

Taking Advantage of Editorial Metadata to Recommend Music

Dmitry Bogdanov and Perfecto Herrera

Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{dmitry.bogdanov,perfecto.herrera}@upf.edu

Abstract. In this work we propose a novel approach to music recommendation based exclusively on editorial metadata. To this end, we propose to use a public database of music releases *Discogs.com*, which contains extensive information about artists, their releases and record labels. We rely on an explicit set of music tracks provided by the user as evidence of his/her music preferences to construct a user profile suitable for distance-based music recommendation. We evaluate the proposed method against two purely metadata-based approaches and one approach partially based on audio content in a listening experiment with 27 participants. The results of subjective evaluation show that the proposed method is competitive to the state-of-the-art recommenders based on commercial metadata, while being easily implemented relying only on open public data.

Keywords: Music recommendation, user modeling, music similarity, editorial metadata, subjective evaluation

1 Introduction

The amount of music available digitally has overwhelmingly increased during the last decade following the growth of the Internet and music technology developments. Nowadays vast amounts of music are available for listeners' access, but still finding relevant and novel music is often a difficult task for them. Thereby, music listeners and music scholars strive for better recommendation systems to facilitate music search and retrieval.

In this context, music recommendation is a challenging topic in the Music Information Research (MIR) community. The state-of-the-art approaches to music recommendation are based on measuring music similarity between artists or particular tracks, and on user profiling, eliciting the information about music preferences. To this end, both metadata and audio content information can be used. Considering metadata, the state-of-the-art approaches to recommend music exploit user ratings, consumption and listening history, which are commonly used for collaborative filtering, and social tags extracted from social tagging services for music such as *Last.fm*¹ or mined from the web pages related to music content [1–6]. Current metadata-based approaches can perform satisfactorily

¹ <http://last.fm>

for listeners' needs when dealing with popular music. However, such approaches have disadvantages. Firstly, due the long-tail problem [2], a system may not have sufficient and correct metadata information for unpopular items. This can significantly limit the quality and the scope of recommendations or even make them completely impossible. Secondly, such approaches are cold-start prone and costly to maintain, requiring a large amount of user ratings, consumption or listening behavior to be processed for collaborative filtering, or large databases of tags. This information is expensive to obtain and maintain, and, moreover, is generally proprietary.

Alternatively, audio content information, extracted from the raw audio signal, can be applied for music recommendation. Such approaches are able to achieve performance close, or even comparable, to successful metadata-based approaches in terms of the relevance of recommendations [7–9], avoiding the problem of the long tail. Nevertheless, they are computationally costly and thus they require a large effort to build and maintain large-scale music collections.

Concerning user profiling, there exist approaches based on user models, which employ classification into interest categories using content-based information [10–13] or hybrid sources [14]. As well, distance-based² approaches, starting from a set of preferred items in a content-based vector space [15, 9], or more complex hybrid probabilistic approaches [16, 17] are proposed.

In the present work, we focus on distance-based music recommendation approaches. Moreover we consider a passive scenario, when recommendations are provided based on knowledge of user preferences rather than on manual user-specified query-by-example. We aim for a lightweight approach suitable for large-scale music collections, in particular containing the long-tail of artists and tracks, while working with publicly available data.

We propose a novel recommendation approach which is based exclusively on editorial metadata. To this end, we propose to use a public database of music releases, *Discogs.com*,³ which contains extensive user-built information on artists, labels, and their recordings. We rely on an explicit set of music tracks provided by a user as evidence of his/her music preferences, the henceforth called “*preference set*”. We construct a user profile suitable for distance-based music recommendation using editorial metadata about the artists from the user’s preference set. More concretely, for each artist we retrieve a descriptive tag cloud, containing information about particular genres, styles, record labels, years of release activity, and countries of release fabrication. We then employ latent semantic analysis [18] to compactly represent each artist as a vector, and match the user’s preference set to a music collection to produce recommendations. We evaluate the proposed approach together with a number of baseline approaches in terms of subjective satisfaction ratings and calculated amount of novel relevant and known trusted recommendations on real listeners.

This paper is organized as follows: In Section 2 we describe the considered approaches. Firstly, the proposed approach working exclusively on editorial

² We pragmatically refer to any music similarity measure with the term “distance”.

³ <http://discogs.com>

metadata (Section 2.1). Secondly, a hybrid baseline approach, which employs content-based semantic distance followed by a simple genre refinement. Thirdly, a metadata-based baseline approach, working on artist tag annotations obtained from the *Last.fm*⁴ social music service. Fourthly, a state-of-the-art commercial recommender on the example of *iTunes Genius*,⁵ which relies on a collaborative filtering information. All three baseline approaches are described in Section 2.2. In Section 3 we present the subjective evaluation of the considered approaches conducted on 27 participants. Section 3.1 provides the characterization of subjects, while Section 3.2 explains the listening experiment instructions, stimuli and procedure. The evaluation results are presented and discussed in Section 3.3. Finally, we conclude this study in Section 4.

2 Studied Approaches

The approaches considered in this work are distance-based. We focus on the use-case of passive recommendations based on user preferences similarly to our previous works [9, 19]. Therefore, the approaches provide track recommendations from a given music collection (the henceforth called *music collection*) starting from a set of tracks, given by the user as evidence of her/her music preferences (a *preference set*), and applying distance measures between the tracks in preference set and the tracks in music collection. To create a preference set, the user is asked to provide a minimal set of music tracks, which she/he believes to be sufficient to grasp or convey her/his music preferences. The tracks can be submitted solely using the essential editorial metadata sufficient to identify them and, additionally, in audio format. The editorial metadata and audio for all provided tracks is then retrieved, if missing. Metadata is cleaned by means of tag cleaning and audio fingerprinting software MusicBrainz.⁶ Thus, we obtain both editorial metadata and audio content for each track from the user's preference set which are suitable to apply both metadata-based and content-based analysis and recommendation procedures.

We employed a large in-house music collection as the source for recommendations. This collection covers a wide range of genres, styles, and arrangements, containing 68K music excerpts (30 sec.) by 16K artists with a maximum of 5 tracks per artist. For consistency, in our experiments we require each of the recommendation approaches to output 15 tracks by different artists (1 track per artist) not being present among the artists in the user's preferences set. To this end, each approach includes an artist filter.

⁴ All tags were obtained on March, 2011.

⁵ <http://www.apple.com/itunes/features/> all experiments were conducted using iTunes 10.3.1 on December, 2011.

⁶ http://musicbrainz.org/doc/MusicBrainz_Picard

2.1 Proposed Approach: Artist Similarity Based On Editorial Metadata (M-DISCOGS)

The approach we proposed works exclusively on editorial metadata found in the *Discogs.com* database. The dump of this database is released under the Public Domain license⁷, which makes it useful for different music applications, and in particular for research purposes of the MIR community. While there exist similar music services, such as public *MusicBrainz*⁸ database, or proprietary *Last.fm* or *AllMusic*⁹, we opt for *Discogs* as it contains the largest catalog of music releases and artists, while being known for accurate moderated metadata, which includes comprehensive annotations of particular releases.

The database contains the extensive information about up to 2,848K releases, 2,195K artists, and 281K labels.¹⁰ In particular, for each artist this information includes a list of aliases, members (in the case an artist is a group), and group memberships (in the case an artist is a single person). Moreover it contains a list of releases authored by the artist, including albums, singles and EPs, and a list of appearances on the releases headed by other artists or compilations. A release corresponds to a particular edition of an album, single, EP, etc., and the releases related to the same album, single, or EP, can be grouped together into a “master release”. Each release contains genre, style, country and year information. Genres are broad categories (such as classical, electronic, funk/soul, jazz, rock, etc.) while styles are more specific categories (such as neo-romantic, tech house, afrobeat, free jazz, viking metal, etc.) In total the database counts up to 15 genre categories and 329 styles.

For each artist in the database¹¹ we create a tag-cloud using genre, style, label, country, and year information related to this artist. To this end, we retrieve three lists of releases (*MAIN*, *TRACK*, *EXTRA*), where the artist occurs as (1) main artist, heading the release, (2) track artist, for example being on a compilation or with a guest appearance on a release, (3) extra artist, being mentioned in the credits of a release (usually related to the activity such as remixing, performing, writing and arrangement, production, etc.).

For each found release related to the artist, we retrieve genre, style, label, country, and year tags. For each of the three lists, we merge releases accordingly to their master release, keeping the genres, styles, and countries, which are present in at least one of the releases (i.e., applying a set union). Concerning the release years, we attempt to approximate the authentic epoch, when the music was firstly recorded, produced, and consumed. As a master release can contain reissues along with original releases, we keep the earliest (the original) year and, moreover, propagate it with descending weights as following:

$$W_{y\pm i} = W_y * 0.75^i, i \in \{1, 2, 3, 4, 5\} \quad (1)$$

⁷ <http://www.discogs.com/data/>

⁸ <http://musicbrainz.org>

⁹ <http://www.allmusic.com>

¹⁰ As on January 3, 2012.

¹¹ In our experiments, we used a *Discogs* monthly dump dated by January, 2011.

where W_y is the original year y , and 0.75 is a decay coefficient. For example, if the original year “y” is 1995, the resulting year-tag weights will be $W_{1995} = 1.0$, $W_{1994} = W_{1996} = 0.75$, $W_{1993} = W_{1997} \approx 0.56$, $W_{1992} = W_{1998} \approx 0.42$, $W_{1991} = W_{1999} \approx 0.32$, $W_{1990} = W_{2000} \approx 0.24$.

Thereafter, we summarize *MAIN*, *TRACK*, and *EXTRA* lists of the artist to a single tag-cloud. We assume a greater importance of tag annotations for the main artist role in comparison to track artists or extra artists; e.g., tags found on an artist’s album are more important than the ones found on a compilation. We empirically assign the weights to these three groups of artist roles: 1.0 for main artists and 0.5 for both track and extra artists. As well, we assign further weights to tags according to their category: 1.0 for genres, styles, and labels, and 0.5 for years and countries, rescaling the artist tag-cloud. In particular, we decided to give equal importance to label information as to genres and styles. The rationale behind grounds on the hypothesis that record label information gives a very valuable clue to a type of music, especially in the long-tail for the case of niche labels.

Finally, we propagate artist tags using the artist relations found in the database, such as aliases and membership relations. We suppose related artists to share similar musical properties and, therefore, assure that artists with low amount of releases will obtain reasonable amount of tags. To this end, for each artist we add a set of weighted tag-clouds of all related artists to the associated tag-cloud. We select a propagation weight of 0.5 and apply only 1-step propagation; i.e. tags will be propagated only between artists sharing a direct relation. Figure 1 presents an example of the proposed annotation procedure.

Following the described procedure we are able to construct tag-clouds for each artist in the *Discogs* database which together form a sparse tag matrix. To simplify this matrix, for each artist we apply additional filtering by means of erasing the tags with weight less than 1% of the artist’s tag with the maximum weight. We then apply latent semantic analysis [18, 20, 21] to reduce the dimensionality of the obtained tag matrix to 300 latent dimensions. Afterwards, Pearson correlation distance [22, 2] can be applied on the resulting topic space to measure similarity between artists.

Once we have matched the annotated artists to the tracks in our music collection and the user’s preference set, we retrieve recommendations applying the tag-based distance by the following procedure. For each track X in the user’s preference set (a recommendation source), we apply this distance to retrieve the closest track C_X (a recommendation outcome candidate) from the music collection and form a triplet $(X, C_X, distance(X, C_X))$. We sort the triplets by the obtained distances, delete the duplicates of the recommendation sources (i.e., each track from the preference set produces only one recommendation outcome), and apply an artist filter. We return, as recommendations, the recommendation outcome candidates from the top 15 triplets. If it is impossible to produce 15 recommendations due to the small size of the preference set (less than 15 tracks) or because of the applied artist filter, we increase the number of possible recommendation outcome candidates per recommendation source.

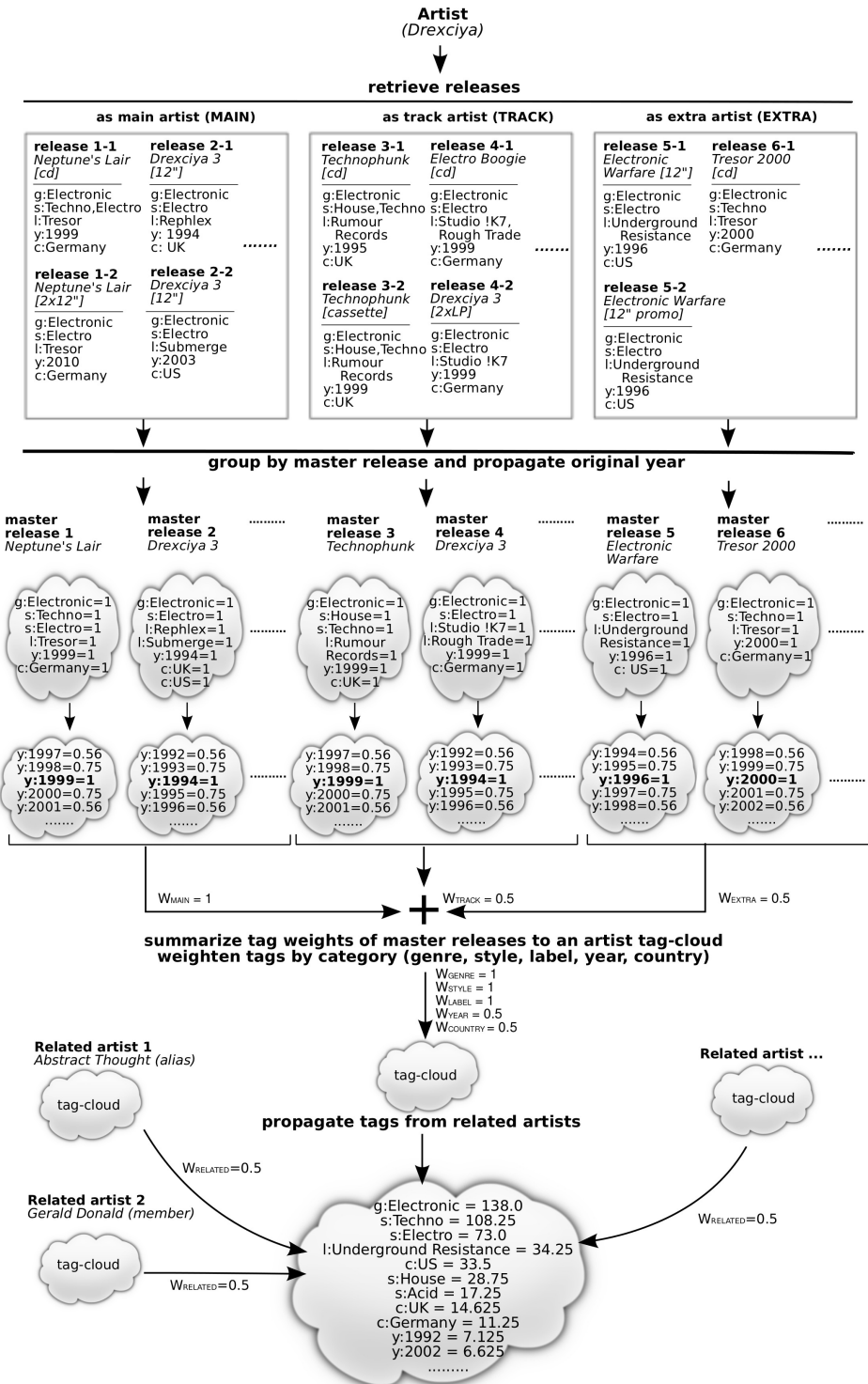


Fig. 1. An example of the proposed artist annotation based on editorial metadata from *Discogs*. Three lists of releases (MAIN, TRACK, EXTRA) are retrieved according to an artist's role. Particular releases are summarized into master releases, merging all found genre, style, label, and country tags, and selecting and propagating original year. Thereafter, tags are weighted to form a tag-cloud of an artist, and summed with the propagated tags of all related artists. Letters "g," "s," "l," "y," "c" stand for genre, style, label, year, and country, respectively.

Pseudo-code of the distance-based recommendation procedure.

```

set IGNORE_ARTISTS to artists in preference set
remove tracks by IGNORE_ARTISTS from music collection
set N_OUTCOMES to 1
set N_RECS to 15

while true:
  set POSSIBLE_RECS to an empty list
  for track X in preference set:
    set X_NNS to N_OUTCOMES closest to X tracks in music collection
    for track C_X in X_NNS:
      append triple(X,C_X,distance(X,C_X)) to POSSIBLE_RECS
  sort POSSIBLE_RECS by increasing distance

  set RECS to an empty list
  for triple(SOURCE,OUTCOME,DISTANCE) in POSSIBLE_RECS:
    if OUTCOME occurs in RECS:
      next iteration
    if SOURCE occurs in RECS >= N_OUTCOMES times:
      next iteration
    append triple (SOURCE,OUTCOME,DISTANCE) to RECS
    if length of RECS list is N_RECS:
      return outcomes from RECS as recommendations
  set N_OUTCOMES to N_OUTCOMES + 1

```

2.2 Baseline Approaches

Content-based Semantic Similarity Refined By Genre Metadata (C-SEM+M-GENRE). As our first baseline, we consider a content-based semantic measure, providing a distance between tracks, filtered by genre metadata. The research presented in [19] has shown that simple filtering by a single genre tag can significantly improve the performance of a content-based-only approach to recommendation. Meanwhile, such genre information is considerably cheap to gather and maintain, it is however sufficiently descriptive for effective filtering.

We employ a semantic distance working on a set of high-level semantic descriptors (genres, musical culture, moods, instrumentation, rhythm, and tempo) inferred by support vector machines (SVMs) from low-level timbral, temporal, and tonal features. This distance has already been evaluated in the context of music similarity and music recommendation based on preference sets [23, 9, 24]. We refer the interested reader to the afore-cited literature for the implementation details of this measure and the evaluation results.

The semantic distance is applied similarly to the above-mentioned procedure for the M-DISCOGS, but in conjunction with a simple filtering: only the tracks of the same genre labels are considered as possible recommendation outcomes. We reproduce genre filtering as described in [19].

Artist Similarity Based On Last.fm Tags (M-TAGS) We consider a purely metadata-based similarity measure working on the artist level. This approach is based on social tags provided by the *Last.fm* API, retrieved for the artists from the user’s preference set and the music collection. Using the API, we obtain a weight-normalized tag list for each artist. The weight ranges in the $[0, 100.0]$ interval, and we select a minimum weight threshold of 10.0 to filter out possibly inaccurate tags. The resulting tags are then assigned to each track in the preference set and the music collection. We then apply latent semantic analysis [18, 20] to reduce dimensionality to 300 latent dimensions. Pearson correlation distance [2] can be applied on the resulting topic space. We retrieve recommendations following the same procedure as for the M-DISCOGS.

Black-box Similarity By iTunes Genius (M-GENIUS) We consider commercial black-box recommendations obtained from the *iTunes Genius* playlist generation algorithm similarly to [8, 19]. Given a music collection and a query, this algorithm is capable to generate a playlist by means of an undisclosed underlying music similarity measure. It works on metadata and partially employs collaborative filtering of large amounts of user data (music sales, listening history, and track ratings) [8].

We randomly select 15 tracks, which are recognizable by *GENIUS*, from the user preference set. For each of the selected tracks (a recommendation source), we generate a playlist, apply the artist filter, and select the top track as the recommendation outcome. We increase the amount of possible outcomes per source when it is impossible to produce 15 recommendations similarly to the M-DISCOGS.

3 Evaluation

3.1 Subjects

We asked 27 voluntary subjects (selected from the authors’ colleagues, their acquaintances and families) to provide their respective preference sets. Moreover, additional information was gathered, including personal data (gender, age, interest for music, musical background), and a description of the strategy and criteria followed to select the music pieces. The participants were not informed about any further usage of the gathered data, such as giving music recommendations.

The age of participants varied between 19 and 46 years ($\mu = 31.22$, $\sigma = 5.57$). All participants showed a very high interest in music (rating with $\mu = 9.43$ and $\sigma = 0.91$, where 0 means no interest and 10 means passionate). In addition, 24 participants play at least one musical instrument. The number of tracks selected by the participants to convey their musical preferences was very varied. It ranged from 8 to 178 music pieces ($\mu = 51.41$, $\sigma = 38.38$) with the median being 50 tracks. The time spent on creating a preference set differed a lot as well, ranging from 12 minutes to 60 hours ($\mu = 8.21$, $\sigma = 16.55$) with the median being two hours. The strategy followed by the participants to gather preference sets

also varied. Driving criteria for the selection of tracks included musical genre, mood, uses of music (listening, dancing, singing, playing), expressivity, musical qualities, and chronological order. We expect our population to represent a wide range of music enthusiasts, considering this information.

3.2 Evaluation Methodology

In general, evaluation of music recommender systems is a complicated, and, so far, not standardized procedure. Existing research works on music recommendation involving evaluations with real participants [10, 1, 25, 7, 8, 26] are significantly limited in the tradeoff condition between the number of participants or by the number of evaluated tracks per approach by a particular user. Moreover, they are often focused on perceived quality of music similarity measures instead of user satisfaction with recommendations. In the latter case, evaluations generally include only one measure of user satisfaction, while the familiarity factor is rarely considered.

We describe our methodology of the conducted evaluation. Participants were asked to perform a blind subjective listening evaluation of the music generated using the 4 different recommendations approaches. To generate recommendations, we used our in-house music collection described in Section 2. For each participant, starting from her/his preference set, four recommendation playlists were generated by four respective approaches. Each playlist consisted of 15 tracks, and never contained more than one track by the same artist, nor contained tracks by artists from the preference set, due to the applied artist filter. All four playlists were then merged, randomized, and their filenames and metadata anonymized, and presented to a participant. This allowed to avoid any response bias due to presentation order, recommendation approach, or contextual recognition of tracks (e.g., by artist names). Furthermore, the participants were not aware of the amount of recommendation approaches, their names and rationales.

To gather feedback on recommendations, we provided a questionnaire for the subjects to express their subjective impressions related to the recommended music. To this end, we used four rating scales, following our previous works [9, 19]: A “*familiarity*” rating ranged from the identification of artist and title (4) to absolute unfamiliarity (0), with intermediate steps for knowing the title (3), the artist (2), or just feeling familiar with the music (1). A “*liking*” rating measured the enjoyment of the presented music with 0 and 1 covering negative liking, 2 being a kind of neutral position, and 3 and 4 representing increasing liking for the musical excerpt. A rating of “*listening intentions*” measured preference, but in a more direct and behavioral way than the “*liking*” scale, as an intention is closer to action than just the abstraction of liking. Again this scale contained 2 positive and 2 negative steps plus a neutral one. Finally, an even more direct rating was included with the name “*give-me-more*” allowing just 1 or 0 to respectively indicate a request for, or a reject of, more music like the one presented. We also asked users to provide title and artist for those tracks rated high in the familiarity scale. The textual meaning of the ratings was presented to the participants together with the allowed rating values.

3.3 Evaluation Results

First, we manually corrected the familiarity rating when the artist/title, provided by the participant, was incorrect. Hence, a familiarity rating of “3” or, more frequently, “4”, was sometimes lowered to 1 or 2. These corrections represented 4.5% of the total familiarity judgments.

The four gathered ratings can be used to characterize different aspects of the considered recommendation approaches. We expect a good recommender system to provide high liking, listening intentions, and “give-me-more” ratings. Moreover, if we focus on music discovery, low familiarity ratings are desired, which will guarantee the novelty of relevant (liked) recommendations. Following [9,19], we recoded the participants’ ratings for each evaluated track into three categories which refer to the type of the recommendation: *hits*, *fails*, and *trusts*. We defined a recommended track to be a hit when it received low familiarity ratings (< 2) and high liking (> 2), listening intentions (> 2), and “give-me-more” ($= 1$) ratings simultaneously. Similarly, trusts were the tracks with high liking, listening intentions, “give-me-more”, but as well high familiarity (> 1). Trusts, provided their overall amount is low, can be useful for a user to feel that the recommender is understanding his/her preferences [8] (i.e., a user could be satisfied by getting a trust track from time to time, but annoyed if every other track is a trust). Fails were the tracks which received low liking (< 3), listening intentions (< 3) and “give-me-more” ($= 0$) ratings. In any other case (e.g., a track received high liking, but low listening intentions and “give-me-more”) the outcome category was considered to be “unclear”, amounting to 17.3% of all recommendations.

We report the percent of hit, fail, trust, and unclear outcomes per recommendation approach in Table 1. According to the results of a chi-square test, an association between the approaches and the outcome categories ($\chi^2(9) = 46.879$, $p < 0.001$) can be accepted. Namely, certain approaches provide hits, fails or trust percents which are statistically different than what a flat distribution (i.e., equiprobable) would yield.

In general, the proposed M-DISCOGS approach performed well comparing to the baselines. The M-DISCOGS provided a considerably low (34.4%) amount of fails, being in between of the metadata-based baselines M-TAGS (with the lowest amount of fails, 32.8%) and M-GENIUS. In contrast, the C-SEM+M-GENRE approach, which is partially content-based, provided the largest (over 41%) amount of fails. Considering hits, the M-TAGS (38.8%) and C-SEM+M-GENRE (37.9%) are the leaders followed by M-GENIUS, and lastly, the M-DISCOGS. That is, our proposed approach provided the least amount of novel relevant recommendations (31.9%). Nevertheless this fact is compensated by the largest amount of trusts, gathered by the M-DISCOGS (16.4%) followed by the M-GENIUS (13.2%), M-TAGS, and the C-SEM+M-GENRE (4.4%). The amount of unclear recommendations ranged as well. As such recommendations consisted of the tracks with inconsistent ratings, we may not expect such tracks to be as relevant as hits and trust categories. Still, such tracks can be useful for certain scenarios (e.g., playlist generation), but are probably not well suited for others (e.g., digital music vending). Considering the extreme case, when

Table 1. Percent of fail, trust, hit, and unclear categories per recommendation approach.

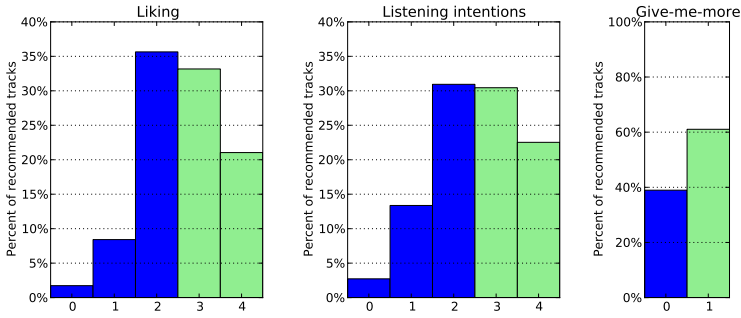
Approach	fail	hit	trust	unclear	hit+trusts
M-TAGS	32.8	38.8	7.4	21.0	46.2
M-DISCOGS	34.4	31.9	16.4	17.3	48.3
M-GENIUS	36.2	35.7	13.2	14.9	48.9
C-SEM+M-GENRE	41.6	37.9	4.4	16.1	42.3

Table 2. Mean ratings per recommendation approach.

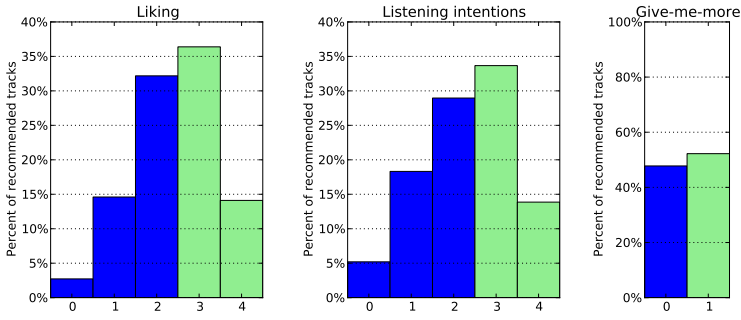
Approach	liking	listening intentions	give-me-more	familiarity
M-DISCOGS	2.63	2.57	0.63	0.83
M-GENIUS	2.60	2.50	0.59	0.80
M-TAGS	2.52	2.45	0.63	0.49
C-SEM+M-GENRE	2.45	2.33	0.52	0.37

fails and unclear categories are both unwanted outcomes, the metadata-based M-GENIUS and M-DISCOGS result as approaches with the least amount of unwanted recommendations (51.1% and 51.7%, respectively), followed by the M-TAGS, and lastly by the partially content-based C-SEML+M-GENRE approach (57.7%). In contrast, considering trusts and hits as wanted outcomes, the M-GENIUS and M-DISCOGS provide their largest amount (48.9% and 48.3%, respectively), followed by the M-TAGS and C-SEM+M-GENRE.

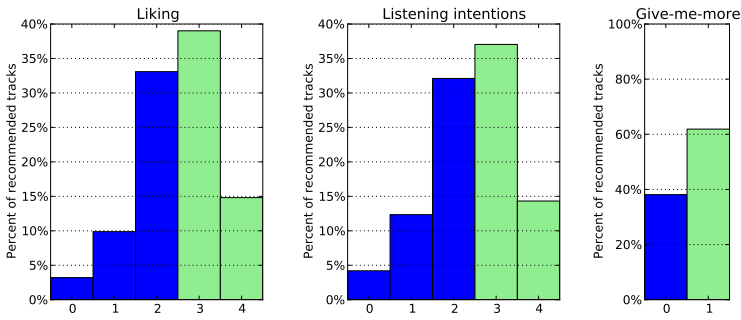
Apart from analysis of the outcome categories, we tested the effect of the recommendation approaches on the liking, listening intentions, and “give-me-more” subjective ratings. To this end, we conducted three separate between-subjects ANOVAs. Tested approaches were shown to have an impact on these ratings ($F(3, 1612) = 3.004$, $p < 0.03$ for the liking rating, $F(3, 1612) = 3.660$, $p < 0.02$ for the intentions rating, and $F(3, 1612) = 3.363$, $p < 0.02$ for the “give-me-more” rating). Pairwise comparisons using Tukey’s test revealed differences only between M-DISCOGS vs C-SEM+M-GENRE for the case of all three ratings, and, in addition, a difference between M-TAGS vs C-SEM+M-GENRE in the case of the “give-me-more” rating. In Figure 2 we present the histograms for the liking, listening intentions, and “give-me-more” ratings. Mean values of these ratings are provided in Table 2. Inspecting the means, we see that all considered approaches performed with a user satisfaction slightly above average. Almost half of the provided recommendations were favorably evaluated, i.e., received high liking and listening intentions ratings (> 2) and a positive “give-me-more” request. An inspection of histograms shows that the proposed M-DISCOGS approach receives the highest amount of maximum ratings for liking and listening intentions ($\simeq 21\%$ and $\simeq 22.5\%$, respectively). In contrast, the amount of received negative ratings is lower. Still, returning to the ANOVA results, the only clear difference in performance, as measured by our 3 indexes, happens between M-



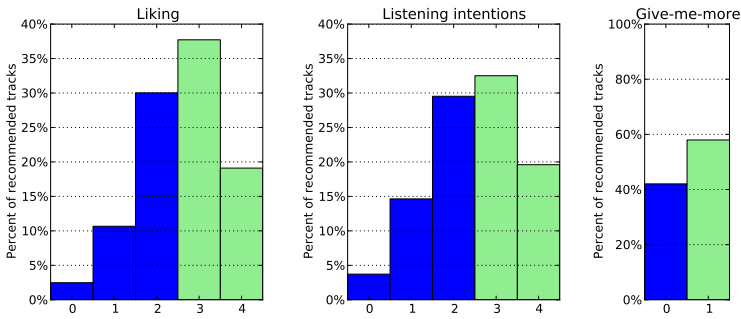
(a) M-DISCOGS



(b) C-SEM+M-GENRE



(c) M-TAGS



(d) M-GENIUS

DISCOGS and C-SEM+M-GENRE. In other words, the proposed M-DISCOGS approach is able to achieve similar liking, listening intentions and willingness to get recommended music as existing (and commercial) state-of-the-art systems. Interestingly, we have evidenced the above-average ceiling in the performance of all considered approaches. This fact highlights a lot of room for improvement of music recommender systems.

4 Conclusions

We have considered and evaluated different distance-based approaches to music recommendation, starting from a set of music tracks explicitly provided by a user as an evidence of his/her musical preferences. We proposed a novel approach working exclusively on editorial metadata taken from publicly available music database, *Discogs.com*. Relying on user-built information about music releases present in this database, we demonstrated how this information can be applied to create descriptive tag-based artist profiles, containing information about particular genres, styles, record labels, years of release activity, and countries. Furthermore, to overcome the problem of tag sparsity, such artist profiles can be compactly represented as vectors in a latent semantic space of reduced dimension. Applying a distance measure between the resulting artist vectors for the tracks in the preference set of a user and the tracks within a music collection, we are able to generate recommendations.

The proposed approach has a number of advantages over common metadata-based approaches. Firstly, our approach is able to provide a compact profile for each artist found in *Discogs* database. Matching these profiles to music collections, large-scale recommendation systems can be built. Secondly, the proposed approach is based only on open public data, meanwhile the majority of successful recommender systems operate on commercially withheld metadata. As a consequence, our approach is easy to create and reproduce. Subjective evaluation of the proposed approach with 27 participants demonstrated performance comparable to the state-of-the-art metadata-based approaches, including an industrial recommender. In particular, our approach provided large amount of trusted and novel relevant recommendations, which suggests that the proposed approach is well suited for music discovery and playlist generation. Although we have considered and evaluated the proposed approach in the context of “passive discovery”, relying on preference sets provided by listeners, we expect our conclusions to be applicable for the query-by-example use-case.

Interestingly, the evaluated content-based approach filtered by simple genre metadata revealed performance comparable to the metadata-based approaches as well. In our previous research [9, 19], we evidenced a high number of trusted recommendations for the metadata-based approaches, and fewer in the case of content-based recommendations similarly to the present study. Moreover, we similarly evidenced the user satisfaction by the evaluated metadata-based approaches to be slightly above average showing a lot of room for improvement.

Future work will be focused on the limitations of the current proof-of-concept study. A number of parameters were chosen empirically in the proposed approach and will require further research to find optimal weights for the release types, tag types, and artist propagation as well as the year propagation decay. Moreover, a hybrid approach expanding the proposed method with audio content information will be of interest.

Acknowledgments. The authors would like to thank all participants involved in the evaluation. This research has been partially supported by the FI Grant of Generalitat de Catalunya (AGAUR) and the Classical Planet (TSI-070100-2009-407, MITYC), DRIMS (TIN2009-14247-C02-01, MICINN), and MIREs (EC-FP7 ICT-2011.1.5 Networked Media and Search Systems, grant agreement No. 287711) projects.

References

1. Firan, C.S., Nejdil, W., Paiu, R.: The benefit of using tag-based profiles. In: Latin American Web Conf. (2007) 32–41
2. Celma, O.: Music recommendation and discovery in the long tail. PhD thesis, UPF, Barcelona, Spain (2008)
3. Baltrunas, L., Amatriain, X.: Towards time-dependant recommendation based on implicit feedback. In: Workshop on Context-aware Recommender Systems (CARS'09). (2009)
4. Jawaheer, G., Szomszor, M., Kostkova, P.: Comparison of implicit and explicit feedback from an online music recommendation service. In: Int. Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec'10), New York, NY, USA, ACM (2010) 47–51 ACM ID: 1869453.
5. Levy, M., Bosteels, K.: Music recommendation and the long tail. In: ACM Conf. on Recommender Systems. Workshop on Music Recommendation and Discovery (Womrad 2010). (2010)
6. Schedl, M., Pohle, T., Knees, P., Widmer, G.: Exploring the music similarity space on the web. *ACM Trans. on Information Systems* **29** (2011) 1–24
7. Magno, T., Sable, C.: A comparison of signal-based music recommendation to genre labels, collaborative filtering, musicological analysis, human recommendation, and random baseline. In: Int. Conf. on Music Information Retrieval (ISMIR'08). (2008) 161–166
8. Barrington, L., Oda, R., Lanckriet, G.: Smarter than genius? human evaluation of music recommender systems. In: Int. Society for Music Information Retrieval Conf. (ISMIR'09). (2009) 357–362
9. Bogdanov, D., Haro, M., Fuhrmann, F., Gómez, E., Herrera, P.: Content-based music recommendation based on user preference examples. In: ACM Conf. on Recommender Systems. Workshop on Music Recommendation and Discovery (Womrad 2010). (2010)
10. Hoashi, K., Matsumoto, K., Inoue, N.: Personalization of user profiles for content-based music retrieval based on relevance feedback. In: ACM Int. Conf. on Multimedia (MULTIMEDIA'03). (2003) 110–119

11. Grimaldi, M., Cunningham, P.: Experimenting with music taste prediction by user profiling. In: ACM SIGMM Int. Workshop on Multimedia Information Retrieval (MIR'04). (2004) 173–180
12. Moh, Y., Orbanz, P., Buhmann, J.M.: Music preference learning with partial information. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. (2008) 2021–2024
13. Moh, Y., Buhmann, J.M.: Kernel expansion for online preference tracking. Proceedings of The Int. Society for Music Information Retrieval (ISMIR) (2008) 167–172
14. Su, J.H., Yeh, H.H., Tseng, V.S.: A novel music recommender by discovering preferable perceptual-patterns from music pieces. In: ACM Symp. on Applied Computing (SAC'10). (2010) 1924–1928
15. Logan, B.: Music recommendation from song sets. In: Int. Conf. on Music Information Retrieval (ISMIR'04). (2004) 425–428
16. Yoshii, K., Goto, M., Komatani, K., Ogata, T., Okuno, H.G.: Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences. In: Int. Conf. on Music Information Retrieval (ISMIR'06). (2006)
17. Li, Q., Myaeng, S.H., Kim, B.M.: A probabilistic music recommender considering user opinions and audio features. *Information Processing & Management* **43** (2007) 473–487
18. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. *Journal of the American society for information science* **41** (1990) 391–407
19. Bogdanov, D., Herrera, P.: How much metadata do we need in music recommendation? a subjective evaluation using preference sets. In: Int. Society for Music Information Retrieval Conf. (ISMIR'11). (2011) 97–102
20. Levy, M., Sandler, M.: Learning latent semantic models for music from social tags. *Journal of New Music Research* **37** (2008) 137–150
21. Sordo, M., Celma, O., Blech, M., Guaus, E.: The quest for musical genres: Do the experts and the wisdom of crowds agree? In: Int. Conf. of Music Information Retrieval (ISMIR'08). (2008) 255–260
22. Gibbons, J.D., Chakraborti, S.: *Nonparametric Statistical Inference*. CRC Press (2003)
23. Bogdanov, D., Serrà, J., Wack, N., Herrera, P.: From low-level to high-level: Comparative study of music similarity measures. In: IEEE Int. Symp. on Multimedia (ISM'09). Int. Workshop on Advances in Music Information Research (AdMIR'09). (2009) 453–458
24. Bogdanov, D., Serrà, J., Wack, N., Herrera, P., Serra, X.: Unifying low-level and high-level music similarity measures. *IEEE Trans. on Multimedia* **13** (2011) 687–701
25. Kim, J., Jung, K., Ryu, J., Kang, U., Lee, J.: Design of ubiquitous music recommendation system using MHMM. Volume 2. (2008) 369–374
26. Lu, C., Tseng, V.S.: A novel method for personalized music recommendation. *Expert Systems with Applications* **36** (2009) 10035–10044