Content-based music recommendation based on user preference examples

Dmitry Bogdanov, Martín Haro, Ferdinand Fuhrmann, Emilia Gómez, Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra
Roc Boronat, 138, 08018 Barcelona, Spain
{firstname.lastname}@upf.edu

ABSTRACT

Recommending relevant and novel music to a user is one of the central applied problems in music information research. In the present work we propose three content-based approaches to this task. Starting from an explicit set of music tracks provided by the user as evidence of his/her music preferences, we infer high-level semantic descriptors, covering different musical facets, such as genre, culture, moods, instruments, rhythm, and tempo. On this basis, two of the proposed approaches employ a semantic music similarity measure to generate recommendations. The third approach creates a probabilistic model of the user's preference in the semantic domain. We evaluate these approaches against two recommenders using state-of-the-art timbral features, and two contextual baselines, one exploiting simple genre categories, the other using similarity information obtained from collaborative filtering. We conduct a listening experiment to assess familiarity, liking and further listening intentions for the provided recommendations. According to the obtained results, we found our semantic approaches to outperform the low-level timbral baselines together with the genre-based recommender. Though the proposed approaches could not reach a performance comparable to the involved collaborative filtering system, they yielded acceptable results in terms of successful novel recommendations. We conclude that the proposed semantic approaches are suitable for music discovery especially in the long tail.

Categories and Subject Descriptors

H.3.3 [**Information Storage And Retrieval**]: Information Search and Retrieval—*information filtering, selection process, retrieval models*; H.5.5 [**Information Interfaces And Presentation**]: Sound and Music Computing—*modeling, systems*

General Terms

Algorithms, Measurement, Human Factors

WOMRAD 2010 Workshop on Music Recommendation and Discovery, colocated with ACM RecSys 2010 (Barcelona, SPAIN)

Keywords

recommender systems, user modeling, evaluation, music recommendation, content-based, collaborative filtering

1. INTRODUCTION

Rapid growth of digital technologies, the Internet, and the multimedia industry has provoked a huge information overload and a necessity of effective information filtering systems, and in particular recommendation systems. In the case of the digital music industry, current major Internet stores contain millions of tracks, which complicates search, retrieval, and discovery of music relevant for a user. At present, the majority of industrial systems provide means for contextual manual search based on information about artist names, album or track titles, and additional semantic properties, which are mostly limited to genres. Using this information music collections are becoming browsable by textual queries and tags.

Besides, current research within the music information retrieval (MIR) community achieved relative success in the task of measuring music similarity [7], striving for facilitation of manual search, and automatization of music recommendation. To this extent, music tracks can be represented in a certain feature space filled in with contextual information, extracted from available metadata, user ratings [18], and social tags [12] (i.e. the contextual approach), or with information, extracted from audio content itself [4, 6, 16, 17, 21] (i.e. the content based approach). Thus, it becomes possible to define many similarity measures (or distances[1]) between tracks in a music collections, and therefore to browse collections and to recommend music using queries-by-example. Still the majority of the content-based distances employ solely rough timbral information, such as Mel frequency cepstral coefficients (MFCCs), and sometimes temporal information. Additionally, current systems provide basic means for personalization, obtaining a user's profile in form of consuming statistics, music ratings, and other types of behavioral information, and operating with this information generally in a collaborative filtering manner [2, 8, 9]. While more sophisticated personalization approaches which explore the nature of preference behavior using both contextual information and audio content information are necessary, they are still in their infancy [13–15, 19, 22] and require more research attention.

Generally, we can discern two types of user interaction

[1] We will pragmatically use the term distance to refer to any dissimilarity measurement between tracks.

with a music retrieval system: (i) music search, when a user has an initial idea of what he/she wants, and operates with metadata to query for a specific artist, album, genre, etc., or provides a query-by-example in the case of similarity-based retrieval, and (ii) music discovery, when a user does not know his/her exact needs and prefers to browse an available music collection on purpose to discover music which is relevant in respect to his/her musical preferences. Querying by example requires a user to explicitly define the "direction of search", and is not perfectly suited for discovery. On the other hand, querying by broad semantic categories (such as genres) can provide an excessive amount of potentially relevant data, containing thousands of tracks. While for both types of interaction contextual information can be used, it is found that contextual approaches perform well on popular items, but fail in the long tail due to the lack of available user ratings, social tags, and metadata for unpopular items [8]. Instead, content-based information extracted from audio can help to overcome this problem.

We focus the present work on content-based music recommendation, concerning both relevance and novelty (i.e. discovery) aspects. We do not consider the issue of balancing both aspects according to a user's current needs. Instead, we present a way to infer user preferences from audio content, and a number of recommendation approaches, which are challenged to provide both relevant and novel recommendations to a user. We propose a procedure to generate such recommendations based on an explicit set of music tracks defined by a given user as evidence of his/her musical preferences. Up to our knowledge this recommendation approach has never been evaluated before. We ask the user to provide such a preference set (Sec. 2.1) in order to extract low-level audio features as well as infer high-level semantic information from the audio of each of the tracks (Sec. 2.2). We then consider three different approaches operating on a semantic domain to summarize the retrieved descriptions and generate music recommendations. Two of them have a music similarity measure in their core (Secs. 2.3.1, and 2.3.2), while the third approach applies a probabilistic model to infer the underlying structure of the user's preferences (Sec. 2.3.3). Alternatively, in order to evaluate the generated recommendations, we employ two approaches, which apply the same ideas on low-level timbral features (Secs. 2.3.4, and 2.3.5), and two contextual ones including a state-of-the-art collaborative filtering recommendation system (Sec. 2.3.6), and a naive genre-based recommender baseline (Sec. 2.3.7). We evaluate all considered approaches by gathering music data from 12 participants (Sec. 3.1), and carrying out a listening experiment to assess familiarity, liking and further listening intentions of the provided recommendations (Sec. 3.2), and present the obtained results (Sec. 3.3). Finally, we draw conclusions about the proposed procedure and discuss future research directions (Sec. 4).

2. METHODOLOGY

2.1 Preference Examples Collection

As a first step, we ask the user to gather the minimal set of music tracks sufficient to grasp or convey her/his music preferences [10] (the user's preference set). We do not promise or mention giving music recommendations in the future, which could bias the selection of representative music. The user provides a folder with the selected tracks in audio format (e.g. mp3), and all the needed information to unambiguously identify and retrieve each track (i.e. artist, piece title, edition, etc.). For the content-based approaches which we will consider, single music pieces are informative by themselves without any additional context, such as artist names and track titles. Still we ask the user to provide this context to be able to make comparison with contextual approaches. We also ask the user for additional information, including personal data (gender, age, interest for music, musical background), a description of the strategy followed to select the music pieces, and the way he/she would describe his/her musical preferences. This information will help us for further analysis.

2.2 Audio Content Analysis

We now describe the procedure of obtaining meaningful low-level and high-level descriptions of each music track from the user's preference set within the used audio content analysis system. We follow [6] to obtain such descriptions. To this extent, for each track we calculate a low-level feature representation using an in-house audio analysis tool[2]. In total it provides over 60 commonly used low-level audio features, characterizing global properties of the given tracks, including timbral, temporal, and tonal features among others. They include inharmonicity, odd-to-even harmonic energy ratio, tristimuli, spectral centroid, spread, skewness, kurtosis, decrease, flatness, crest, and roll-off factors, MFCCs, spectral energy bands, zero-crossing rate, spectral and tonal complexities, transposed and untransposed harmonic pitch class profiles, key strength, tuning, chords, beats per minute and onsets.

We do not use the described low-level features explicitly in the approaches we will consider, except for MFCCs, used to construct two of the baseline systems. Instead, we use them to infer semantic descriptors. For that reason, we perform a regression by suitably trained classifiers producing different semantic dimensions such as genre, culture, moods, and instrumentation. We use standard multi-class support vector machines (SVMs) [20], employ 14 ground truth music collections (including full tracks and excerpts) and execute 14 classification tasks corresponding to these data. The regression results form a high-level descriptor space, which contains the probability estimates for each class of each SVM classifier. With the described procedure we obtain 56 high-level descriptors, including categories of genre, culture, moods, instruments, rhythm and tempo. For more detailed information regarding the list of low-level features, the collections used for regression, and SVM implementation see [6] and references therein.

2.3 Recommendation Approaches

We now consider different approaches to music recommendation, which are based on the retrieved descriptions of the user's preference set. The approaches we propose include three methods working on semantic descriptors. In comparison, we consider two low-level baseline approaches working on MFCCs, and two contextual ones.

All approaches are used to retrieve 20 music tracks from a given music collection as the recommendations for the user except one of the contextual approaches (Sec. 2.3.6), which operates on *Last.fm*[3] music collection.

[2] `http://mtg.upf.edu/technologies/essentia`

[3] `http://last.fm`

2.3.1 Semantic distance from the mean (SEM-MEAN)

As the simplest approach, we propose the representation of the user as a single point in the semantic descriptor space. As such, we compute the mean point for the user's preference set. We employ the semantic distance, presented and validated in [6]. It has been shown to perform with positive user satisfaction, being comparable to well-known low-level timbral distances, based on MFCCs, while operating in a high-level semantic space. More concretely, the distance operates directly on the retrieved semantic descriptors, and is defined as a weighted Pearson correlation distance [1, 6]. Given a music collection, we rank the tracks according to the semantic distance to the user point (i.e. the mean point of the user's preference set) and return 20 nearest tracks as recommendations.

2.3.2 Semantic distance from all tracks (SEM-ALL)

Alternatively, we do not simplify the user representation to one point but instead consider all tracks from the user's preference set. We use the same semantic distance as for SEM-MEAN. For each track from the user's preference set, we compute the distances to the tracks in a given music collection, and mark 20 nearest tracks as candidates. We then rank all selected candidates according to the obtained distances, omit possible duplicates, and return the tracks corresponding to the lowest 20 distances as recommendations. In this case, we take into account all possible areas of preferences, explicitly specified by the user, while searching for the most similar tracks.

2.3.3 Semantic Gaussian mixture model (SEM-GMM)

Finally, we propose the representation of the user as a probability density of his/her preferences on the semantic space. For that purpose, we use the retrieved semantic descriptors, and employ a Gaussian mixture model (GMM) [5], which estimates a probability density as a weighted sum of a given number of simple Gaussian densities (components). We initialize the GMM by k-mean clustering, and train the model using the expectation-maximization algorithm. The number of centers for the k-means are estimated by Bayesian information criterion [5]. For computational reasons, we consider a number of components in the range between 1 and 20. Once we have our model trained, we compute probability density for each of the tracks in a given music collection. We rank the tracks according to the obtained density values, and return 20 most probable tracks as recommendations under the assumption of a uniform distribution of the tracks in the universe within the semantic space.

The advantage of SEM-GMM approach is that the model takes the relevance of the semantic attributes within the user's preferences into account, accenting areas preferred by the user in the semantic space. Thus, the recommended tracks would generally comprise of the most characteristic semantic properties, inferred from the user's preference set. Meanwhile, SEM-ALL is blind to the underlying semantic structure of preferences, and SEM-MEAN only provides very rough approximation. Still, in the case when the user's tracks are evenly distributed in the semantic space, SEM-GMM may have insufficient expressive power due to the assigned limit of Gaussian components, discriminating certain preference areas. Nonetheless we assume gaussianity of the user's preference patterns.

2.3.4 Timbral distance from all tracks (MFCC-ALL)

For comparison purposes and as our first baseline we modify the SEM-ALL approach to use a common low-level timbral distance [16] instead of the semantic one. To this extent, we use MFCCs and model each music track as a single Gaussian with full covariance matrix. A closed form symmetric approximation of the Kullback-Leibler divergence is then used as a distance. Thereby, we can regard the MFCC-ALL approach as a counterpart of the distance-based approach to music recommendation proposed by Logan [14] in which the Earth-Mover's Distance between MFCC clusters is used as a distance measure.

2.3.5 Timbral Gaussian mixture model (MFCC-GMM)

Alternatively, as in the SEM-GMM approach, we construct a probabilistic model using a GMM. Instead of the semantic descriptors, we use a population of mean MFCC vectors (one vector per track) to train the model.

2.3.6 Collaborative filtering with Last.fm (LASTFM)

In addition to the described content-based approaches, we consider a contextual baseline approach based on music similarity inferred from collaborative filtering information. We did not have at hand any data of this kind on our own, and therefore we opted for the usage of black box recommendations, provided by *Last.fm*. It is an established music recommender with an extensive number of users, and a large music collection, providing means for both monitoring listening statistics and social tagging [11].

We manually generate a list of recommendations browsing *Last.fm*. The procedure we follow for that purpose partially emulates human user behavior while discovering new music. During the retrieval procedure we did not open any account for *Last.fm*, therefore we consider such recommendations unbiased to possible personalization, which can be provided for the registered accounts. We randomly preselect 20 music tracks from the user's preference set, and query the *Last.fm* website for each of the preselected tracks. To this extent, for each query track, we search a corresponding *Last.fm* track page[4]. If the track page is found, we pass to the "Similar Music" page[5]. This page provides a ranked list of tracks similar to the query track. From the list we select the first track which is available for pre-listen online, by a different artist than the query track. Otherwise, if the corresponding track page is not found, or the "Similar Music" page is not available for the query track due to insufficient collaborative filtering data (e.g., when the query track is an unpopular long-tail track with low number of listeners), we search for the corresponding artist page[6] and proceed to the "Similar Artists" page[7]. This page provides a ranked list of artists, similar to the artist of the query track. We apply an artist filter to the list as the query artist name can have variations. Thereafter we select the top-ranked artist from the list, go to the corresponding artist page, and select the first track, which is available for pre-listen online, from the "Top Tracks" section. This section provides two lists of the most popular tracks by the artist, relying on short-term last

[4] for example, see `http://www.last.fm/music/Mastodon/_/The+Czar`

[5] `http://www.last.fm/music/Mastodon/_/The+Czar/+similar`

[6] `http://www.last.fm/music/Baby+Ford/`

[7] `http://www.last.fm/music/Baby+Ford/+similar`

week period, or long-term last 6 months period of listening statistics. We opted for the last 6 months period. If no prelistens are found, we proceed iteratively to the next similar artist's top tracks, until we find one. If no similar artist contains previewable tracks, we skip the query track.

2.3.7 Random tracks by the same genre (GENRE)

Finally, as a simple and low-cost contextual baseline, we provide random recommendations, which rely on genre categories of the user's tracks. As in the LASTFM approach, we preselect 20 music tracks from the user's preference set. For each of the tracks we obtain a genre category of this track from the *Last.fm* track page, or artist page. As such, we select the first genre tag we encounter, which is presented in a given music collection (we assume, that all tracks are tagged with a genre category). Thereafter, we return a random track of this genre tag from the collection.

3. EXPERIMENTS AND RESULTS

3.1 User Data Analysis

We worked with a group of 12 users (8 males and 4 females). They were aged between 25 and 45 years old (average $\mu = 32.75$ years old and standard deviation $\sigma = 5.17$ years old) and showed a very high interest in music (rating around $\mu = 9.58$, with $\sigma = 0.67$, where 0 means no interest in music and 10 means passionate about music). Ten of the twelve users play at least one musical instrument, including violin, piano, guitar, singing, synthesizers and ukulele.

The number of tracks selected by the users to convey their musical preferences was very varied, ranging from 19 to 178 music pieces ($\mu = 73.25$, $\sigma = 46.07$). The time spent for this task also differed a lot, ranging from half an hour to 180 hours ($\mu = 30.41$, $\sigma = 54.19$).

It is interesting to analyze the provided verbal descriptions about the strategy followed to select the music tracks. Some of the users were selecting one song per artist, while some others did not apply this restriction. They also covered various uses of music such as listening, playing, singing or dancing. Other users mentioned musical genre, mood, expressivity, musical parameters, lyrics and chronological order as driving parameters for selecting the tracks. Furthermore, some users implemented an iterative strategy by gathering a very large amount of music pieces from their music collection and performing a further refinement to obtain the final selection.

Finally, each user provided a set of labels to define their musical preferences. Most of them were related to genre, mood and instrumentation, some of them to rhythm and few to melody, harmony or expressivity. Other labels were attached to lyrics, year and duration of the piece. The users' preferences covered a wide range of musical styles (from classical to country, jazz, rock, pop, electronic, folk) and musical properties (e.g. acoustic vs. synthetic, calm vs. danceable, tonal and dissonant).

3.2 Recommendation Evaluation

In order to evaluate the considered approaches, we performed subjective listening tests on our 12 subjects. The entire process used an in-house collection of 100K music excerpts (30 sec.) by 47K artists (approximately 2 tracks per artist) covering a wide range of musical dimensions (different genres, styles, arrangements, geographic locations, and epochs). For each user we generated 7 recommendation playlists, using each of the three proposed approaches and two low-level plus two contextual baseline approaches. Each playlist consisted of 20 music tracks, returned by the respective approach specifics (Sec. 2.3). No playlist contained more than one song from the same artist. All playlists were merged into a single list of 140 tracks, with all the tracks randomly ordered to avoid any response bias because of presentation order or because of recommendation approach. The file names were anonymized, and all metadata was deleted from the files as well, to make contextual identification of the tracks impossible. Also the participants were not aware of the amount of recommendation approaches, their names and their rationales.

Table 1: The percent of fail, trust, hit, and unclear categories per recommendation approach. Note that the results for the LASTFM approach were obtained on a different underlying music collection.

Approach	fail	hit	trust	unclear
SEM-MEAN	49.167	31.250	2.500	17.083
SEM-ALL	42.500	34.583	3.333	19.583
SEM-GMM	48.750	30.000	2.500	18.750
MFCC-ALL	64.167	15.000	2.083	18.750
MFCC-GMM	69.583	11.667	1.250	17.500
LASTFM	16.667	41.250	25.417	16.667
GENRE	53.750	25.000	1.250	20.000

A questionnaire was given for the subjects to express different subjective impressions related to the recommended music. 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 "gimmemore" allowing just 1 or 0 to respectively indicate a request for, or a reject of, more music like the one presented. The users were also asked to provide title and artist for those tracks rated high in the familiarity scale. We manually corrected this scale when the given artist/title was wrong (hence a familiarity rating of "3" or, more frequently, "4", was sometimes lowered to 1. These corrections represented just 3% of the total familiarity judgments.

3.3 Results

Considering the subjective scales used, a good recommendation system should provide high-liking/listening intentions/request for the greater part of retrieved tracks and in particular for low-familiarity tracks. Therefore, we recoded the user's ratings into 3 main categories, referring to the type of the recommendation: hits, fails and trusts. Hits were those tracks having a low familiarity rating (< 2) and a high (> 2) liking rate. Fails were those tracks having a low (< 3) liking rating. Trusts were those tracks that got a high familiarity (> 1) and a high (> 2) liking rate. Trusts, provided their overall amount is low, can be useful for a user to feel that the recommender is understanding his/her preferences [3] (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). Using the liking, the intentions and the "gimmemore" Boolean rating we respectively computed three different recommendation outcome measures. Then we combined the three into a final recommendation outcome that required absolute coincidence of them in order to consider it to be a hit, a fail or a trust. A 18.3% of all the recommendations were then considered as "unclear" (e.g., a case that, using the liking, it was a hit, but using the other two indexes it was a fail), and were excluded from further analyzes. An interesting additional result is that many of the unclear outcomes correspond to high-liking ratings that turned into 0 in the gimmemore scale. This pattern was more frequent for the recommendations generated using the GMM-MFCC (6.6%) than for any other approaches, being the GENRE the least changed (2.9%). Contrastingly, the opposite change (low-liking becoming positive "gimmemore") was nearly absent in the ratings.

The percent of each category per recommendation approach is presented in Table 1. An inspection of it reveals that the approach yielding more hits (41.2%) and trusts (25.4%) is LASTFM (not surprisingly the trusts found with other approaches were scarce, below 4%). The three approaches based on semantic descriptors (SEM-ALL, SEM-MEAN and SEM-GMM) yielded more than 30% of hits, and the remaining ones could not supply more than 25%. The existence of an association between recommendation approach and the outcome of the recommendation could be accepted, according to the result of the Pearson chi-square test ($\chi^2(18) = 351.7$, $p < 0.001$).

Additionally, three separate between-subjects ANOVA were performed in order to test the effects of the recommendation approaches on the three subjective ratings. The effect was confirmed in all of them ($F(6, 1365) = 55.385$, $p < 0.001$ for the liking rating, $F(6, 1365) = 48.89$, $p < 0.001$ for the intentions rating, and $F(6, 1365) = 43.501$, $p < 0.001$ for the "gimmemore" rating). Pairwise comparisons using Tukey's test revealed the same pattern of differences between the recommendation approaches, irrespective of the 3 tested indexes. This pattern highlights the LASTFM approach as the one getting the highest overall ratings, it also groups together the MFCC-GMM and MFCC-ALL approaches (those getting the lowest ratings), and the remaining approaches also clustered in-between.

Finally, a measure of the quality of the hits was computed doing $(\text{liking} - \text{familiarity}) * \text{intentions}$. Selecting only the hits, an ANOVA on the effect of recommendation method on this quality measure revealed no significant differences attributable to the method. Therefore, once a hit is selected, there is no recommendation method granting better or worst recommendations than any other. The same pattern was revealed by solely using the liking as a measure of the quality of the hits.

4. CONCLUSIONS

In this work we presented three content-based approaches to music recommendation, which are based on an explicit set of music tracks provided by a user as evidence of his/her musical preferences (the user's preference set). Our approaches work on semantic descriptors (inferred from low-level audio features in diverse classification tasks) covering musical dimensions such as genre and culture, moods and instruments, and rhythm and tempo. More concretely, we proposed two approaches which apply a high-level semantic distance to retrieve tracks from a given collection. These approaches compute the distance either from the mean point of the preference set, or from all tracks in the preference set. Alternatively, we proposed a model-based approach, which creates a probabilistic model to infer the underlying structure of the user's preferences. For that purpose, we employed a GMM to model the preferences within the semantic domain. We evaluated the proposed approaches against a number of baselines in a subjective evaluation with 12 users. As such baselines, we considered two approaches operating on low-level timbral features (MFCCs) instead of the proposed semantic descriptors. The first approach employs a state-of-the-art timbral distance, while the second one creates a GMM within the timbral domain. Moreover, in contrast to the content-based methods, we included two contextual recommenders in our evaluation. One of them naively retrieves random tracks from a given music collection by a genre criterion. The other employs *Last.fm* as a source for collaborative filtering information about music similarity.

The evaluation results revealed the user's preference of the proposed semantic approaches over the low-level timbral baselines. This concerns both the compared distance-based approaches as well as the probabilistic models. Regarding the semantic distance employed in our approaches, this fact supports and complements the outcomes from the previous research on semantic music similarity measures [6], in which a number of similarity measures were evaluated in a subjective experiment but on a set of random tracks not necessarily preferred by participants. In that experiment a comparable performance of the semantic and low-level timbral distances was revealed, meanwhile the semantic distance surpassed the other methods in objective evaluations. Considering these previous results and the present outcomes, we may conclude that the high-level semantic description outperforms the low-level timbral description in the task of music recommendation.

In contrast, the proposed approaches are found to be inferior to the considered collaborative filtering recommender in terms of both the number of successful novel recommendations (hits) and the trusted recommendations. This result can be partly explained by the fact that the recommendations generated by the latter approach used the *Last.fm* music collection, which could entail an evaluation bias. Considering this fact, we can hypothesize a lower performance of the collaborative filtering approach on our in-house collection. Still the collaborative filtering approach yielded only 7% more hits than our best proposed semantic method. In particular, we expect the proposed approaches to be suitable for music discovery in the long tail which has a lack of contextual information, and incorrect or incomplete metadata.

Interestingly, the naive genre-based recommender, while being worse than our proposed approaches, still outperformed the timbre-based baselines. This could be partially explained by the fact that genre was one of the driving criteria for selecting users' preference sets, and that genre entails more information and diversity than timbral information extracted from MFCCs. We also did not find benefits of using our semantic GMM-based approach comparing to the semantic distance-based approaches, probably due to the insufficient size of training data (only one mean MFCC vector per track was computed in our experiments).

In general, we conclude that though the considered content-based approaches to music recommendation do not reach the satisfaction and novelty degree of the collaborative filtering approach, the difference in performance diminishes to a great extent while using semantic descriptors. We may hypothesize a better performance, comparable with the collaborative filtering approach, once the amount and quality of semantic descriptors is increased. Consequently, future research will be devoted to the extension of the inherent semantic descriptor space, used by the proposed approaches, as well as the improvement of the underlying classifiers, and the distance measure. Furthermore, we plan to assess the potential benefit of user profiling by explicitly given preference examples in form of music tracks over more broad contextual categories (favorite artists, albums, genres, and even activities), and implicit information such as listening behavior statistics.

5. ACKNOWLEDGMENTS

The authors would like to thank all participants involved in the evaluation. This research has been partially funded by the FI Grant of Generalitat de Catalunya (AGAUR).

6. REFERENCES

[1] M. B. Abdullah. On a robust correlation coefficient. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 39(4):455–460, 1990.

[2] L. Baltrunas and X. Amatriain. Towards time-dependant recommendation based on implicit feedback. In *Workshop on Context-aware Recommender Systems (CARS'09)*, 2009.

[3] L. Barrington, R. Oda, and G. Lanckriet. Smarter then genius? human evaluation of music recommender systems. In *10th International Society for Music Information Retrieval Conference (ISMIR'09)*, 2009.

[4] L. Barrington, D. Turnbull, D. Torres, and G. Lanckriet. Semantic similarity for music retrieval. In *International Symposium on Music Information Retrieval (ISMIR'07)*, 2007.

[5] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[6] D. Bogdanov, J. Serrà, N. Wack, and P. Herrera. From low-level to high-level: Comparative study of music similarity measures. In *International Workshop on Advances in Music Information Research (AdMIRe'09)*, 2009.

[7] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4):668–696, 2008.

[8] O. Celma. *Music recommendation and discovery in the long tail*. PhD thesis, UPF, Barcelona, Spain, 2008.

[9] C. S. Firan, W. Nejdl, and R. Paiu. The benefit of using tag-based profiles. In *Latin American Web Conference*, pages 32–41, 2007.

[10] M. Haro, A. Xambó, F. Fuhrmann, D. Bogdanov, E. Gómez, and P. Herrera. The musical avatar - a visualization of musical preferences by means of audio content description. In *Audio Mostly (AM '10)*, Pitea, Sweden, 2010. ACM.

[11] N. Jones and P. Pu. User technology adoption issues in recommender systems. In *Networking and Electronic Commerce Research Conference*, 2007.

[12] M. Levy and M. Sandler. Music information retrieval using social tags and audio. *IEEE Transactions on Multimedia*, 11(3):383–395, 2009.

[13] Q. Li, S. H. Myaeng, and B. M. Kim. A probabilistic music recommender considering user opinions and audio features. *Information Processing & Management*, 43(2):473–487, Mar. 2007.

[14] B. Logan. Music recommendation from song sets. In *Proc ISMIR*, page 425–428, 2004.

[15] C. C. Lu and V. S. Tseng. A novel method for personalized music recommendation. *Expert Systems with Applications*, 36(6):10035–10044, 2009.

[16] E. Pampalk. *Computational models of music similarity and their application in music information retrieval*. PhD thesis, Vienna University of Technology, Mar. 2006.

[17] T. Pohle, D. Schnitzer, M. Schedl, P. Knees, and G. Widmer. On rhythm and general music similarity. In *10th International Society for Music Information Retrieval Conference (ISMIR'09)*, 2009.

[18] M. Slaney and W. White. Similarity based on rating data. In *International Symposium on Music Information Retrieval (ISMIR'07)*, 2007.

[19] J. Su, H. Yeh, P. S. Yu, and V. S. Tseng. Music recommendation using content and context information mining. *IEEE Intelligent Systems*, 25(1):16–26, 2010.

[20] V. Vapnik. *The Nature of Statistical Learning Theory (Information Science and Statistics)*. Springer, 2nd edition, Nov. 1999. Published: Hardcover.

[21] K. West and P. Lamere. A model-based approach to constructing music similarity functions. *EURASIP Journal on Advances in Signal Processing*, 2007:149–149, 2007.

[22] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno. Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences. In *International Conference on Music Information Retrieval (ISMIR'06)*, 2006.