songs from pieces that are older and therefore have a longer (playing) history. In contrast to the aspect of hotness, recentness does not require an artist to be recently popular, just a temporal closeness to the present.

Novelty:

If a music recommender keeps on suggesting tracks/artists known by the user, he or she will not be satisfied, even if the recommended items are perfectly suited otherwise. Hence, presenting recommendations novel for the user is a vital requirement for a recommender system.

Serendipity:

Serendipity means that a user is surprised in a positive way since he discovered an item he did not expect or was not aware of. Being able to make serendipitous recommendations is hence a well desired property for recommender systems [9]. In the context of music retrieval, we believe that the listener's music preference and taste as well as aspects of artist and song popularity have to be taken into account when we aim at providing serendipitous results. For instance, a fan of medieval folk metal might be rather disappointed and bored if the system recommends the band "Saltatio Mortis", which is well known for this music style. In contrast, for a user occasionally enjoying "Metallica" and "Bob Dylan", the former mentioned band may be a serendipitous recommendation.

A SERENDIPITOUS MUSIC RETRIEVAL MODEL

We regard similarity and diversity as orthogonal aspects. A music retrieval system should hence take into account the user's preference to retrieve music items that are similar according to a particular set of aspects (e.g., rhythm and timbre), albeit also ensuring a degree of diversity by including items that are dissimilar according to another set of aspects (e.g., artist name, song lyrics, or tags).

The proposed model for serendipitous music retrieval given a seed/query i, which can either be an artist a or a track t (by artist a)[1], is described by the following retrieval function:

$$r(i) = \min_j \left(\sum_{s \in S} w_s \cdot \delta_s(i,j) - \sum_{d \in D} w_d \cdot (1 - \delta_d(i,j)) \right) \cdot$$
$$(w_p \cdot p(j, reg(l)) + w_h \cdot h(j, reg(l)) + w_r \cdot r(j)) \tag{1}$$

$\delta_s(i,j)$ represents the sth $[0,1]$-normalized dissimilarity function out of the set S of aspects to measure similarity; D accordingly represents diversity aspects. The similarity and diversity aspects can be measured on the track or artist level or take both into account. w_s and w_d are weights that allow to control the importance of each similarity/diversity aspect in S and D, respectively. These weights can be either defined manually by the user, learned online via relevance feedback, or inferred from the user's past listening behavior. Also a mixture of these three strategies seems reasonable. The same holds for the weights w_f, w_h, w_n, and w_r, which control the influence of the factors familiarity, hotness, novelty, and recentness, respectively.

$f(j, reg(l))$ is a measure of familiarity of artist or track j ($[a,t]$), given a particular region $reg(l)$ of the world. This

---

[1]The seed items i and the potential results j might hence be seen as tuples $[a,t]$.

region may be defined by a location l, for instance, given by longitude and latitude values. It can alternatively be given as a political or cultural region. For example, popularity in a particular region can be estimated by analyzing shared folders in Peer-to-Peer networks of users in that region or from page count estimates of search engines [18]. The hotness of an artist or song $h(j, reg(l))$ in a region $reg(l)$ can be inferred from traditional music charts such as the "Billboard Hot 100" [3]. Unfortunately, such charts suffer from two major shortcomings. First, from a global point of view, there are many countries in which such charts are not available. Second, the computation of the rankings in different countries vary in terms of distribution channels included, for instance, online digital music distribution, classical record sales, or airplays. Alternatively, hotness can be inferred from `last.fm` playcounts. Also query strings from Peer-to-Peer networks were shown to relate to recent artist popularity [14].

Here we propose to estimate familiarity and hotness with respect to a given location from music-related microblogging activity. We derived typical listening patterns for 790 major cities of the world by applying natural language processing techniques to geo-localized `Twitter` streams. Using a data set of track and artist names, we then infer time-dependent (hotness vs. familiarity) and location-dependent ($reg(l)$) popularity estimates on the artist or track level. Artist-to-genre assignments, for example gathered from `allmusic.com`, enable the prediction of cultural listening patterns. To this end, the listening data is normalized, aggregated at the level of cities or countries C, and represented via a genre distribution vector for each c ($\in C$), denoted as g_c. Linking a collection of 48,800 artists with the set of `allmusic`'s 18 major genres[2], we can further compute the most "mainstreamy" and the most independent countries in terms of their music listening behavior. To this end, we compute the deviation of g_c from a mean global genre distribution vector. The result is visualized in Figure 1, where the y-axis illustrates the excess or shortfall of each genre in the countries depicted, respective to the global music taste. Hence, a value of 1.5 for a (genre,country)-pair signifies that the genre is listened to 150% more frequently in that country than the global consumption of this genre suggests. Likewise, a value of -1.0 (a shortfall of 100%) refers to the fact that a particular genre is never listened to in the country under consideration.

The novelty $n(j)$ is particularly important for serendipitous recommendation, as already explained above. A straightforward definition may use a binary attribute describing whether or not the user has item j in her collection. This is of course only an approximation for whether a user knows an item, as it neglects various channels of music consumption (e.g., listening to analog disk records or tapes, music streaming (although in this case the streaming provider might offer an API), listening to music at a party or to somebody else's music collection). Nevertheless, this definition may serve as a good proxy for the actual novelty of the music piece under consideration.

---

[2]The used genres are avantgarde (av), blues (bl), celtic (ce), classical (cl), country (co), easylistening (ea), electronica (el), folk (fo), gospel (go), jazz (ja), latin (la), newage (ne), rap (ra), reggae (re), rnb (rn), rock (ro), vocal (vo), and world (wo).

The recentness measure $r(j)$ might be expressed by an exponentially decaying function $e^{-(now-rd(j))}$, where $rd(j)$ is the release date/year of track j. The release date can be gathered from the ID3 tag of an audio file or from music databases such as `MusicBrainz` [2].

CONCLUSIONS AND FUTURE WORK
In this paper, we first outlined requirements for building a serendipitous music retrieval system. More precisely, we elaborated on the factors similarity, diversity, familiarity, hotness, novelty, and recentness. We then presented a retrieval model that takes into account all these factors, and we indicated how measures for the individual factors can be defined.

We are currently developing a prototypical music retrieval system that employs the proposed model. Having ready algorithms for computing the different factors, we aim at applying them on a large scale, using a real-world music repository. To this end, we are experimenting with a set of 2.3 million tracks, for which we compute content-based similarity scores with our top-performing[3] signal-based similarity algorithm [19]. As an alternative to content-based similarity, we construct a text-based similarity measure from `last.fm` tags, web pages about music artists [17], and microblogs [16]. Each of these measures can be used as a similarity or diversity function. Familiarity and hotness are derived using social media mining techniques. Recentness is computed using editorial metadata provided by the record labels. Novelty will be assessed in our foreseen retrieval system using a hybrid approach that takes into account several indicators to predict whether a piece is known to a user or not. Such indicators include listening information from music streaming services, such as `last.fm` or `Spotify`, existence of the track in the user's digital music collection, and mentions of the track or artist in the user's microblog posts.

Particularly challenging will be the evaluation of the system as it requires assessing the user's satisfaction in various dimensions and according to different usage scenarios. We might elaborate evaluation experiments based upon some ideas from [21]. Even though validation of the proposed system will be challenging, we are confident that this work will help bring the vision of serendipitous music retrieval a bit closer to reality.

ACKNOWLEDGMENTS

REFERENCES
1. http://musicmachinery.com/2009/05/25/artist-similarity-familiarity-and-hotness (access: Sep 2011).

2. http://musicbrainz.org (access: Oct 2011).

3. http://www.billboard.com/charts/hot-100 (access: Oct 2011).

4. http://www.music-ir.org/mirex (access: Sep 2011).

5. http://www.youjuke.org (access: Oct 2011).

6. Aucouturier, J.-J., and Pachet, F. Scaling Up Music Playlist Generation. In *Proc. IEEE ICME* (Aug 2002).

7. Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Lüke, K.-H., and Schwaiger, R. InCarMusic: Context-Aware Music Recommendations in a Car. In *Proc. EC-Web* (Aug–Sep 2011).

8. Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., and Slaney, M. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proc. IEEE 96* (Apr 2008).

9. Celma, O. *Music Recommendation and Discovery in the Long Tail*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2008.

10. Knees, P., Pohle, T., Schedl, M., and Widmer, G. Combining Audio-based Similarity with Web-based Data to Accelerate Automatic Music Playlist Generation. In *Proc. ACM SIGMM MIR* (Oct 2006).

11. Lïbeks, J., and Turnbull, D. Exploring "Artist Image" using Content-based Analysis of Promotional Photos. In *Proc. ICMC* (Jun 2010).

12. M. Schedl et al. Towards Automatic Retrieval of Album Covers. In *Proc. ECIR* (Apr 2006).

13. McFee, B., and Lanckriet, G. Heterogeneous Embedding for Subjective Artist Similarity. In *Proc. ISMIR* (Oct 2009).

14. N. Koenigstein et al. Spotting Out Emerging Artists Using Geo-Aware Analysis of P2P Query Strings. In *Proc. ACM SIGKDD* (Aug 2008).

15. Pohle, T., Knees, P., Schedl, M., Pampalk, E., and Widmer, G. "Reinventing the Wheel": A Novel Approach to Music Player Interfaces. *IEEE Transactions on Multimedia 9* (2007).

16. Schedl, M., Knees, P., and Böck, S. Investigating the Similarity Space of Music Artists on the Micro-Blogosphere. In *Proc. ISMIR* (Oct 2011).

17. Schedl, M., Pohle, T., Knees, P., and Widmer, G. Exploring the Music Similarity Space on the Web. *ACM Transactions on Information Systems 29*, 3 (Jul 2011).

18. Schedl, M., Pohle, T., Koenigstein, N., and Knees, P. What's Hot? Estimating Country-Specific Artist Popularity. In *Proc. ISMIR* (Aug 2010).

19. Seyerlehner, K., Schedl, M., Pohle, T., and Knees, P. Using Block-Level Features for Genre Classification, Tag Classification and Music Similarity Estimation. In *Extended Abstract to MIREX/ISMIR* (Aug 2010).

20. Whitman, B., Flake, G., and Lawrence, S. Artist Detection in Music with Minnowmatch. In *Proc. IEEE NNSP* (Sep 2001).

21. Yuan Cao Zhang, Diarmuid O Seaghdha, Daniele Quercia, Tamas Jambor. Auralist: Introducing Serendipity into Music Recommendation. In *Proc. WSDM* (Feb 2012).

[3]Our approach ranked first in the MIREX 2010 and 2011 task of *Audio Music Similarity and Retrieval*.
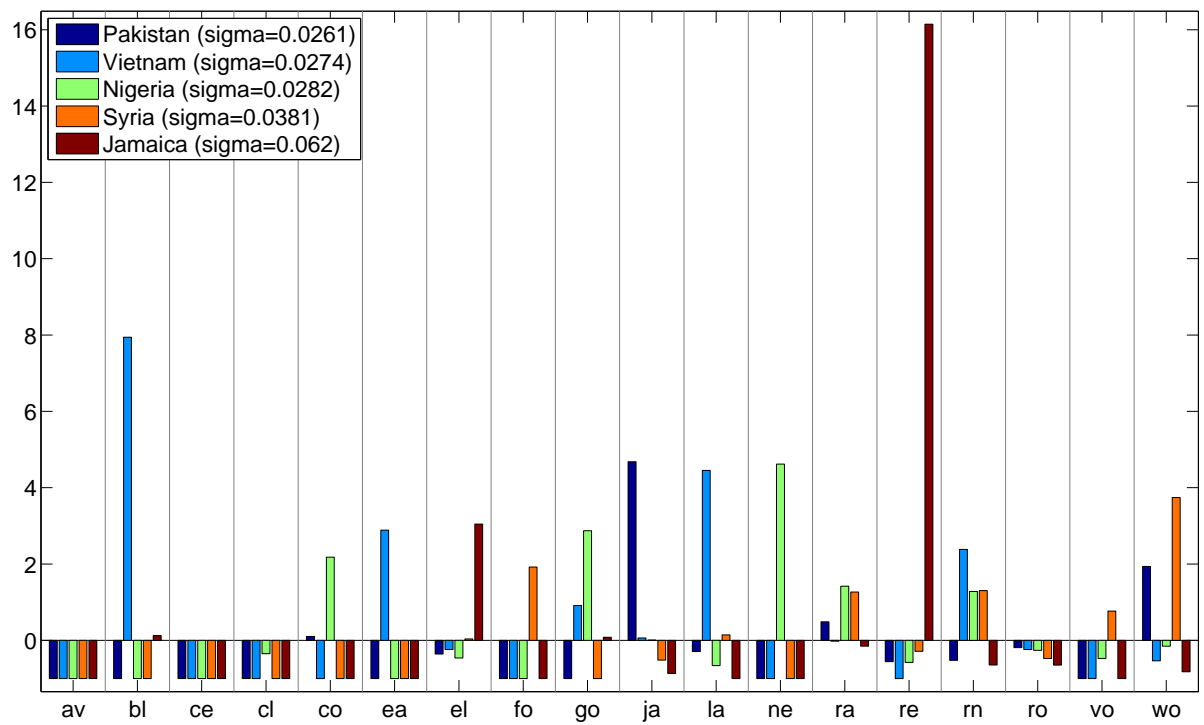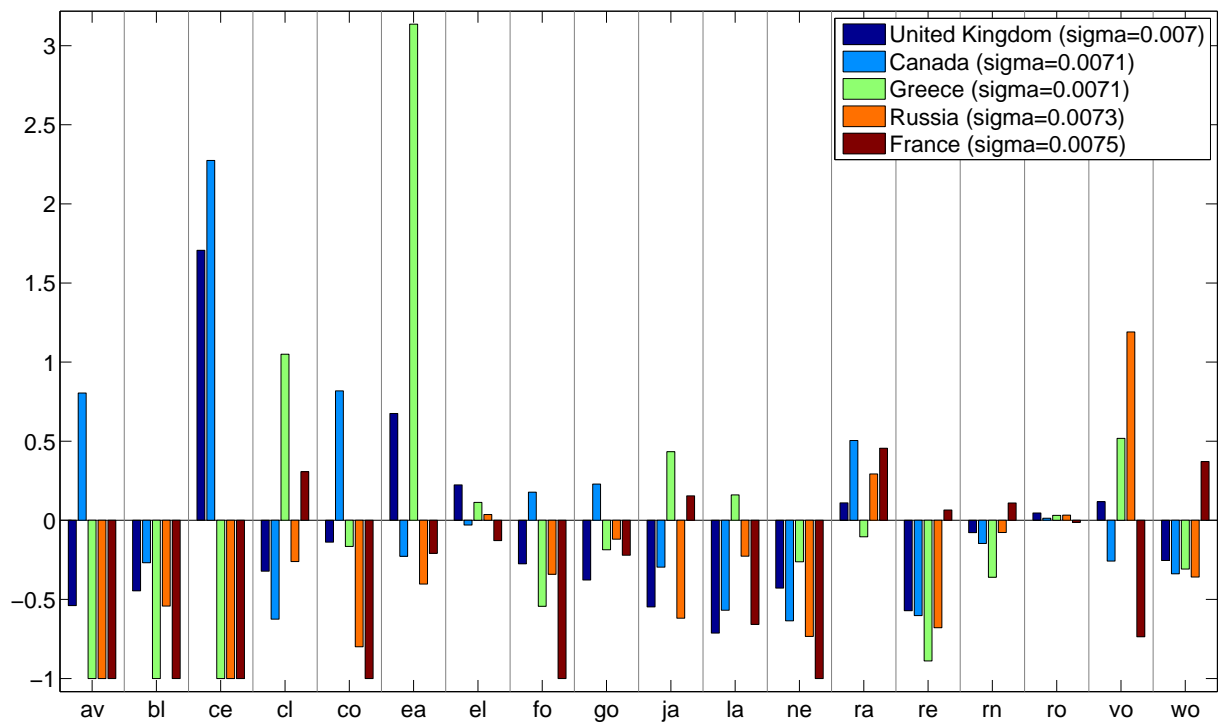
Figure 1. Relative deviations from mean global genre distribution for countries with most (top) and least (bottom) representative listening behavior.