

Modeling Artist Preferences of Users with Different Music Consumption Patterns for Fair Music Recommendations

Dominik Kowald*
Know-Center GmbH
Graz, Austria
dkowald@know-center.at

Elisabeth Lex*
Graz University of Technology
Graz, Austria
elisabeth.lex@tugraz.at

Markus Schedl
Johannes Kepler University Linz
Linz, Austria
markus.schedl@jku.at

ABSTRACT

Music recommender systems have become central parts of popular streaming platforms such as Last.fm, Pandora, or Spotify to help users find music that fits their preferences. These systems learn from the past listening events of users to recommend music a user will likely listen to in the future. Here, current algorithms typically employ collaborative filtering (CF) utilizing similarities between users' listening behaviors. Some approaches also combine CF with content features into hybrid recommender systems [2].

Problem & objective. While music recommender systems can provide quality recommendations to listeners of mainstream music artists, recent research [6, 9] has shown that they tend to discriminate listeners of unorthodox, low-mainstream artists. This is foremost due to the scarcity of usage data of low-mainstream music as music consumption patterns are biased towards popular artists [2, 6].

Thus, the objective of our work is to provide a novel approach for modeling artist preferences of users with different music consumption patterns and listening habits. We focus on three user groups: (i) LowMS (i.e., listeners of unorthodox, niche music), (ii) HighMS (i.e., listeners of mainstream music), and (iii) MedMS (i.e., listeners of music that lies in between). The main problem we address in this work is how to exploit variations in listening habits to avoid discrimination of users, whose listening behavior differs significantly from the mainstream. With that, we aim to realize fair music recommendations in the sense that recommendations are not biased towards the mainstream.

Approach & method. In our work, we model user listening behavior on the level of music artists to describe a user's music taste. Since a user's music artist preferences may change over time [7], we take temporal drifts of a user's music listening habits into consideration.

To do so, we utilize the Base-Level Learning (BLL) equation from the cognitive architecture ACT-R [1] to model music listening habits. The BLL equation accounts for the time-dependent decay of item exposure in human memory. It quantifies the usefulness of a piece of information based on how frequently and how recently it was accessed by a user and models this time-dependent decay using a power-law distribution. We have utilized the BLL equation in our previous works to recommend tags in social bookmarking systems [4] and to recommend hashtags in Twitter [5].

*Both authors contributed equally to this work.

Group	$ U $	$ A $	$ LE $	$ A/U $	$ MS $
LowMS	1,000	82,417	6,915,352	239	.125
MedMS	1,000	86,249	7,900,726	496	.379
HighMS	1,000	92,690	8,251,022	1,194	.688

Table 1: Dataset statistics of the LowMS, MedMS, and HighMS Last.fm user groups. Here, $|U|$ is the number of distinct users, $|A|$ is the number of distinct artists, $|LE|$ is the number of listening events, $|A/U|$ is the average number of artists listened by a user and $|MS|$ is the average mainstreamness value per user group.

In the present paper, we build upon these results, and we adopt the BLL equation to model the listening habits of users in our three groups and predict their music artist preferences. We name our approach BLL_u and demonstrate the efficacy of BLL_u using the *LFM-1b* dataset [8], which contains listening histories of more than 120,000 Last.fm users, amounting to 1.1 billion individual listening events over nine years¹.

Additionally, the dataset contains demographic data such as age and gender as well as a "mainstreamness" factor [10], which relates a user's artist preferences to the aggregated preferences of all users (i.e., the mainstream). Based on this factor, we assign the users in our dataset to one of the three groups: (i) LowMS, (ii) MedMS, and (iii) HighMS. Thus, the 1000 users with the lowest mainstreamness are in the LowMS group, the 1000 users with a mainstreamness value centered around the median are in the MedMS group, and the users with the highest values are in the HighMS group. We summarize the dataset statistics of these groups in Table 1 and evaluate our proposed BLL_u approach for all three user groups.

Contributions & results. The contributions of our work are two-fold. Firstly, we propose our BLL_u approach that is designed to model and predict artist preferences to provide personalized, fair music recommendations. As our notion of fairness is related to popularity bias (i.e., of the mainstream), we model the user's preference for an artist by considering how often this individual user has listened to this artist. Additionally, since music preferences are dynamic, we incorporate the user's temporal drifts of artist preferences into our model. Secondly, we evaluate our approach on three different groups of Last.fm users based on the distance of their listening behavior to the mainstream: (i) LowMS, (ii) MedMS, and (iii) HighMS.

For our evaluation, we follow good practice in the field of information retrieval and recommender systems by splitting our user groups into train and test sets. We employ a time-based split, and we put the 1% most recent listening events of each user into the test

¹The dataset is freely available via <http://www.cp.jku.at/datasets/LFM-1b/>

set and keep the remaining listening events for training. For our three user groups, this procedure leads to three test sets with 68,651 listening events for LowMS, 78,511 listening events for MedMS, and 82,030 listening events for HighMS. For each user, we aim to predict the artists in these listening events².

Figure 1 illustrates the results of our evaluation in form of recall/precision plots. Here, we compare BLL_u to four baselines: (i) TOP , which recommends the most popular artists of all users, (ii) CF_u , which recommends artists using collaborative filtering, (iii) POP_u , which recommends the most popular artists of a specific user, and (iv) $TIME_u$, which recommends the artists a particular user has listened to most recently. We find that for all groups, BLL_u leads to the best accuracy results for predicting music artists and provides especially good results for the LowMS group. Interestingly, we also find that the time-based approach $TIME_u$ provides even better accuracy results than BLL_u when only predicting 1 or 2 artists.

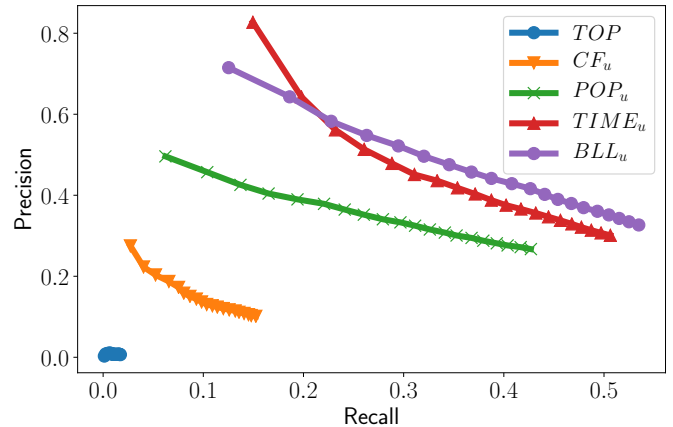
Future work. We plan to extend our analysis to include more sophisticated mainstreamness measures based on rank-order correlation or Kullback-Leibler divergence [9] since our current mainstreamness measure is rather simplistic. Furthermore, we aim to integrate our findings into fair music recommendation algorithms (e.g., for songs), with particular attention to avoid discrimination of the low mainstreamness group, since standard collaborative filtering approaches do not provide suitable music recommendations for this user group [10]. For example, we plan to integrate the BLL-preference values we obtain for a specific user and a particular artist via our approach as a context dimension into a matrix factorization-based approach such as the one presented in [3].

Keywords. User Groups; Music Preference Prediction; Fair Music Recommendations; Discrimination; Time-Aware Recommendations

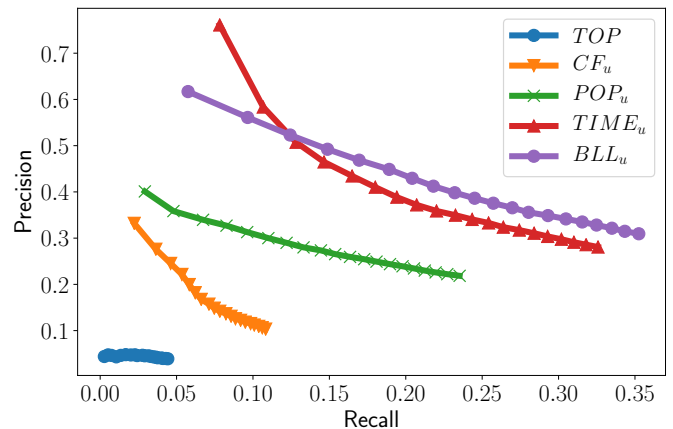
REFERENCES

- [1] John R Anderson, Daniel Bothell, Michael D Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. 2004. An integrated theory of the mind. *Psychological review* 111, 4 (2004), 25 pages.
- [2] Óscar Celma. 2010. *Music Recommendation and Discovery – The Long Tail, Long Tail, and Long Play in the Digital Music Space*. Springer.
- [3] Noam Koenigstein, Gideon Dror, and Yehuda Koren. 2011. Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of RecSys’2011*. ACM, 165–172.
- [4] Dominik Kowald and Elisabeth Lex. 2016. The Influence of Frequency, Recency and Semantic Context on the Reuse of Tags in Social Tagging Systems. In *Proceedings of Hypertext’2016*. ACM, New York, NY, USA, 237–242.
- [5] Dominik Kowald, Subhash Chandra Pujari, and Elisabeth Lex. 2017. Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach. In *Proceedings of WWW’2017*. ACM, 10 pages.
- [6] Aäron van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep Content-based Music Recommendation. In *Proceedings of NIPS’2013*. Curran Associates Inc., USA, 2643–2651.
- [7] C. H. Park and M. Kahng. 2010. Temporal Dynamics in Music Listening Behavior: A Case Study of Online Music Service. In *Proceedings of ACIS’2010*. IEEE, 573–578.
- [8] Markus Schedl. 2016. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings of the 2016 Conference on Multimedia Retrieval*. ACM, 103–110.
- [9] Markus Schedl and Christine Bauer. 2018. An Analysis of Global and Regional Mainstreamness for Personalized Music Recommender Systems. *Journal of Mobile Multimedia* 14 (2018), 95–112.
- [10] Markus Schedl and David Hauger. 2015. Tailoring music recommendations to users by considering diversity, mainstreamness, and novelty. In *Proceedings of SIGIR’2015*. ACM, 947–950.

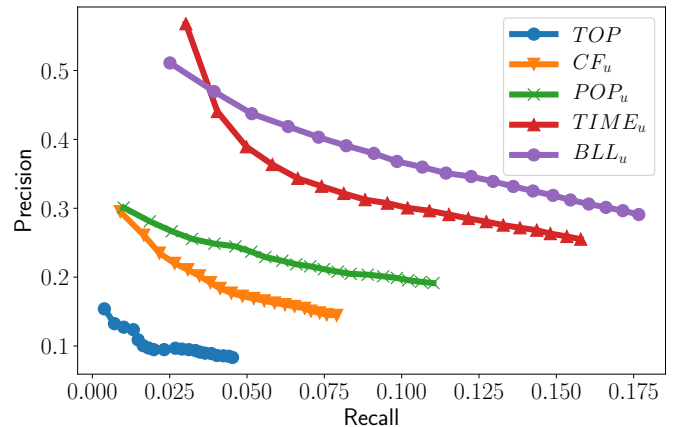
²Our evaluation framework is freely available via <https://github.com/learning-layers/TagRec>



(a) User group: LowMS



(b) User group: MedMS



(c) User group: HighMS

Figure 1: Recall/precision plots of the baselines and our BLL_u approach for the three user groups LowMS, MedMS, and HighMS, and for $k = 1 \dots 20$ predicted artists. We see that BLL_u provides the best prediction accuracy results for all user groups.