

Sequences of Diverse Song Recommendations

An exploratory study in a commercial system

Nava Tintarev
Delft University of Technology
Delft, The Netherlands

Christoph Lofi
Delft University of Technology
Delft, The Netherlands

Cynthia C. S. Liem
Delft University of Technology
Delft, The Netherlands

ABSTRACT

This paper presents an exploratory study of the perceptions users have of diversity and ordering in playlist recommendations. There is a match between the diversification approach used in the system, and importance that users placed on the item properties. Surprisingly, participants had no expectations of the songs being in a particular order in a playlist. We discuss possible explanations for this finding, refining the research agenda to consider which ordering choices are perceptible to users, and influence user satisfaction.

1 INTRODUCTION

Recommender systems suggest users items to consume, try, or buy, by learning from our past interactions, inferring our interests, and making predictions. A criticism of recommender systems has been that they "over-personalize": censoring user choices over time and effectively polarising users' preferences [2, 7].

One approach that addresses this relative uniformity of recommendation is diversification of items within a list. Diversifying the items in lists has been found to increase user satisfaction [10], to help users to find target items faster [3], and to increase novelty of the items that are recommended [9].

Previous work has considered how to automatically generate playlists [1, 4, 5, 8], but to the best of our knowledge this work did not also consider ordering and diversity, or whether the improvements in accuracy were beneficial from a user perspective.

To support discovery, music recommenders need to strike a delicate balance between novelty and relevance. The challenge is thus to define the measure of similarity in a way that maintains relevance while at the same time sufficiently diversifying items within the recommendation sequence.

This paper raises the question of what happens in domains, like music, where the items are usually consumed in sequence. A sequence not only affords the recommender system more chances to make accurate recommendations, but to mix familiar and unfamiliar items. However, this situation also creates new challenges for recommendation quality and user satisfaction (i.e., two given items are good recommendations when considered in isolation, but create a poor experience when consumed in sequence). In contrast, recommendation lists are commonly seen as top-N recommendations where the user selects one or only a handful of items from a list of N items. Therefore, this paper presents an exploratory study

with users of a live recommender systems which presents playlists, focusing on the perspective of users, addressing:

- How users perceive diversity: Which items should be kept familiar, which should be diversified?
- How users perceive the role of ordering: What expectations do they have?

This work is a novel first step toward understanding user requirements for diverse sequence construction.

2 USER STUDY: SONG LIST EVALUATION

The study was distributed as an online survey for users of Spotify and the Discover Weekly service. The *Discover Weekly* service is a recommender system which delivers a new personalized playlist to a user every week. This service is selected only as an example of a recommendation engine for which sufficient user preferences are collected (and not constructed artificially for research purposes) to produce user-tailored recommended lists of songs. Participants were able to listen to the songs when completing the survey.

2.1 Participants

28 participants started the experiment, and 20 completed it (8 participants excluded). Of these, 8 participants were female, and 12 male. Most of the users were established users of Spotify, with at least 3-6 months of use. Only 1 participant had used the system for less than 3 months, and 17 of the participants had used the system for more than a year. Participants took a mean of 7 minutes to complete the study.

2.2 Results

Given the exploratory nature of the study, and the small sample size these results are descriptive (i.e., no statistical comparisons).

Overall Diversity. Half of the participants (10) agreed with the statement "The recommendations consist of a good variety of songs. 5=Strongly Agree, 1=Strongly Disagree" (mean= 3.15, SD= 1.09). However, seven participants said that they Disagree or Strongly Disagree. Two participants supplied comments: "They seem to be all of the same genre."; "The variety leans really [sic] pop or emo heavy I find, even though I listen to a variety of music."

This suggests that while many users may be happy with the variety and range of their playlists, there are users for which the way the service applies diversity is not fully satisfactory. This can be due to limitation in the amount or type of preference data available for a given user, but can also be due to other factors that determine the composition of the list. The users mentioned genre and context as factors that impacted how they expected diversity to be applied.

To further understand what constitutes satisfactory, or unsatisfactory perceived diversity, we investigated the properties of the

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

UMAP '17, July 9–12, 2017, Bratislava, Slovakia

© 2017 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-4635-1/17/07.
DOI: <http://dx.doi.org/10.1145/3079628.3079631>

Feature	Percentage of songs
Genre	68.33 (32.38)
Artist	21.83 (15.58)
Title	17.33 (18.50)

Table 1: Percentage of songs (out of a list of 30) that are familiar with regard to each feature.

Feature	% Users	% Responses
Fam. Genre	65.00	23.64
Fam. Artist	20.00	7.27
Fam. Title	5.00	1.82
New Music	95.00	34.55
Variety	75.00	27.22
Ordering	15.00	5.45

Table 2: Percentage of responses to the question “Which aspects are important for you to include in this playlist?”. Since a participant could select several options as important % Responses normalizes for the number of responses.

lists they were given. Specifically, we asked for participants to indicate the number of songs containing familiar artists, familiar titles, and familiar genres. For example, for artists we asked “Roughly, how many of the songs recommended to you this week have an Artist that seems familiar (or ‘rings a bell’) to you? (0-30 songs)”. We see in Table 1 that genre is the most familiar factor, followed by artist, and (song) title. Although title has the lowest percentage, this is equivalent to 5 out of 30 songs.

Diversity Preferences. In the previous section we saw how diversity is applied by the system across our participants, and we know that for most of the participants this results in good perceived variety and range. These results do not give us an indication of how this way of applying diversity relates how to people perceive different types (genre, artist, title) of diversity. We therefore asked participants to tell us which of the item features were important for this particular playlist (“Which aspects are important for you to include in this playlist?”). We asked about: familiar artists, familiar titles, familiar genres, ordering of the items, overall variety, and the presence of new music. Participants were able to select several of these if they felt several of them were important. Table 2 summarizes the percentage of how many participants selected each feature, as well as the overall percentage of responses (taking into account that each user could give several responses).

We see here that the proportion of importance for genre, artist and title, is similar to the proportions said to be familiar in the sets for these features (c.f. Table 1).

About two-thirds of participants expected to see familiar genres, suggesting that playlists that contain some familiar genres may be well received by a majority of users. In contrast, only 4 participants marked that familiar artists were important, and only 1 thought that familiar titles were important. Despite the smaller sample and the subjective nature of self reporting, this gives an indication that the current diversification strategy reflects users’ diversity needs.

We also see that three-quarters of participants felt that the list needed to be diverse, and that almost all participants felt that the songs had to be novel to them. Therefore for this service novelty appears to be more important than diversity within the list.

Ordering. In response to the statement, “The songs are listed in the ideal order” (Strongly Agree=5, Strongly Disagree=1), many participants did not perceive an ordering (mean = 2.85, SD = 0.75). User comments suggests that this is not something they expect: “Hunh. Never thought about it as I listen on shuffle usually”; “Never paid attention to the order, often listen in shuffle”; “Are they in a particular order?”.

3 CONCLUSIONS AND FUTURE WORK

This work is part of a larger agenda of understanding user requirements for constructing *sequences of items*, in which familiar and unfamiliar items can be mixed, and considers ordering. *Surprisingly, participants did not expect the recommended list to be in a particular order, nor did they find this property to be important.* Users do not seem to be consciously seeking a sequential narrative (i.e., they might be using the system for background listening), even if songs can (and are likely to) be consumed in a sequence.

Consequently, this pilot study enables us to formulate new hypotheses about user perceptions of ordering in diverse sequences. Firstly, when users consider ordering they may be looking for something in particular, like an increase in tempo across songs, or musical/thematic progression (which have been found to be important in user generated playlists [6]). It may also be more important to identify constraints on ordering than to seek an optimal ordering. There are orderings that would be unsuitable for most listeners, like playing songs in very different genres, or highly different tempos in direct succession.

Future work will study which types of ordering choices are perceptible to users, and influence satisfaction. It will investigate the role of varying the distance of songs, and the number of intermediate songs that are needed to create a “smooth transition”.

REFERENCES

- [1] Natalie Aizenberg, Yehuda Koren, and Oren Somekh. 2012. Build Your Own Music Recommender by Modeling Internet Radio Streams. In *WWW*. 1–10.
- [2] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348 (2015), 1130–1132.
- [3] Derek Bridge and John Paul Kelly. 2006. Ways of Computing Diverse Collaborative Recommendations. In *Adaptive Hypermedia and Adaptive Web-based Systems*. 41–50.
- [4] Shuo Chen, Joshua Moore, Thorsten Joachims, and Douglas Turnbull. 2012. Playlist Prediction via Metric Embedding. In *SIGKDD*. 714–722.
- [5] Balzs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based recommendations with recurrent neural networks. In *International Conference on Learning Representations (ICLR)*.
- [6] Brian McFee and Gert RG Lanckriet. 2011. The Natural Language of Playlists. In *International Conference on Music Information Retrieval (ISMIR)*. 537–542.
- [7] Dimitar Nikolov, Diego FM Oliveira, Alessandro Flammini, and Filippo Menczer. 2015. Measuring online social bubbles. *PeerJ Computer Science* 1 (2015), e38.
- [8] Roberto Turrin, Andrea Condorelli, Paolo Cremonesi, Roberto Pagano, and Massimo Quadrana. 2015. Large scale music recommendation. In *Workshop on Large-Scale Recommender Systems (LSRS 2015) at ACM RecSys*.
- [9] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 13–22.
- [10] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists Through Topic Diversification. In *World Wide Web (WWW)*.