VISUALIZATION OF USER INTERESTS IN ONLINE MUSIC SERVICES

Jingxian Zhang and Dong Liu

CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System
University of Science and Technology of China, Hefei 230027, China
E-mail: junezjx@mail.ustc.edu.cn, dongeliu@ustc.edu.cn

ABSTRACT

Online music services have been popular for end users to obtain music, where user interests, as reflected by their downloading records, are crucial for service providers to understand users and thus to provide personalization. However, the raw downloading records are of huge volume and difficult to analyze intuitively. We study a visualization approach to analyzing downloading records so as to present user interests. To reveal the underlying relevance between music tracks, we utilized not only the metadata of music (especially genres), but also collaborative relevance that is voted by users. To present time varying user interests, we designed several new figures, namely Bean plot, Instrument plot, and Transitional Pie plot, that are capable in displaying different aspects of user interests variation. We have performed experiments with a real-world data set, and the results show the effectiveness of our proposed visualization method. Our work is also inspiring for visualization of time varying data in other applications.

***Index Terms*—** Collaborative relevance, online music, time varying, user interests, visualization.

1. INTRODUCTION

With the rapid development of the Internet and smart phones, getting music from online music services is becoming more and more convenient for people. Examples of online music services include Apple iTunes (www.apple.com/itunes), Nokia OVI (www.music.ovi.com), Pandora Internet Radio (www.pandora.com), and so on. In these services, personalization has been a key feature to satisfy users' needs and to provide better user experiences, therefore, users' downloading records have been extensively analyzed so as to understand users' interests or preference.

Typically, in music and other content services, users' interests are not static but rather time varying. And these interests and their temporal drifts could be revealed by analyzing downloading records of users. Visualization could be a promising way to render such information. However, music downloading records are of huge volume and raise some challenges to be visualized. First, music tracks themselves are not easy to visualize as music has multiple features including metadata (release year, genre, artist, etc.) and acoustic features. Second, to analyze and display user interest drifts, the relevances between music tracks are necessary, but how to quantify and visualize the relevances is a difficulty. Last but not the least, visualization of temporal and drifting data remains a challenge especially taken into account the interpretability and perceivability. To the best of our knowledge, there has not been any systematic visualization of music downloading records that is capable in displaying user interest drifts.

In this paper, we study a visualization approach to the analysis of music downloading records, in order to present user interests especially user interest drifts. To describe music tracks and quantify the relevances between them, we utilize not only the metadata of music (notably genres), but also *collaborative relevance*, defined as the votes given by users (virtually successive downloads). Indeed, we have been inspired by the content-based and collaborative filtering approaches in recommender systems [1]. To present time varying user interests, we designed several new types of plots, namely Bean plot, Instrument plot, and Transitional Pie plot, that are capable in displaying different aspects of user interest drifts. In Bean plot, one user's downloading records are split into sessions and visualized in an intuitive manner. Instrument plot contains temporal drifts as well as relevances and displays much more diverse interest patterns. Transitional Pie plot is good at displaying the summary statistics of users' downloading records in terms of distribution and variation of interests. We have performed experiments with a real-world data set provided by an online music service, and users' interest drifts can be easily perceived.

The rest of the paper is organized as follows. Section 2 introduces Bean plot. Section 3 presents the definition of collaborative relevance and introduces Instrument plot and Transitional Pie plot. Section 4 describes experiments and results. Related work is presented in Section 5, followed by conclusions in Section 6.

This work was supported by Natural Science Foundation of China (NSFC) under No.61303149 and No.61331017, and by the Fundamental Research Funds for the Central Universities under No.WK2100060011.

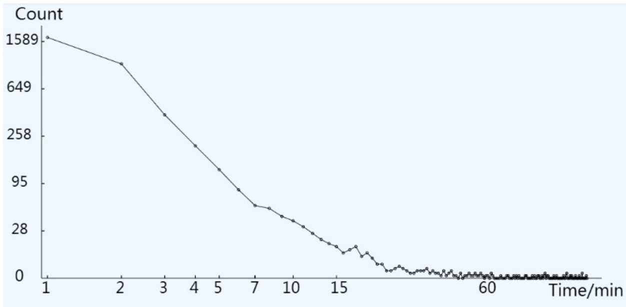


Fig. 1. Statistical distribution of time gaps between successive downloads (log-log plot of *Time* and *Count*)

2. BEAN PLOT

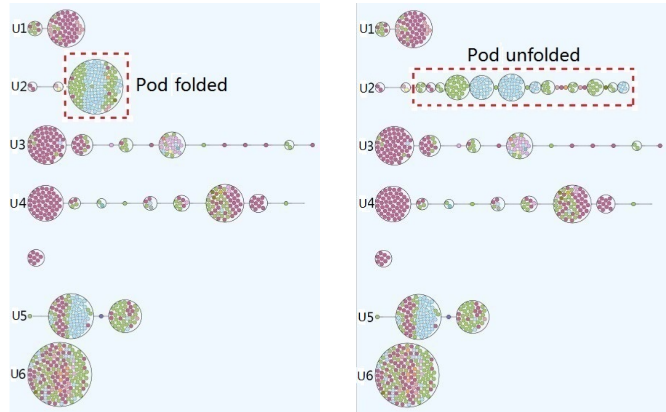
Downloading records that are visualized in this paper can be described as triplets: <user ID, track ID, download timestamp>. For each user, downloading history is indeed temporal data, in which each track may have metadata and we consider two fields of them: genre and release year. Mostly, temporal data are arranged in line in chronological order. But this makes clusters of features obscure and periodical rules uneasy to see. So in this section, we define download sessions to make user interests explicit. And we proposed Bean plot to visualize download sessions. Tracks are divided by download sessions in Bean plot and form figures like beans and pods.

2.1. Download Sessions

Session is a common term in online music services. We hope to present clear download periods by assigning tracks into different download sessions. In order to get the proper rule for dividing download sessions, we counted the time gaps between two successively downloads for all users and got the statistical distribution shown in Fig. 1. In the figure, minute is taken as the minimum unit. Fig. 1 can be seen as a log-log plot of piecewise power-law distribution, which is quite common in online services. It could be figured out from Fig. 1 that most tracks were downloaded in a minute's time after the prior track was downloaded. More than 98% of tracks have download gaps that less than one hour with the successively downloaded tracks. So we may conclude that a time gap of less than one hour dominates the graph. Thus we define a download session as where each track (except the first one) have a download time gap of less than or equal to one hour to its prior track.

2.2. Bean Plot

After dividing download records into sessions, we proposed Bean plot to present download periods explicitly. In Bean plot, tracks that belong to a download session are put in one



(a) Folded Pods (b) Unfolded Pod

Fig. 2. Bean plot showing downloading records of 6 users (U1∼U6 in the plot). (a) Each bean (small, color filled circle) represents a track and each pod (disc with variable size) represents a download session. The color of bean indicates the genre of the track, following the color codes shown in Fig. 3. Pods are arranged in chronological order. (b) Once clicked, one pod will unfold to multiple smaller pods also arranged in chronological order, each of which has single genre.

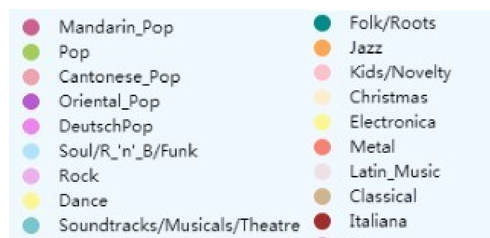


Fig. 3. Color coding of genres (showing partial)

disc to reveal clusters and interest patterns. Tracks of the same download session are just like beans in a pod, so we name it Bean plot.

Bean plot (see Fig. 2 (a)) arrange tracks in line by download sessions. For a user, pods that contain one or more beans lie in a line, with each bean representing a music track the user has downloaded and each pod standing for a download session. Different colors stand for different genres, and the color mapping rule is shown in Fig. 3 (the same color mapping is used for all the following plots). Beans in a pod are arranged in rows from left to right, each row arranging beans from top to bottom in chronological order. If two tracks are not in a download session and they have a download time gap of more than an hour but less than or equal to 24 hours, there will be a three-unit distance between them. Otherwise, there will be a five-unit distance between them in line.

Furthermore, we designed an interaction on the Bean plot to enable zoom-in examination of sessions. When a pod is clicked, it will unfold to several smaller pods by genre in chronological order as shown in Fig. 2 (b), presenting details of user's download session. Each smaller pod represents a "sub"-session, in which successively downloaded tracks are of the same genre.

3. INSTRUMENT PLOT AND TRANSITIONAL PIE PLOT

Bean plot emphasizes download sessions and utilizes pods to visualize interest patterns, but it doesn't take another important characteristic, relevances between music tracks, into consideration. A wealth of information could be revealed by the relevances between tracks and contribute to the analysis of user interests. Clear about the usefulness of relevances among tracks, how to define relevance become the problem of primary importance. In this section, we propose *collaborative relevance* among music tracks as a solution of this problem and then take it into consideration in Instrument plot and Transitional Pie plot.

3.1. Collaborative Relevance

There are several approaches to quantify the relevances between music tracks, and one of them is based on the features of music. However, music has multiple features, many of which are neither structured nor explicit. Even for explicit ones such as genre, it remains a difficulty to quantify relevances based on them.

Inspired by the collaborative filtering in recommender systems [2], we propose collaborative relevances that are voted by users. The basic assumption of collaborative filtering is that: if many people like both items A and B and a new user likes A, then he/she probably also likes B. In other words, the relevance between two items is reflected by the number of users who like both. We use the same idea with two remedies: first, "users who like both" is replaced by "users who downloaded both within a short period," since user interests may drift in our work; second, not only direct relevance but also indirect relevance is defined to deal with data scarcity.

We define two kinds of collaborative relevance: *direct collaborative relevance* ($diRele$) and *indirect collaborative relevance* ($inRele$) (see Fig. 4). If there are N users and S_i is the set of tracks that the ith user has downloaded (i is from 1 to N). For two different tracks a and b, we define function $R(a, b, i)$, $diRele(a, b)$ and $inRele(a, b)$ as

$$R(a,b,i) = \begin{cases} 1 & \text{if } a \text{ and } b \text{ are both in } S_i \text{ and} \\ & \text{downloaded within an hour} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$diRele(a,b) = \sum_{i=1}^{N} R(a,b,i) \tag{2}$$

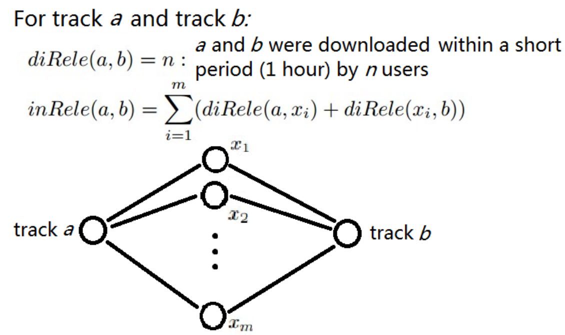


Fig. 4. *Direct collaborative relevance* ($diRele$) and *indirect collaborative relevance* ($inRele$) between two tracks

$$inRele(a,b) = \sum_{i=1}^{N} \sum^{x_i \neq a,b} (R(a, x_i, i) + R(x_i, b, i)) \tag{3}$$

$inRele$ can't show as much relevance as $diRele$, so we define relevance ($Rele$) between two tracks as

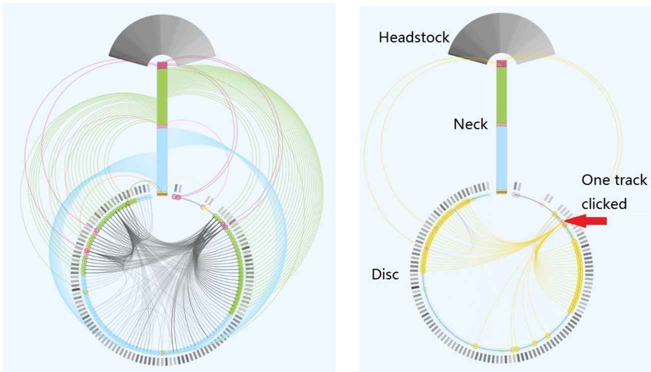
$$Rele(a,b) = diRele(a,b) + 0.25 * inRele(a,b) \tag{4}$$

3.2. Instrument plot

Similar to Bean plot, Instrument plot arranges tracks in chronological order. In addition to this, Instrument plot also presents tracks by statistical distribution and takes collaborative relevance into consideration. It adds Bezier curves, whose greylevels show the degree of collaborative relevance, to connect tracks. We designed the layout of this plot to mimic Pipa (a Chinese traditional instrument), and thus named it Instrument plot (as shown in Fig. 5).

Tracks user has downloaded are arranged in a disc (body of the instrument) in chronological order. The gray strips beside each small round stand for the release year of tracks (higher greylevel presents older release year). The distance between tracks stands for division of sessions and is in accord with settings in Bean plot. In each download session, we connect every two tracks that have collaborative relevance by Bezier curves. Greylevel of the Bezier curves are related to $Rele$ of tracks linearly. The statistical distribution of genre are presented as the neck of the instrument, showing user's download preference. The statistical distribution of release years (still, higher greylevel stands for older release year) are shown as the headstock of the instrument. For each music track in the body of the instrument, we allocate a position in the neck of the instrument and draw a semi-circle between the two points.

When a track is clicked in Instrument plot, tracks related to it will be highlightened as an interaction. Here we define



(a) (b)

Fig. 5. Instrument plot. (a) Disc of the instrument shows tracks in time order, where color indicates genre following Fig. 3. Bezier curves inside the disc with different graylevels display the collaborative relevances between tracks. Neck of the instrument shows statistical distribution of genres, which are connected to the tracks on the disc. Headstock shows statistical distribution of release years of tracks. (b) Once clicked, one track and all its related tracks will be highlighted.

related as (1) having same genre with the track (2) having collaborative relevance with the track. Thus the relevance between tracks could be more impressive and interest patterns could be clearer.

3.3. Transitional Pie plot

Transitional Pie plot is derived from pie chart, with the adding of transfer characteristics and collaborative relevance. Pie chart is widely used and performs well in showing statistical distribution, but temporal information hardly can be presented by it. By designing Transitional Pie plot, we reserved the merit of pie chart and added elements to present temporal information and connection of music tracks.

In Transitional Pie plot, tracks are arranged in a disc by statistical distribution of genres. For each genre, tracks are arranged in chronological order and each track has its corresponding position. We draw a Bezier curve between two successively downloaded tracks outside the disc if they are of different genres. The curves have gradient color identifying that tracks have transferred from one genre to another. Inside the disc, we draw Bezier curves to show $Rele$ between tracks, where greylevels show the strength of them.

From Bezier curves outside the disc, transfer characteristics of downloading are depicted by the density and color of curves. Intensive curves in one area present the frequent shifts between genres. Area with little curves denotes lasting interest in genres. Bezier curves inside the disc depict collaborative relevance. By analyzing transfer characteristics outside the disc along with collaborative relevance inside it,

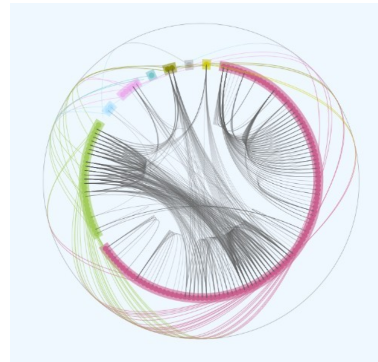


Fig. 6. Transitional Pie plot. Similar to pie plot, the disc shows statistical distribution of genres. Within each genre, tracks are arranged in time order. Bezier curves inside the disc display the collaborative relevances as in Fig. 5. The curves outside the disc display the transitions between genres, i.e. each curve connects two successively downloaded tracks if they are of different genres.

Transitional Pie plot provides a comprehensive view to identify interest patterns.

4. EXPERIMENTAL RESULTS

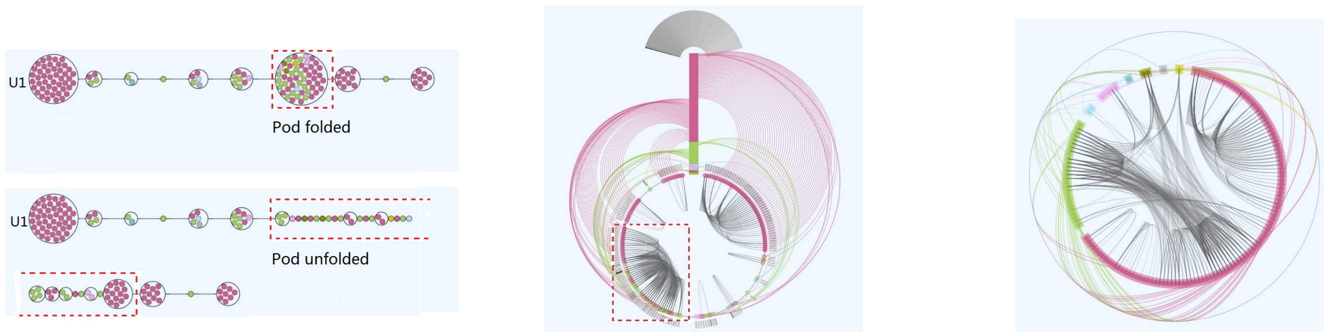
4.1. Implementation

We implemented a visualization interface that enables user interactions by HTML5 and JavaScript techniques. In the interface, there is a main canvas labeled *Music Graph* for plots, where buttons are provided to switch between three plots and to change users. Another canvas *Music Information* displays some more information of users or tracks that are clicked. We have submitted a video clip in addition to this paper which captured the plots and interactions of our implemented visualization. Please check the video for more information.

4.2. Results and Analyses

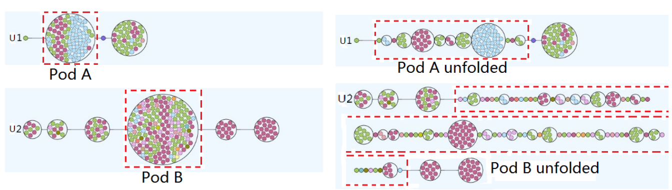
4.2.1. Results of Single User

Fig. 7 presents three plots of a user. In Bean plot (Fig. 7 (a)), in first half of session 7, no obvious layers could be seen, but in the next half, a lengthways layer is obvious. We could see that the user had a lasting interest in mandarin pop in the second half of session 7 and session 1, 8, 9. As for Instrument plot (Fig. 7 (b)), it could be figured out that the user prefers *mandarin pop* music from the headstock. Intense Bezier curves between mandarin pop and pop tracks outside the disc in Transitional Pie plot (Fig. 7 (c)) shows the rapid alternation between these two genres. Both Fig. 7 (b) and (c) shows dense curves inside the disc, presenting even though the user downloaded tracks of two genres which altered rapidly, there are relevance among these tracks.



(a) Folded and Unfolded Bean plot (b) Instrument plot (c) Transitional Pie plot

Fig. 7. Three plots of one user. As the first impression, mandarin pop (coded purple) is the major interest of this user. From the Bean plot, especially the unfolded pod in (a), we observe alternate genres of purple and green, which can also be perceived from the outer-disc curves in (c). However from the collaborative relevances shown in (b) and (c), we notice the "purple" and "green" tracks of this user are highly relevant. Note the circled part in (b) that corresponds to the pod in (a), in this pod (session), actually user's interest does not change.



(a) Folded Pods (b) Unfolded Pods

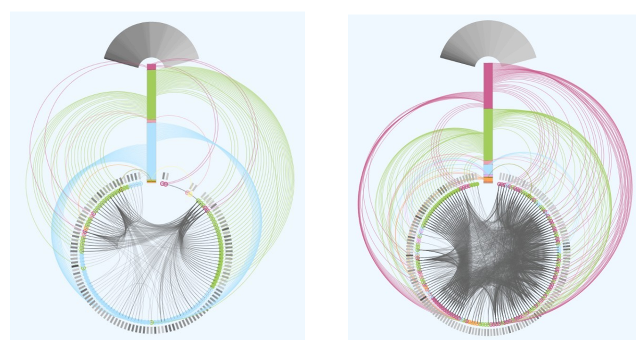
Fig. 8. Bean plots of two users. (a) Pod A has obvious lengthways layers of different genres, while pod B looks jumbled. They represent two patterns of interest drifts. (b) Unfolded pods show the two patterns more clearly.

We may conclude that the user has a main interest in *mandarin pop* music. His interest has drifted to different genres, but the Bezier curves show the drifts are interpretable.

4.2.2. Comparison between Users

As is shown in Bean plot in Fig. 8 (a), lengthways layers could be obviously seen in pod A, while there are no evident layers in pod B. When we unfold pod A and B, we could get the result shown in Fig. 8 (b). There are apparent sessions of interest shift from mandarin pop to soul in unfolded pod A, while in unfolded pod B there are no lasting interest of a genre except for the largest purple pod, which matches the conclusion got from folded pods in Fig. 8 (a).

From the headstocks of Fig. 9 (a) and (b), we find that user A prefers older tracks. From the necks, we observe that user A loves "green" and "blue" while user B prefers "purple" and "green". Inside the body of string instrument in Fig. 9 (b), dense Bezier curves connect tracks, especially in the first half. So we could conclude, even though these tracks are of different genres and no lasting interest patterns are shown, the



(a) User A (b) User B

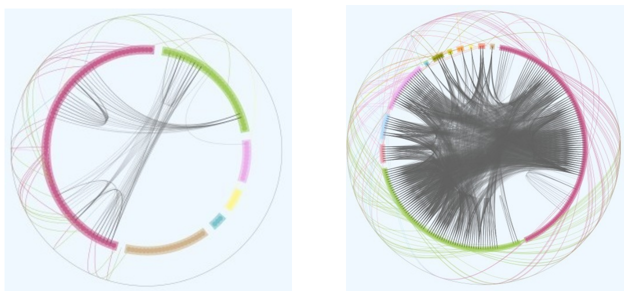
Fig. 9. Instrument plots of two users. We observe that user A has clearly 3 segments of genres (green, blue, and green again) that all last for a while. Her interest drifts from green to blue and then from blue to green, with not much relevances between green and blue. In contrast, the tracks of user B have diverse genres, but are highly relevant to each other. Thus we may claim that user B's interest is more concentrated.

user's interest doesn't drift unreasonably.

As is shown in Transitional Pie plot (Fig. 10 (b)), there are intense transfer curves between mandarin pop and pop tracks, showing the rapid alternation of genres in time order. Inside the disc, dense relevance curves tell us that even though the user downloaded tracks of several genres which altered rapidly, there are strong relevance among them.

4.2.3. Discussion

Among the three figures, Bean plot focuses on sessions and expresses user's interest shift. The layers in pods show interest shift obviously. The pods arranged in the line also present



(a) User A (b) User B

Fig. 10. Transitional Pie plots of two users. User A has less transitions between genres, i.e. each genre stably lasts for a while. User B has intense transitions between genres especially between purple and green, but the tracks of user B have higher relevances than those of user A.

the density of user's download. Instrument plot shows connection between time-order data and statistical distribution. Collaborative relevance and connection between chronological arrangement and statistical distribution are presented, revealing extra information that Bean plot and Transitional Pie plot don't show. User's preference of genre and year are also shown macroscopically. Transitional Pie plot is sensitive to frequent download alternation among genres. Transfer characteristics of download are obviously depicted from the Bezier curves outside the disc. Also, the collaborative relevance inside the disc along with the transfer characteristics outside the disc could connect tracks of different genres and better interpret user's interests.

5. RELATED WORK

There are many researches focus on visualizing temporal data in many aspects, but few of them visualized users' temporal information in online music service. An interactive visualization methodology is presented for dynamic social networks to analyze the community structure in the US House of Representatives [2]. Temporal rings and scatterplots are used to visualize online customer opinions in hotel reviews [3], which makes opinion pattern obvious and enables a side-by-side comparison. Linear and cyclical way to visualize time are presented to visualize multi-dimensional data [4].

Song similarity is a very high-dimensional measure that can incorporate many different aspects of music like melody, harmony, genre, etc. Highlevel semantic descriptions of music tracks are used to model user preference [5]. Three methods of visualization, disc visualization, rectangular visualization and tree-map visualization were designed to visualize personal music libraries [6]. As for the color coding part, Holm found that a default color-genre mapping can be suitable for a given country or region[7].

6. CONCLUSION

In this paper, we proposed a visualization approach to analyzing user interests in online music services. The visualization was based on users' downloading records. Inspired by the collaborative filtering for recommender systems, we defined *collaborative relevance* (including direct and indirect collaborative relevances) to help interpret users' interest patterns. Collaborative relevance can amend the analysis of user interests when the metadata (notably genres) had changed but actually user's interests not. We designed three plots, namely Bean plot, Instrument plot, and Transitional Pie plot to visually present user interest drifts. The three plots have different emphases and display different aspects of user interests. We performed experiments with a real-world data set and the results demonstrate the effectiveness of our proposed methods.

As future work, we plan to perform user study to investigate the usability of our designed plots and try to improve them. Combining the visualization of music itself and that of downloading records is also an interesting topic. Moreover, how to design visually pleasing figures to display aggregated data (of multiple users) is still an open problem and will interest the service providers.

7. REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowledge Data Engineering*, vol. 17, no. 6, pp. 734–749, June 2005.

[2] K. Reda, C. Tantipathananandh, A. Johnson, J. Leigh, and W. Berger, "Visualizing the evolution of community structures in dynamic social networks," *Comput. Graph. Forum*, vol. 30, pp. 1061–1070, 2011.

[3] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu, "Opinionseer: Interactive visualization of hotel customer feedback," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1109–1118, 2010.

[4] J. Kehrer and H. Hauser, "Visualizing the evolution of community structures in dynamic social networks," *IEEE Trans. Vis. and Comput. Graph.*, vol. 19, pp. 1061–1070, 2013.

[5] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambo, E. Gomez, and P. Herrera, "A content-based system for music recommendation and visualization of user preferences working on semantic notions," in *9th Int. Workshop on Content-based Multimedia Ind.*, 2011, pp. 249–252.

[6] M. Torrens, P. Hertaog, and J.-L Arcos, "Visualizing and exploring personal music libraries," in *Proc. of 5th Int. Conf. on Music Info. Retrieval*, 2004, pp. 421–424.

[7] J. Holm, A. Aaltonen, and H. Siirtola, "Associating colours with musical genres," *Journal of New Music Research*, vol. 38, no. 1, pp. 87–100, 2009.