A Content-based System for Music Recommendation and Visualization of User Preferences Working on Semantic Notions

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

Abstract

The amount of digital music has grown unprecedentedly during the last years and requires the development of effective methods for search and retrieval. In particular, content-based preference elicitation for music recommendation is a challenging problem that is effectively addressed in this paper. We present a system which automatically generates recommendations and visualizes a user's musical preferences, given her/his accounts on popular online music services. Using these services, the system retrieves a set of tracks preferred by a user, and further computes a semantic description of musical preferences based on raw audio information. For the audio analysis we used the capabilities of the Canoris API. Thereafter, the system generates music recommendations, using a semantic music similarity measure, and a user's preference visualization, mapping semantic descriptors to visual elements.

1. Introduction

Rapid growth of the Internet and the multimedia industry during the last decade has become a reason for a huge information overload. In particular, in the case of digital music, the amount of tracks available in the largest Internet stores, such as *iTunes Store*, *Amazon MP3*, and *7digital*, already exceeds 10 millions[1]. This requires effective means for browsing and search in music collections and, more importantly, tools for automatized user preference elicitation and music recommendation. While contextual data including user ratings and tags is a powerful source of information to provide such means, it might not be effective on less popular items. In general, these items, being situated in the so-called long-tail [5], suffer from the lack of contextual data generated by users. Here, content-based information extracted from audio can help to overcome this problem [6, 1].

In this work, we present a system for music recommendation and visualization of user preferences. At its core, our system exploits solely content-based information extracted from the audio signal. More concretely, it considers high-level semantic descriptions of music tracks, inferred from low-level timbral, temporal, and tonal features. We follow the ideas presented in [7, 2], where a set of audio tracks is explicitly given by a user as evidence of her/his musical preferences (i.e. the *preference set*) to create a user model. This model has been shown to be suitable for music recommendation and visualization. In contrast to explicit gathering, the present system obtains such a preference set in an implicit way, given a user profile on popular online music services, such as *Last.fm*[2] and *SoundCloud*[3]. Semantic descriptors including genres, musical culture, moods, instrumentation types, rhythm, and tempo information are extracted from the preference set of a user and then exploited to recommend similar music and visualize musical preferences.

2. The system

Here we describe the implementation of the present system. Structurally, its workflow can be divided into data gathering, audio analysis, music recommendation, and preference visualization. The block diagram of the entire system is shown in Fig. 1. The user interface is designed in the form of a web page[4].

[1] `http://en.wikipedia.org/wiki/Comparison_of_online_music_stores`, retrieved on January 27th, 2011.

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

[3] `http://soundcloud.com`

[4] A demo of the system is available online: `http://mtg.upf.edu/project/musicalavatar`

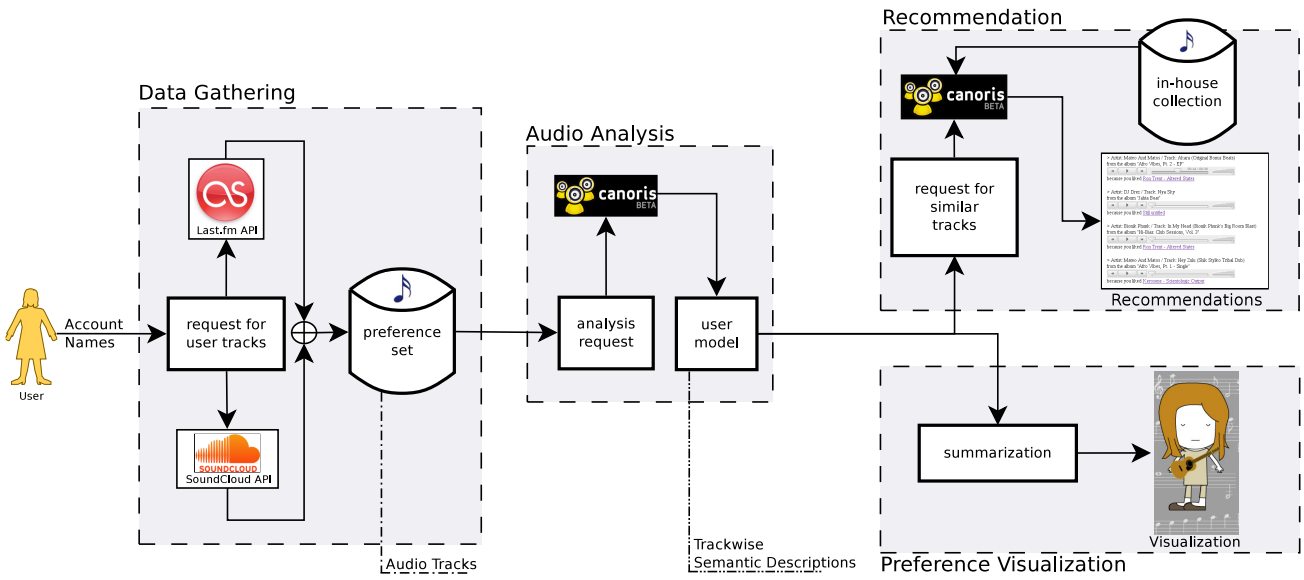


Figure 1. Block diagram of the entire system.

2.1. Data gathering

The user specifies her/his account name on *Last.fm* and/or *SoundCloud* services, from which the preferred tracks should be retrieved. *Last.fm* is an established music recommender with an extensive number of users and a large playable music collection. It provides means for both monitoring listening statistics and social tagging [8]. In particular, users can explicitly mark specific tracks as their favorites or access their track listening statistics. This information is available via the *Last.fm* API[5]. *SoundCloud* is a platform which allows users (mostly musicians) to collaborate, promote, and distribute their music. Specifically, it allows users to upload their own tracks or mark tracks as their favorites. This information is available via the *SoundCloud* API[6]. Our system is currently limited to the users of these musical services, but will be further extended with an option to upload preferred tracks to our server.

Different types of tracks can be used to infer the user's preference set:

- Tracks marked as favorites by the user on *Last.fm*.
- Tracks listened most by the user according to their *Last.fm*'s statistics.
- Tracks marked as favorites by the user on *SoundCloud*.
- Tracks uploaded by the user on *SoundCloud*.

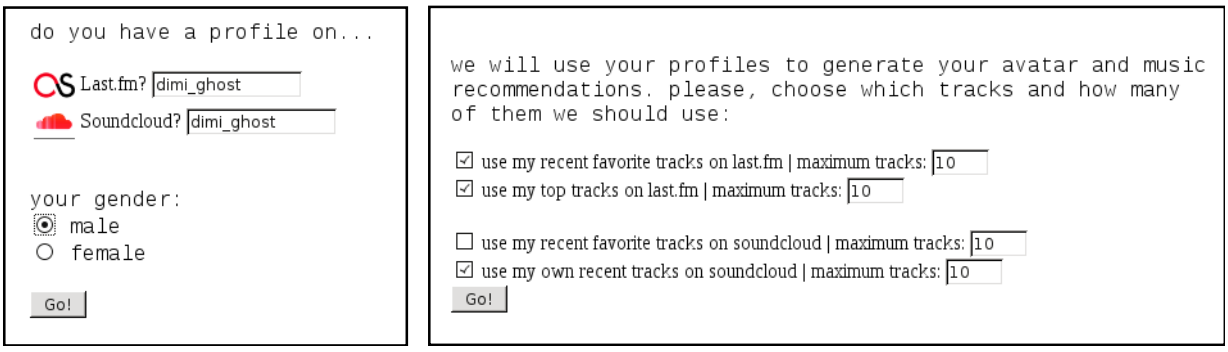
[5] http://www.last.fm/api
[6] http://soundcloud.com/developers

The type of tracks to use and their amount can be specified by the user. Fig. 2 presents two screenshots of the user interface. The system retrieves the URLs of the tracks to be included in the preference set using the *Last.fm* and *SoundCloud* APIs. Using these URLs, audio fragments (30 sec.) of the track previews are downloaded[7].

2.2. Audio analysis

We follow the research work presented in [7, 2] to infer the semantic descriptions for each track from the user's preference set. This procedure implies extraction of low-level timbral, temporal, and tonal audio features and thereafter running a number of classification tasks, using support vector machines trained on ground truth information about genres, musical culture, moods, instrumentation, rhythm, and tempo. Technically, we use the *Canoris*[8] API to obtain such semantic descriptions. *Canoris* is a recently launched web service developed by the UPF's Music Technology Group[9] for the analysis and synthesis of sound and music. In particular, *Canoris* provides audio analysis and content-based similarity search capabilities. For each uploaded audio file a wide set of analysis features can be retrieved, ranging from frame-based low-level features (such as MFCCs) up to semantic descriptions (such as mood or genre labels) of the file as a whole. The latter include the aforementioned classification procedure when computing a number of high-level descriptors. To this extent, our system obtains track-

[7] The system considers solely the tracks with available previews.
[8] http://canoris.com/
[9] http://mtg.upf.edu



(a) (b)

Figure 2. Screenshots of the user interface when specifying (a) the account names, and (b) the type of tracks to be retrieved.

wise semantic descriptions for the user's preference set (i.e. the user model).

2.3. Music recommendation

To generate recommendations, we employ an in-house music collection of 50.000 music excerpts, covering a wide range of musical genres. This collection was analyzed via the *Canoris* API to retrieve the same semantic descriptions as used for the preference set. We follow [2] to search for the tracks from our in-house music collections which are similar to the user's preference set. We use a semantic distance, proposed in [3, 4] and implemented in the core of the similarity engine provided by the *Canoris* API. This similarity engine supports queries-by-example to retrieve the most similar files based on the analysis results. In the aforecited literature, the proposed distance was shown to outperform common timbral approaches to music similarity and music recommendation in both objective and subjective evaluations.

The system searches for tracks inside the in-house collection with the smallest semantic distance to any of the tracks in the preference set (see Fig. 3). An additional parameter is introduced to limit the number of possible recommendation outcomes per each track from the preference set. Thereafter, the recommendation outcomes are presented to the user, including metadata of the tracks, audio previews, and the reason why a particular track was recommended (i.e. recommendation sources). An example of the recommendation output provided to a user is presented in Fig. 4(a).

2.4. Preference visualization

We follow [7] to visualize a user's musical preferences in the form of a humanoid cartoon character, the *Musical Avatar*. Selected descriptors of each track are summarized across all tracks in the preference set. The resulting descriptors represent degrees of the user preference for different genres, moods, and instrumentation types. These descriptors are then mapped to visual elements of the avatar, which are implemented using Processing[10] [9]. We refer the interested reader to [7] for further details. The generated visualization is then provided to the user (see Fig. 4(b)).

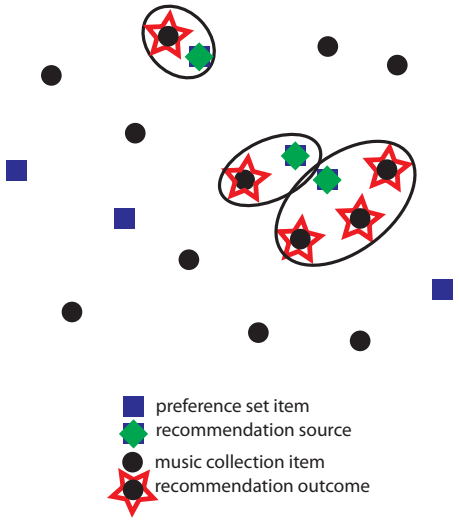
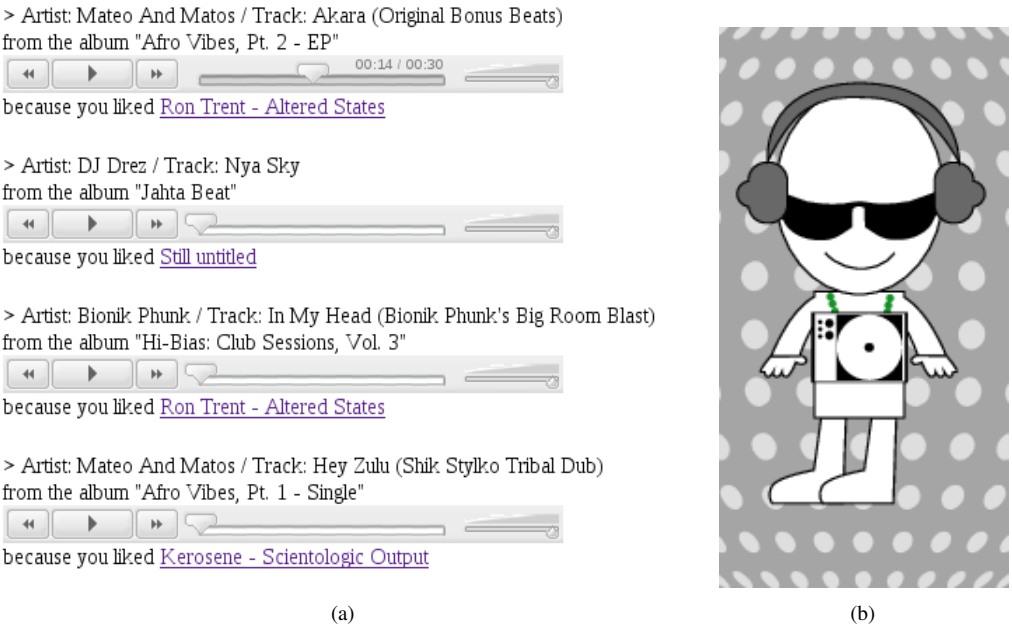


Figure 3. Graphical representation of the recommendation approach in the semantic space. The music collection items with the smallest distances to the preference set items are selected as recommendation outcomes.

[10] http://processing.org



(a) (b)

Figure 4. Screenshots of (a) the recommendation output and (b) the generated preference visualization returned to a user.

3. Conclusions

We presented a system suited for music recommendation and user preference visualization. Our system operates on content-based information extracted from audio, and, more concretely, it exploits high-level semantic descriptions of music tracks automatically inferred by a number of classifiers. The system employs the *Last.fm* and *Soundcloud* APIs to generate semantic user models for the members of these music services. To this end, the system extracts audio fragments of the tracks preferred by the users and computes semantic descriptions of these tracks via the *Canoris* API. Finally, the system generates musical recommendations, relying on a semantic similarity measure between music tracks. In addition, the inferred musical preferences of the users are visualized in the form of *Musical Avatars*.

4. Acknowledgments

This research has been partially funded by the FI Grant of Generalitat de Catalunya (AGAUR).

References

[1] L. Barrington, R. Oda, and G. Lanckriet. Smarter than genius? human evaluation of music recommender systems. In *International Society for Music Information Retrieval Conference (ISMIR'09)*, pages 357–362, 2009.

[2] D. Bogdanov, M. Haro, F. Fuhrmann, E. Gómez, and P. Herrera. Content-based music recommendation based on user preference examples. In *ACM Conference on Recommender Systems. Workshop on Music Recommendation and Discovery (Womrad 2010)*, Sept. 2010.

[3] D. Bogdanov, J. Serrà, N. Wack, and P. Herrera. From low-level to high-level: Comparative study of music similarity measures. In *IEEE International Symposium on Multimedia (ISM'09). International Workshop on Advances in Music Information Research (AdMIRe'09)*, pages 453–458, 2009.

[4] D. Bogdanov, J. Serrà, N. Wack, P. Herrera, and X. Serra. Unifying low-level and high-level music similarity measures. *IEEE Transactions on Multimedia*. In press.

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

[6] O. Celma and P. Herrera. A new approach to evaluating novel recommendations. In *ACM Conference on Recommender Systems (RecSys'08)*, pages 179–186, 2008.

[7] 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)*. ACM, 2010.

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

[9] C. Reas and B. Fry. *Processing: a programming handbook for visual designers and artists*. MIT Press, Sept. 2007.