## RESEARCH

# Modeling Popularity and Temporal Drift of Music Genre Preferences

Elisabeth Lex[*,†], Dominik Kowald[*,‡] and Markus Schedl[§]

In this paper, we address the problem of modeling and predicting the music genre preferences of users. We introduce a novel user modeling approach, $BLL_u$, which takes into account the popularity of music genres as well as temporal drifts of user listening behavior. To model these two factors, $BLL_u$ adopts a psychological model that describes how humans access information in their memory. We evaluate our approach on a standard dataset of Last.fm listening histories, which contains fine-grained music genre information. To investigate performance for different types of users, we assign each user a mainstreaminess value that corresponds to the distance between the user's music genre preferences and the music genre preferences of the (Last.fm) mainstream. We adopt $BLL_u$ to model the listening habits and to predict the music genre preferences of three user groups: listeners of (i) niche, low-mainstream music, (ii) mainstream music, and (iii) medium-mainstream music that lies in-between. Our results show that $BLL_u$ provides the highest accuracy for predicting music genre preferences, compared to five baselines: (i) group-based modeling, (ii) user-based collaborative filtering, (iii) item-based collaborative filtering, (iv) frequency-based modeling, and (v) recency-based modeling. Besides, we achieve the most substantial accuracy improvements for the low-mainstream group. We believe that our findings provide valuable insights into the design of music recommender systems.

## 1. Introduction

Music recommender systems play a pivotal role in popular streaming platforms such as Last.fm,[1] Pandora,[2] or Spotify[3] to help users find music that suits their taste. Existing music recommender systems typically employ collaborative filtering algorithms based on the users' interactions with music items (i.e., listening behavior or ratings), sometimes in combination with content features (e.g., acoustic features of songs) in the form of hybrid music recommender systems (Celma, 2010; Schedl et al., 2018b).

**Problem.** While music recommender systems can provide quality recommendations to listeners of popular music, related research (Schedl and Bauer, 2018; van den Oord et al., 2013) has shown that they tend to fail listeners who prefer niche artists and genres. A reason for that is the scarcity of usage data of such types of music as music consumption patterns are biased towards popular artists (van den Oord et al., 2013; Celma, 2010; Celma and Cano,

2008). In this paper, we introduce a novel user modeling and genre prediction approach for users with different music consumption patterns and listening habits. We focus on three user groups: (i) LowMS, i.e., listeners of niche music, (ii) HighMS, i.e., listeners of mainstream (MS) 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 improve personalization for all three user groups. We investigate this problem by predicting the music genres a user is going to listen to in the future.

**Approach and methods.** We model the users' listening behavior in terms of fine-grained music genre preferences. To that end, we use behavioral data in the form of listening events, i.e., the listening history of which genres a user has listened to in the past. Our approach is based on the Base-Level Learning (BLL) equation from the cognitive architecture ACT-R (Anderson and Schooler 1991; Anderson et al., 2004) that 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 recently a user accessed it in the past. This time-dependent decay takes the shape of a power-law distribution. Related work has employed the

\* Both authors contributed equally to this work

† Graz University of Technology, Graz, AT

‡ Know-Center GmbH, Graz, AT

§ Johannes Kepler University (JKU) Linz and Linz Institute of Technology (LIT) AI Lab, Linz, AT

Corresponding author: Dominik Kowald (dkowald@know-center.at)

BLL equation to recommend Web links (Fu and Pirolli, 2007), to recommend scientific talks at conferences (Maanen and Marewski, 2009), to recommend tags in social bookmarking systems (Kowald and Lex, 2016), and to recommend hashtags (Kowald et al., 2017b).

In this work, we build upon these results and adopt the BLL equation to model the listening habits of users in our three groups to predict their music genre preferences. We demonstrate the efficacy of our approach on the *LFM-1b* dataset (Schedl, 2016), which contains listening histories of more than 120,000 Last.fm users, amounting to 1.1 billion individual listening events over nine years. The music in this dataset is categorized according to a fine-grained taxonomy that consists of 1,998 music genres and styles. Additionally, the dataset contains demographic data such as age and gender as well as a "mainstreaminess" factor (Bauer and Schedl, 2019) that relates the listening preferences of each user to the aggregated preferences of all Last.fm users in the dataset. Based on this factor, we assign the users in our dataset to one of the three groups, i.e., (i) LowMS, (ii) MedMS, and (iii) HighMS. This allows us to evaluate our proposed $BLL_u$ approach for different types of users.

**Contributions and findings.** The contributions of our work are two-fold. Firstly, we propose the $BLL_u$ approach for modeling popularity and temporal drift of music genre preferences. Secondly, we evaluate $BLL_u$ on three different groups of Last.fm users, which we separate based on the distance of their listening behavior to the mainstream: (i) LowMS, (ii) MedMS, and (iii) HighMS.

We find that for all three groups, $BLL_u$ provides the highest accuracy for predicting music genre preference, compared to five baselines: (i) group-based modeling (i.e., *TOP*), (ii) user-based collaborative filtering (i.e., $CF_u$), (iii) item-based collaborative filtering (i.e., $CF_i$), (iv) frequency-based modeling (i.e., $POP_u$), and (v) recency-based modeling (i.e., $TIME_u$). Moreover, $BLL_u$ gives the highest accuracy improvements for the LowMS group. Finally, we also validate our findings in a cold-start setting, in which we only evaluate users with a small number of listening events. Here, we also find that our $BLL_u$ approach provides the best prediction accuracy results.

**Structure of this paper**. This paper is organized as follows: In Section 2, we review related work, and in Section 3, we describe the dataset as well as statistical analyses about genre mainstreaminess, popularity, and temporal drift of music genre preferences. Also, this section includes the methodology and the proposed approach for modeling music genre preferences. In Section 4, we present the experimental setup as well as the evaluation results. Finally, Section 5 concludes this paper and gives an outlook into future work.

## 2. Related Work

At present, we identify three strands of related research: (i) research on music preferences in light of psychology, (ii) temporal dynamics of music preferences, and (iii) personalization for music recommendation.

**Research on music preferences in light of psychology.** Research in music psychology (North and Hargreaves, 2008) has shown that a range of factors impact music preferences (Schedl et al., 2015), such as emotional state (Cantor and Zillmann, 1973; Juslin and Sloboda, 2001; Rodà et al., 2014), a user's current activity, their self-view and self-esteem (North and Hargreaves, 1999), the cognitive functions of music (e.g., music as a way to communicate and to self-reflect) (Schäfer and Sedlmeier, 2010), as well as personality (Cattell and Anderson, 1953; Arnett, 1992; Dollinger, 1993; Rentfrow and Gosling, 2003; George et al., 2007; Delsing et al., 2008; Dunn et al., 2012; Schedl et al., 2018a).

For instance, Rentfrow and Gosling (2003) showed that the Big Five personality traits (i.e., openness to experience, agreeableness, extraversion, neuroticism, and conscientiousness) influence genre preferences in music and that music preferences can be categorized along specific dimensions (e.g., reflective & complex, intense & rebellious, upbeat & conventional, and energetic & rhythmic music); the structure of music preferences is also discussed by Delsing et al. (2008). Greenberg et al. (2015) found that a person's cognitive approach (i.e., their tendency towards empathy versus systemizing versus balancing both) impacts their music genre preferences. A user's music preference is also impacted by familiarity (Pereira et al., 2011; Schubert, 2007). This has been attributed to the so-called *mere exposure effect* (Peretz et al., 1998), which means that prior exposure can positively influence music liking. In our work, we also incorporate prior exposure (in this case, to a music genre) into our model.

**Temporal dynamics of music preferences.** Music preferences are often dynamic due to variations in user taste (Kim et al., 2018), or evolving music taste (Moore et al., 2013). One can distinguish between research on long-term temporal dynamics of listening behavior and short-term dynamics. Studies investigating long-term dynamics research on, for example, how music preferences of children and young adults evolve (Hargreaves et al., 2015; Leadbeater, 2014), or how user tastes change over time and how artists develop (Moore et al., 2013).

Studies investigating short-term dynamics typically assess users' listening behaviors (Aizenberg et al., 2012; Park and Kahng, 2010) on a fine-granular basis (e.g., time of the day) to detect patterns and periodicity in listening behavior, or in the case of Krause and North (2018), to study the relationship between music preferences and seasons of the year. The latter approaches are typically intended to help create predictive models of music preferences to create playlist recommendations for music streaming services, among others. As we describe in detail in Section 3, in our data, we observe interesting temporal dynamics in users' genre listening histories. Specifically, the time-dependent decay of number of plays per genre follows a power-law distribution, so our users tend to listen to genres to which they have recently listened.

**Personalization for music recommendation.** A number of aspects make personalization in music recommender

systems challenging, such as, e.g., the variability of listening intent and purpose of music consumption, insufficient ratings and usage data, as well as users' tendency to appreciate recommendations of items that have been previously recommended (Schedl et al., 2018b), but also the dependence of music preferences on the user's personality traits or emotional state. In this vein, Selvi and Sivasankar (2019) extracted the user's emotional context from social media messages as well as their current time context and incorporated both to generate personalized music recommendations. Ferwerda et al. (2015) used a specific personality-enriched dataset that provided links to users' listening histories on Last.fm to leverage personality traits to predict a user's genre preferences. Zheng et al. (2018) proposed a tag-aware dynamic music recommendation framework that represents musical tracks via user-generated tags and generates time-sensitive recommendations. Koenigstein et al. (2011) incorporated a temporal analysis of user ratings assigned to music pieces and item popularity trends into a matrix factorization approach to mitigate the issue of insufficient item ratings. The latter is a common problem that causes (music) recommender systems to suffer from bias towards popular items. Due to insufficient amounts of usage data for less popular items, many recommendation algorithms cannot provide useful recommendations for consumers of less popular and niche items (Abdollahpouri et al., 2019; Celma, 2010; van den Oord et al., 2013). Recent work (Vall et al., 2019) has yet provided evidence that deep-learning-based methods (i.e., recurrent neural networks) seem to be less biased towards popular items.

In our work, we use only listening histories as a data source to model user preferences and to generate recommendations. As we show in Section 3, we observe that all users in our dataset tend to consume items they have listened to frequently and recently in the past, where the time-dependent decay of this item consumption count follows a power-law distribution. Correspondingly, the Base-Level Learning (BLL) equation from the cognitive architecture ACT-R (Anderson and Schooler, 1991; Anderson et al., 2004) describes a time-dependent decay of item exposure in human memory in the form of a power-law distribution. Leveraging these similarities between characteristics of music consumption patterns and cognition models (i.e., ACT-R in our case), we propose here to use the BLL equation to describe listeners' behavioral music consumption traces.

## 3. Data and Method
In this section, we present the dataset we use for our study and statistical analyses we carry out. We outline the approach of this work and the baselines, which we employ to validate our proposed method.

### 3.1 Dataset and Statistical Analyses
First, we describe the Last.fm dataset, as well as the selected genre mapping procedure. We report statistical analyses for (i) music genre popularity, (ii) average pairwise user similarity, (iii) popularity of music genre preferences, and (iv) temporal drifts of music genre preferences.

**Dataset description and availability.** For our study, we use a dataset gathered from the online music service Last.fm, namely the *LFM-1b* dataset.[4] *LFM-1b* contains listening histories of more than 120,000 users, totaling to about 1.1 billion individual listening events accrued between January 2005 and August 2014. Each listening event is characterized by a user identifier, artist, album, track name, and a timestamp (Schedl 2016). Besides, the *LFM-1b* dataset contains user-specific demographic data such as country, age, gender as well as additional features such as mainstreaminess, which is defined as the overlap between the user's listening history and the aggregated listening history of all Last.fm users in the dataset. More precisely, the mainstreaminess of a user corresponds to the average distance between all artists' relative frequencies in the user's listening profile and the artists' relative frequencies among all users in the dataset (Schedl and Hauger, 2015).

**Mapping listening events to music genres.** Since we are interested in modeling and predicting music genre preferences, we enhance the listening events in the *LFM-1b* dataset with additional genre information. Therefore, we use an extension of the *LFM-1b* dataset, termed LFM-1b User-Genre-Profile (i.e., *LFM-1b UGP*) dataset (Schedl and Ferwerda, 2017), which describes the genres of an artist in a listening event by exploiting social tags from Last.fm.

Among others, *LFM-1b UGP* contains a weighted mapping of 1,998 music genres and styles available in the online database Freebase[5] to Last.fm artists. In part, this taxonomy includes particular descriptors such as "Progressive Psytrance" or "Melodic Black Metal", and therefore allows for a fine-grained representation of musical styles. The weightings correspond to the relative frequency of tags assigned to artists in Last.fm. For example, for the artist "Metallica" the top tags and their corresponding relative frequencies are "thrash metal" (1.0), "metal" (.91), "heavy metal" (.74), "hard rock" (.41), "rock" (.34) and "seen live" (.3). This means that the tag "thrash metal" is the most popular genre tag assigned to "Metallica" and thus, its weighting is 1.0. From this list, we remove all tags that are not part of the 1,998 Freebase genres (i.e., "seen live" in our example) as well as all tags with a relative frequency smaller than .5 (i.e., "hard rock" and "rock" in our example). Thus, for "Metallica", we end up with three genres, namely "thrash metal", "metal" and "heavy metal" that we assign to all listening events of the artist "Metallica". Overall, this process gives us, on average, 2–3 genres per artist (i.e., mean = 2.466). Furthermore, 96.25% of the genres are assigned to more than one artist.

**User groups based on mainstreaminess.** The *LFM-1b* dataset contains a mainstreaminess value for each user, which defines the distance from this user's music genre preferences to the music genre preferences of the (Last.fm) mainstream. To study different types of users, we split the dataset into three equally sized groups based on their mainstreaminess (i.e., low, medium, and high). We sort the users in the dataset based on their mainstreaminess value and assign the 1,000 users with the lowest values to the

LowMS group, the 1,000 users with the highest values to the HighMS group, and the 1,000 users with a value that lies around the average mainstreaminess (=.379) to the MedMS group.

Here, we consider only users with at least 6,000 and at most 12,000 listening events, a choice we made based on the average number of listening events per user in the dataset (i.e., 9,043) as well as the kernel density distribution of the data. With this method, on the one hand, we exclude users with too little data available for training our algorithms (i.e., users with <6,000 listening events), and on the other hand, we exclude so-called power listeners (i.e., users with >12,000 listening events) who might distort our results.

Furthermore, this high average number of listening events per user also means that we have enough listening events (i.e., between 6.9 to 8.2 million) to train and test the music genre preference modeling and prediction approaches, even if we only consider 1,000 users per group. **Table 1** summarizes the statistics and characteristics of these three groups.

*(i) LowMS.* The LowMS group represents the $|U|$ = 1,000 least mainstream users. They have an average mainstreaminess value of $\overline{MS}$ = .125. This group contains $|A|$ = 82,417 distinct artists, $|LE|$ = 6,915,352 listening events, $|G|$ = 931 genres and $|GA|$ = 14,573,028 genre assignments.

*(ii) MedMS.* The MedMS group represents the $|U|$ = 1,000 users whose mainstreaminess values are between the ones of LowMS and HighMS groups (i.e., their mainstreaminess values lie around the average). This group has an average mainstreaminess value of $\overline{MS}$ = .379. Most statistics of this group lie between those of the LowMS and HighMS users (for example, the number of genre assignments per listening event $|GA|/|LE|$ = 2.565), except for the average age, which is the highest for the MedMS users ($\overline{Age}$ = 25.352 years).

*(iii) HighMS.* This group represents the $|U|$ = 1,000 most mainstream users in the *LFM-1b* dataset ($\overline{MS}$ = .688). These users are not only the youngest ones ($\overline{Age}$ = 21.486 years) but also listen to the highest number of distinct genres on average ($\overline{G_u}$ = 186.010). Also, this user group exhibits the highest number of distinct genres ($|G|$ = 973).

**Average pairwise user similarity.** Finally, the boxplots in **Figure 1** show the average pairwise user similarity in the three user groups. We calculate these scores based on the genre distributions of the users and using the cosine similarity metric. We see that users in the LowMS group have a very individual listening behavior (mean user similarity = .118), while users in the HighMS group tend to listen to similar music genres (mean user similarity = .691). Again, the users in the MedMS group lie in between (mean user similarity = .392). Given these results, we expect a collaborative filtering approach based on user similarities to deliver good genre prediction results for the HighMS group.

**Popularity of music genre preferences.** In **Figure 2**, we compare the music genre popularity distributions of the LowMS, MedMS, and HighMS groups. To this end, we plot the number of listening events for the groups' top-30 genres. We find that there are some dominating genres with more than 2 million LE counts in the HighMS group, while the genre distribution is much more evenly distributed in the LowMS group with a LE count of around 500,000 for the most popular genres. We can describe the genre distribution of the MedMS group as an intermediate of the LowMS and HighMS distribution. We analyze the actual top-30 genres in these groups, and while the most popular genres Rock and Pop dominate the other genres
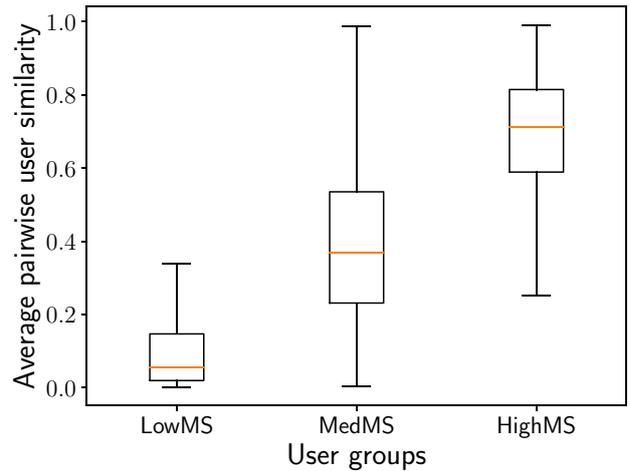


**Figure 1:** Boxplots show the average pairwise user similarity in our user groups using the cosine similarity metric computed on the users' genre distributions. While users in the LowMS group show a very individual listening behavior, users in the HighMS group tend to listen to similar music genres.

**Table 1:** Dataset statistics for the LowMS, MedMS, and HighMS Last.fm user groups. Here, $|U|$ is the number of distinct users, $|A|$ is the number of distinct artists, $|G|$ is the number of distinct genres, $|LE|$ is the number of listening events, $|GA|$ is the number of genre assignments, $|GA|/|LE|$ is the number of genre assignments per listening event, $\overline{G_u}$ is the average number of genres a user $u$ has listened to, $\overline{MS}$ is the average mainstreaminess value, and $\overline{Age}$ is the average age of users in the group.

| User Group | $|U|$ | $|A|$ | $|G|$ | $|LE|$ | $|GA|$ | $|GA|/|LE|$ | $\overline{G_u}$ | $\overline{MS}$ | $\overline{Age}$ |
|---|---|---|---|---|---|---|---|---|---|
| LowMS | 1,000 | 82,417 | 931 | 6,915,352 | 14,573,028 | 2.107 | 85.771 | .125 | 24.582 |
| MedMS | 1,000 | 86,249 | 933 | 7,900,726 | 20,264,870 | 2.565 | 126.439 | .379 | 25.352 |
| HighMS | 1,000 | 92,690 | 973 | 8,251,022 | 22,498,370 | 2.727 | 186.010 | .688 | 21.486 |

in the HighMS group (LE count of Rock = 2,269,861), in the LowMS group, it is not as dominant (LE count of Rock = 685,998). Furthermore, we find several genres that are not popular in the MedMS and HighMS groups but are popular in the LowMS group, such as Ambient and Black Metal.

Based on the dataset characteristics, we expect that a group-based modeling approach, which models a user's music genre preferences utilizing the most-frequently listened genres of all users in the group, performs fine for HighMS in relation to other modeling techniques, while for the LowMS group, a personalized modeling technique would be preferable. In the MedMS group, we expect both modeling approaches to work well due to the group being an intermediate of the HighMS and LowMS groups.

**Temporal drift of music genre preferences.** Next, we investigate the temporal drift of music genre preferences. The plots (a), (b), and (c) of **Figure 3** show the effect of time on the genre listening behavior of our LowMS, MedMS, and HighMS user groups. We plot the relistening
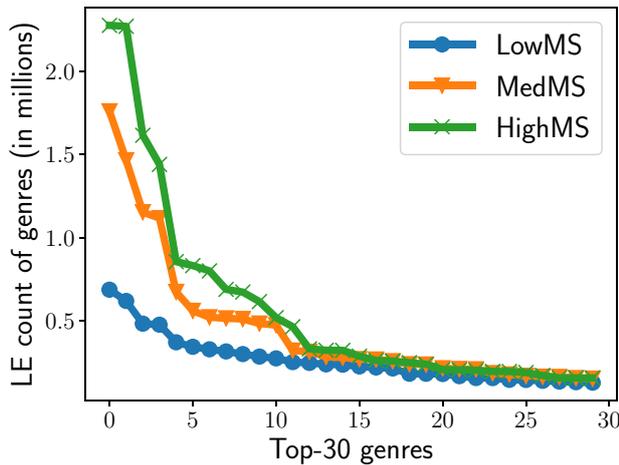


**Figure 2:** Number of listening events LE (in millions) for the top-30 genres of our LowMS, MedMS, and HighMS Last.fm user groups. We find that there are some dominating genres in the HighMS group, while the genre distribution in the LowMS group is more evenly distributed.

count of music genres over the time (in hours) since the last listening events of these genres on a log-log scale. For example, if a user $u$ has listened to artists with genre $g$ twice in a time interval of 1 hour, then the relistening count for "1 hour" is incremented by 1. We repeat this process for all listening events, which gives us a relistening count for each hour. We observe similar results for all three groups, which means that the shorter the time since the last listening event of a genre $g$, the higher its relistening count. In all three plots, we see a peak after 24 hours, which indicates that people tend to listen to similar music genres daily at the same time. However, we also see that when people have not listened to a genre for a longer period, i.e., one month (around 750 hours), the relistening count of this genre drastically drops.

Finally, we also plot the linear regression lines of the empirical data in the plots of **Figure 3**. In the log-log-scaled plots, we can observe a good fit of the data, which indicates that the data likely follows a power-law distribution (cf. Anderson and Schooler, 1991). This claim is supported by the high $R^2$ values of the fits, which are between .870 and .895. Concerning the slopes $\alpha$ of the lines, which describe how strongly temporal listening drifts influence the user groups, we observe values between $-1.480$ and $-1.587$. We can use these values as the $d$ parameter of the BLL equation (Anderson et al., 2004), cf. Equation 6.

Taken together, we observe interesting temporal effects in all three user groups: Last.fm users tend to listen to genres they have listened to recently. Moreover, we find that this temporal drift of music genre preferences follows a power-law distribution. Correspondingly, we can model this drift with the BLL equation.

### 3.2 Modeling and Prediction of Music Genre Preferences

In this section, we describe five baseline approaches (i.e., $TOP$, $CF_u$, $CF_i$, $POP_u$, and $TIME_u$) as well as our approach based on the BLL equation for modeling and predicting music genre preferences (i.e., $BLL_u$).

**Group-based baseline: $TOP$.** Motivated by our analysis in **Figure 2**, the $TOP$ approach models a user $u$'s music genre preferences using the overall top-$k$ (e.g., top-30) genres



(a) User group: LowMS
Linear regression: $R^2 = .870$, $\alpha = -1.480$

(b) User group: MedMS
Linear regression: $R^2 = .894$, $\alpha = -1.574$

(c) User group: HighMS
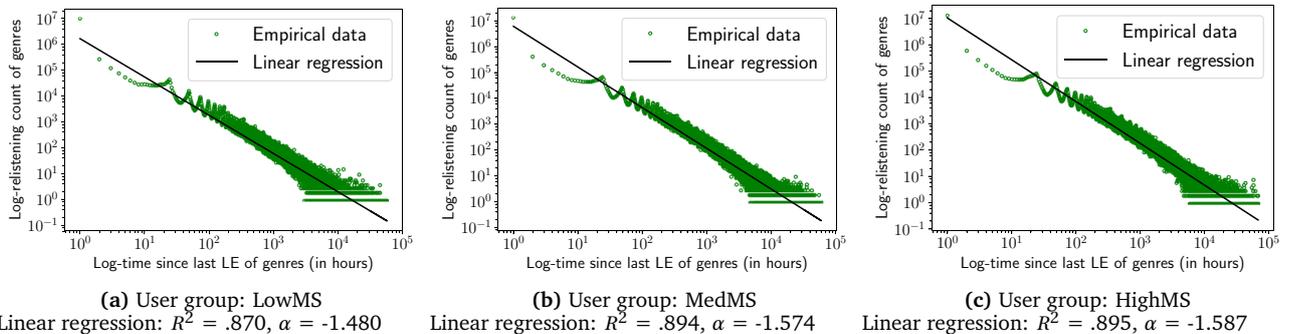Linear regression: $R^2 = .895$, $\alpha = -1.587$

**Figure 3:** The effect of time on genre relistening behavior for the LowMS, MedMS, and HighMS Last.fm user groups. For all three groups, we find that the shorter the time since the last listening event of a genre, the higher its relistening count. Additionally, we plot the linear fits of the data and report the corresponding $R^2$ estimates as well as the slopes $\alpha$. We can observe a very good fit of the data, which indicates that the data likely follows a power-law distribution.

of all users in the user group $UG_u$ (i.e., LowMS, MedMS, HighMS) to which $u$ belongs. This is given by:

$$\widetilde{G_u^k} = \operatorname*{argmax}_{g \in G}^{k}(|GA_{g,UG_u}|) \qquad (1)$$

where $argmax^k$ refers to the "arguments of the maxima" function for the top-$k$ genres with maximum values, $\widetilde{G_u^k}$ denotes the set of $k$ predicted genres for user $u$, and $|GA_{g,UGu}|$ corresponds to the number of times $g$ occurs in all genre assignments $GA$ of $UG_u$. Thus, we describe this approach as a group-based modeling technique since it reflects the preferences of the whole user group LowMS, MedMS or HighMS. As our analysis in **Figure 2** shows that the genre distribution in the HighMS group is the least evenly distributed one, we expect the *TOP* approach to provide good prediction accuracy results for the HighMS group while performing worse for the LowMS group in relation to other modeling techniques.

**User-based collaborative filtering baseline: $CF_u$.** User-based collaborative filtering-based approaches aim to find similar users for a target user $u$, i.e., the set of neighbors $N_u$. $N_u$ is calculated using the cosine similarity between $u$'s genre distribution and the genre distributions of all other users. Then, the top-20 users are defined as $N_u$. Finally, $CF_u$ predicts the genres these similar users in $N_u$ have listened to (Shi et al., 2014), which is formally given by:

$$\widetilde{G_u^k} = \operatorname*{argmax}_{g \in G}^{k}\left( \sum_{v \in N_u} sim(G_u, G_v) \cdot |GA_{g,v}| \right) \qquad (2)$$

where $sim(G_u, G_v)$ is the cosine similarity between the genre distributions of user $u$ and neighbor $v$, and $|GA_{g,v}|$ indicates how often $v$ has listened to genre $g$. Since $CF_u$ relies on user similarities, we expect it to provide good results for the HighMS group compared to other modeling approaches (see also **Figure 1**).

**Item-based collaborative filtering baseline: $CF_i$.** Similar to $CF_u$, $CF_i$ is a collaborative filtering-based approach, but instead of finding similar users for the target user $u$, it aims to find similar items (i.e., music artists). Then it predicts the genres that are assigned to these similar artists as given by:

$$\widetilde{G_u^k} = \operatorname*{argmax}_{g \in G}^{k}\left( \sum_{a \in A_u} \sum_{s \in S_a} sim(G_a, G_s) \cdot |GA_{g,s}| \right) \qquad (3)$$

Here, $A_u$ is the set of artists $u$ has listened to, $S_a$ is the set of similar artists for an artist $a$, $sim(G_a, G_s)$ is the cosine similarity between the genres assigned to $a$ and the genres assigned to a similar artist $s$, and $|GA_{g,v}|$ indicates how often genre $g$ was assigned to artist $a$ (hence, in our case either 0 or 1). Again, a neighborhood size $|S_{Au}| = 20$ leads to the best genre prediction results, and we also set $A_u$ to the set of the 20 artists that $u$ has listened to most frequently.

**Frequency-based baseline: $POP_u$.** The $POP_u$ approach is a personalized music genre preference modeling technique, which predicts the $k$ most frequently listened to (i.e., most popular) genres in the listening history of a user $u$. $POP_u$ corresponds to the modeling approach presented in (Schedl and Ferwerda, 2017) and is given by the following equation:

$$\widetilde{G_u^k} = \operatorname*{argmax}_{g \in G_u}^{k}(|GA_{g,u}|) \qquad (4)$$

where $G_u$ is the set of genres $u$ has listened to[6] and $|GA_{g,u}|$ denotes the number of times $u$ has listened to tracks with genre $g$ (i.e., the frequency). Thus, it ranks the genres $u$ has listened to in the past by popularity. Therefore, in relation to other modeling algorithms, we expect $POP_u$ to generate good genre predictions for all users in our three user groups, but especially for HighMS, in which the popularity feature is the most important one (see **Figure 2**).

**Recency-based baseline: $TIME_u$.** Our analysis presented in **Figure 3** motivates the personalized and recency-based music genre preference modeling, where we find that people tend to listen to genres to which they have listened just very recently. Thus, $TIME_u$ predicts the most recently listened to genres that are present in the listening history of a user $u$, which is given by:

$$\widetilde{G_u^k} = \operatorname*{argmin}_{g \in G_u}^{k}(t_{u,g,n}) \qquad (5)$$

where $t_{u,g,n}$ is the time since the last (i.e., the $n^{th}$) listening event of $g$ by $u$. Since we find that the temporal drift of music genre preferences is an important feature for all our three user groups, $TIME_u$ should provide good prediction accuracy results for LowMS, MedMS, and HighMS in relation to other modeling approaches.

**Our approach based on the BLL equation: $BLL_u$.** To combine the frequency-based modeling method $POP_u$ with the recency-based modeling method $TIME_u$, we utilize the BLL equation from the declarative memory module of the cognitive architecture ACT-R (Anderson et al., 2004). The BLL equation quantifies the importance of information in human memory (e.g., a word or a music genre) by considering how recently (i.e., temporal drift) and frequently (i.e., popularity) it was used in the past. In our setting, we define it as follows:

$$B_{u,g} = ln\left( \sum_{j=1}^{n} t_{u,g,j}^{-d} \right) \qquad (6)$$

Here, $g$ is a genre user $u$ has listened to in the past, and $n$ is the number of times $u$ has listened to $g$. Further, $t_{u,g,j}$ is the time since the $j^{th}$ listening event of $g$ by $u$, and $d$ is the power-law decay factor that accounts for the feature of the temporal drift of music genre preferences.

We set $d$ to the slopes $\alpha$ identified in the analysis of **Figure 3** (i.e., 1.480 for LowMS, 1.574 for MedMS, and

1.587 for HighMS). The resulting base-level activation values $B_{u,g}$ are normalized using a simple softmax function in order to map them onto a range of [0,1] where they sum to 1 (Kowald et al., 2017b):

$$B'_{u,g} = \frac{\exp(B_{u,g})}{\sum\limits_{g' \in G_u} \exp(B_{u,g'})} \quad (7)$$

Again, $G_u$ is the set of distinct genres listened to by $u$. Finally, $BLL_u$ predicts the top-$k$ genres $\widetilde{G_u^k}$ with the highest $B'_{u,g}$ values for $u$:

$$\widetilde{G_u^k} = \underset{g \in G_u}{\mathrm{argmax}}^k (B'_{u,g}) \quad (8)$$

**Comparison of approaches. Table 2** shows how the five baselines, as well as $BLL_u$, cover our four features of interest, i.e., (i) personalization, (ii) collaboration, (iii) popularity, and (iv) temporal drift.

Here, our $BLL_u$ approach is the only one that covers the features of personalization, popularity, and temporal drifts. Moreover, *TOP*, $CF_u$, and $CF_i$ are the only approaches that consider collaboration among users and, thus, investigate the listening events of all users. We further examine which feature combination works best for predicting genres in our setting in the next section of this paper.

## 4. Experiments and Results

In this section, we outline the experimental setup (see Section 4.1) and in Section 4.2, we present the results of our study on evaluating the usefulness for modeling music genre preferences using the BLL equation.

### 4.1 Experimental Setup

To measure the accuracy of our music genre preference modeling approaches, we conduct a study, in which we predict the genres assigned to the artists a user is going to listen to in the future.

**Evaluation protocol.** We split the datasets into train and test sets (Cremonesi et al., 2008) and make sure that our

evaluation protocol preserves the temporal order of the listening events, which simulates a real-world scenario in which we predict (genres of) future listening events based on past ones (Kowald et al., 2017b; Seitlinger et al., 2015). This also means that a classic $k$-fold cross-validation evaluation protocol with random splits is not useful.

Therefore, we put the most recent 1% of the listening events of each user into the test set and keep the remaining listening events for training. We do not use a classic 80/20 or 90/10 split as the number of listening events per user is large (i.e., on average 7,689 per user). Furthermore, although we only use the most recent 1% of listening events per user, this process leads to three large test sets with 69,153 listening events for LowMS, 79,007 listening events for MedMS, and 82,510 listening events for HighMS. On average, there are 76 listening events per user for which we predict the assigned genres.

In **Figure 4**, we present boxplots showing the average duration in days per user we have available in our three test sets. We see that the average duration per user is evenly distributed across all three user groups with a median value of 11.8 days, which is also around 1% of the median value of the overall average duration per user (i.e., the sum of training and test durations). This corresponds to the 1% of the listening events per user we use for the test sets. Thus, we are going to predict the genres a user is going to listen to in this period.

Following this evaluation protocol, our goal is to validate whether our BLL-based approach (i.e., $BLL_u$) provides better prediction accuracy results than the five baseline approaches (i.e., *TOP*, $CF_u$, $CF_i$, $POP_u$, and $TIME_u$). When investigating the numbers shown in **Table 1**, we also see that our prediction task is not trivial since $|GA|/|LE|$, i.e., the number of genre assignments per listening event (=what should be predicted), is much smaller than $\overline{G_u}$, i.e., the average number of genres a user $u$ has listened to (=what could be predicted).
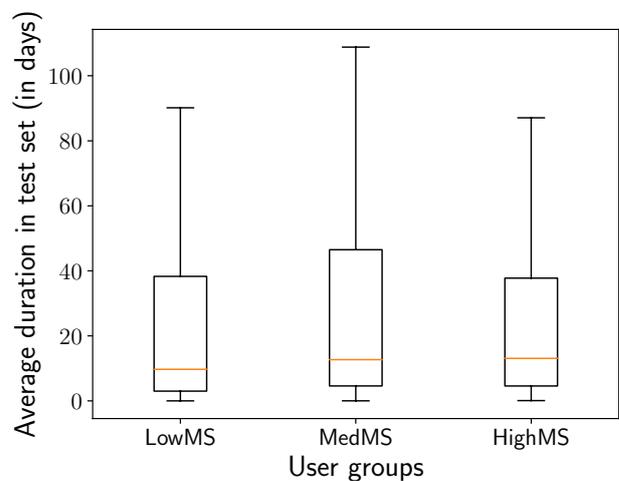


**Figure 4:** Boxplots showing the average duration in days per user we have available in our three test sets. Across all three users groups, the average duration per user is evenly distributed with a median value of 11.8 days.

**Table 2:** Comparison of our five baselines as well as our approach based on the BLL equation for modeling and predicting music genre preferences. In this table, a "✓" indicates that a specific approach covers a specific feature. While *TOP*, $CF_u$ and $CF_i$ also consider collaboration among users (i.e., investigate listening events of all users), our $BLL_u$ approach is the only one that is personalized and accounts for the features of popularity as well as temporal drifts.

| Feature | TOP | $CF_u$ | $CF_i$ | $POP_u$ | $TIME_u$ | $BLL_u$ |
|---|---|---|---|---|---|---|
| Personalization | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Collaboration | ✓ | ✓ | ✓ | | | |
| Popularity | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Temporal drifts | | | | | ✓ | ✓ |

**Evaluation metrics.** To measure the prediction quality of the approaches, we use the following six state-of-the-art metrics (Baeza-Yates and Ribeiro-Neto, 2011):

*(i) Recall: R@k.* Recall is calculated as the number of correctly predicted genres divided by the number of relevant genres (i.e., from the test set). It is a measure of the completeness of the predictions.

*(ii) Precision: P@k.* Precision is calculated as the number of correctly predicted genres divided by the number of predictions $k$ and is a measure of the accuracy of the predictions. We report recall and precision for $k = 1$ … 10 predicted genres in the form of recall/precision plots.

*(iii) F1-score: F1@5.* F1-score is the harmonic mean of recall and precision. If 10 genres are predicted, the F1-score typically reaches its highest value for $k = 5$. Thus, we report it for $k = 5$.

*(iv) Mean Reciprocal Rank: MRR@10.* MRR is the mean of reciprocal ranks of all relevant genres in the list of predicted genres.

*(v) Mean Average Precision: MAP@10.* MAP is the mean of the average precision scores at all ranks where relevant genres are predicted. With this, it also takes the ranking of the correctly predicted genres into account.

*(vi) Normalized Discounted Cumulative Gain: nDCG@10.* nDCG is another ranking-dependent metric. It is based on the Discounted Cumulative Gain (DCG) measure (Järvelin et al., 2008).

We report MRR, MAP, and nDCG for $k = 10$ predicted music genres, where these metrics reach their highest values.

**Evaluation framework.** For reasons of reproducibility, we conduct the prediction study using our recommendation benchmarking framework *TagRec* (Kowald et al., 2017a), which provides the evaluation protocol and metrics described in this section. Furthermore, we also implement the modeling approaches described in Section 3.2 using *TagRec*. It is freely available via our Github repository.[7]

### 4.2 Results and Discussion

In this section, we report and discuss our prediction accuracy results on evaluating the usefulness of our BLL-based music genre preference modeling approach (i.e., $BLL_u$) compared to five baseline approaches: (i) group-based modeling (i.e., *TOP*), (ii) user-based collaborative filtering ($CF_u$), (iii) item-based collaborative filtering ($CF_i$), (iv) frequency-based modeling (i.e., $POP_u$), and (v) recency-based modeling (i.e., $TIME_u$).

**Table 3** summarizes our evaluation results for the three user groups (i.e., LowMS, MedMS, and HighMS), the four evaluation metrics (i.e., $F1@5$, $MRR@10$, $MAP@10$, and $nDCG@10$) as well as the six approaches (i.e., *TOP*, $CF_u$, $CF_i$, $POP_u$, $TIME_u$, and $BLL_u$). Additionally, in **Figure 5**, we show the recall/precision plots of the approaches for $k = 1…10$ predicted genres (i.e., $R@k$ and $P@k$).

Based on the features introduced in **Table 2**, we discuss these results concerning the influence of (i) personalization, (ii) collaboration, (iii) popularity, and (iv) temporal drift. Furthermore, we compare the results of our $BLL_u$ approach for our user groups and different numbers of predicted genres in **Figure 6** as well as show the performance of the approaches in a cold-start setting in **Figure 7**. Finally, we also discuss the implications of our findings for personalized music recommendation.

**Influence of personalization.** The personalized approaches (i.e., $POP_u$, $CF_u$, $CF_i$, $TIME_u$, and $BLL_u$) outperform the group-based *TOP* approach in the LowMS setting. This is in line with our analysis presented in **Figure 2**, where we

**Table 3:** Genre prediction accuracy results of our study comparing our $BLL_u$ approach with a group-based baseline (*TOP*), a user-based collaborative filtering baseline ($CF_u$), an item-based collaborative filtering baseline ($CF_i$), a frequency-based baseline ($POP_u$) and a recency-based baseline ($TIME_u$). For all three user groups (i.e., LowMS, MedMS, and HighMS), the combination of popularity and temporal drift of music genre preferences in the form of $BLL_u$ provides the best results for all metrics. According to a t-test with $\alpha = .001$, "***" indicates statistically significant differences between $BLL_u$ and all other approaches for all user groups.

| User group | Evaluation metric | TOP | $CF_u$ | $CF_i$ | $POP_u$ | $TIME_u$ | $BLL_u$ |
|---|---|---|---|---|---|---|---|
| LowMS | F1@5 | .108 | .311 | .341 | .356 | .368 | **.397*** |
| | MRR@10 | .101 | .389 | .425 | .443 | .445 | **.492*** |
| | MAP@10 | .112 | .461 | .505 | .533 | .550 | **.601*** |
| | nDCG@10 | .180 | .541 | .590 | .618 | .625 | **.679*** |
| MedMS | F1@5 | .196 | .271 | .284 | .292 | .293 | **.338*** |
| | MRR@10 | .146 | .248 | .264 | .274 | .272 | **.320*** |
| | MAP@10 | .187 | .319 | .336 | .351 | .365 | **.419*** |
| | nDCG@10 | .277 | .419 | .441 | .460 | .452 | **.523*** |
| HighMS | F1@5 | .247 | .273 | .266 | .282 | .228 | **.304*** |
| | MRR@10 | .188 | .232 | .229 | .242 | .201 | **.266*** |
| | MAP@10 | .246 | .304 | .298 | .314 | .267 | **.348*** |
| | nDCG@10 | .354 | .413 | .402 | .429 | .357 | **.462*** |

**(a)** User group: LowMS        **(b)** User group: MedMS        **(c)** User group: HighMS
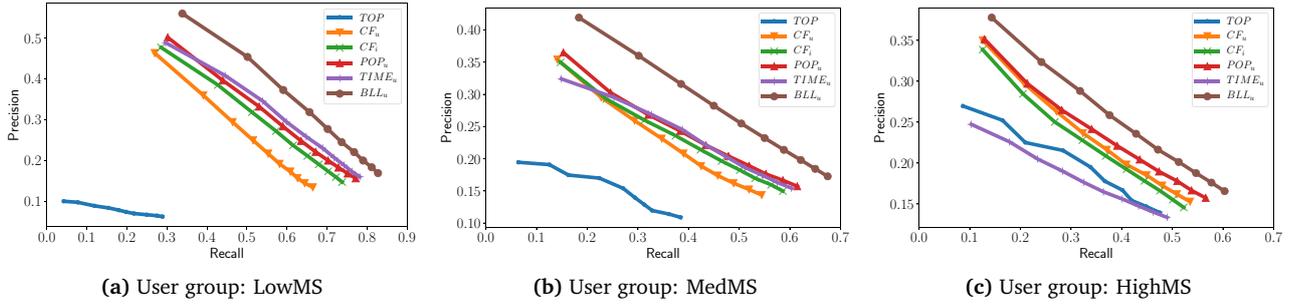
**Figure 5:** Recall/precision plots of the baselines and our $BLL_u$ approach for the three user groups LowMS, MedMS, and HighMS. We see that $BLL_u$ provides the best results for all groups and for all $k = 1…10$ predicted genres.
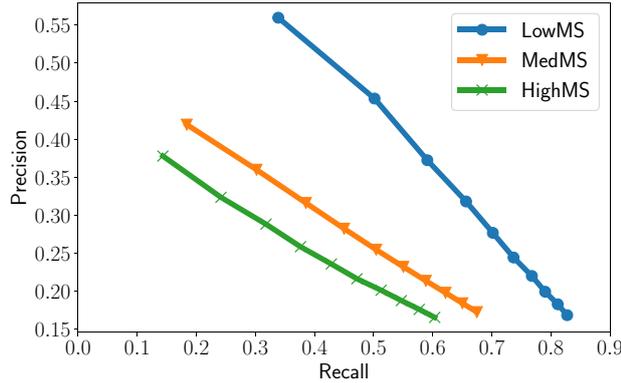


**Figure 6:** Recall/precision plot of our $BLL_u$ approach for $k = 1…10$ predicted genres for the three user groups LowMS, MedMS and HighMS. We see that $BLL_u$ provides good prediction accuracy results for all groups but especially in the LowMS setting. This shows that our approach is especially useful for predicting the music genre preferences of users with low mainstreaminess values.
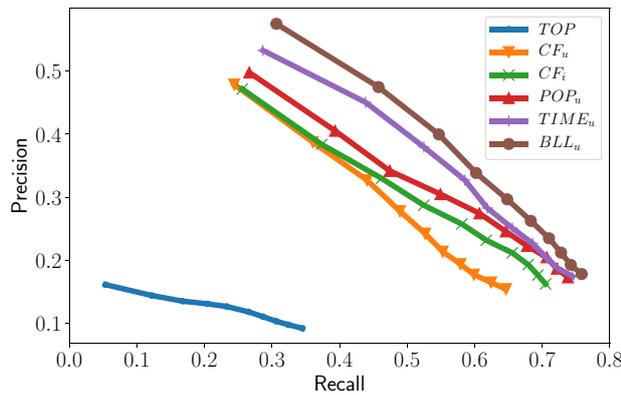


**Figure 7:** Recall/precision plot for our $BLL_u$ approach and our five baselines in a cold-start setting. We see that $BLL_u$ also provides the best results in cases where users only have a few listening events available for training.

found that the music genre popularity distribution in the LowMS group is the most evenly distributed one.

The same is true for the MedMS group, in which we observe a very similar performance of $CF_u$, $CF_i$, $POP_u$, and $TIME_u$. However, in the HighMS setting only the four personalized approaches, which utilize the popularity feature (i.e., $POP_u$, $CF_u$, $CF_i$, and $BLL_u$) outperform $TOP$. This shows that the influence of personalization on the

prediction accuracy becomes more important as the mainstreaminess of the users decreases (i.e., in the LowMS setting).

**Influence of collaboration.** We investigate the genre prediction accuracy of three approaches (i.e., $TOP$, $CF_u$, and $CF_i$) that consider collaboration among users, i.e., that analyze the listening events of all users. Here, the personalized $CF_u$ and $CF_i$ approaches provide better results than the non-personalized $TOP$ approach for all three user groups.

Furthermore, $CF_u$ provides its best results for the HighMS group. This is in line with our analysis presented in **Figure 1**, which shows that the average pairwise user similarity is the highest for high-mainstream users. This is also the reason why $CF_i$ does not outperform $CF_u$ in the HighMS but outperforms it in the LowMS and MedMS settings.

**Influence of popularity.** We evaluate four popularity-based approaches. The first approach provides non-personalized genre predictions based on the preferences of all users (i.e., $TOP$), and the second offers personalized predictions based on user similarities (i.e., $CF_u$). The third approach provides personalized predictions using item similarities (i.e., $CF_i$), and the fourth produces personalized genre predictions based on the preferences of the individual user (i.e., $POP_u$). While the prediction accuracy of $TOP$ increases with the level of mainstreaminess, the prediction accuracy of $POP_u$ decreases with the level of mainstreaminess. The prediction accuracy of $CF_u$ and $CF_i$ are relatively stable over all three user groups, with the only exception that $CF_u$ provides better results than $CF_i$ in the HighMS setting.

Thus, in the HighMS group, $TOP$ provides a higher prediction accuracy than in the other two groups. These results are in line with our analysis presented in **Figure 2**, where we find that there are some dominating genres in the HighMS group, which explains the good results of $TOP$, $CF_u$, and $POP_u$ in this setting. When further comparing $CF_u$ with $CF_i$, we see that $CF_i$ outperforms $CF_u$ in the LowMS and MedMS settings.

**Influence of temporal drift.** Our analysis in **Figure 3** reveals that users in Last.fm tend to listen to genres which they have listened to very recently. In other words, time is

important for all three user groups. However, as shown in **Table 3** and **Figure 5**, $TIME_u$ provides the weakest accuracy results for HighMS and good prediction accuracy results for LowMS and MedMS. Thus, for HighMS, popularity is a more important feature than recency.

$BLL_u$ outperforms $TIME_u$ in all experiments. This means that our personalized modeling approach, which also considers the features of popularity and temporal drifts, can provide accurate genre predictions for all three groups in relation to other modeling techniques.

**Accuracy of $BLL_u$ for different values of $k$.** In **Figure 6**, we show the recall/precision results of $BLL_u$ for $k = 1...10$ predicted genres for the three user groups. We observe apparent differences in the accuracy value ranges when comparing the three groups. While $BLL_u$ outperforms the five baselines in all three settings (with significant differences between $BLL_u$ and all other approaches according to a t-test with $\alpha = .001$), the accuracy estimates are much higher in the LowMS group (i.e., $R@10 = .827$ and $P@1 = .559$) than in the MedMS group (i.e., $R@10 = .674$ and $P@1 = .419$) and the HighMS group (i.e., $R@10 = .603$ and $P@1 = .377$). This shows that our approach is especially useful to predict the genre preferences of users with low inclination to listen to mainstream music.

**Performance in cold-start setting.** Since recommender systems are often faced with situations in which users only have a few interactions available to train the underlying recommendation algorithms, we also evaluate our $BLL_u$ approach in a cold-start setting (Schein et al., 2002). For this, we extract the 1,000 users with the lowest number of LEs from the LFM-1b dataset. As we need to make sure that we have at least 1 LE per user available for training the algorithms, this procedure leads to 1,000 users with a minimum of 2 LEs and a maximum of 46 LEs per user. For these users, we have precisely 1 LE in the test set, for which we predict the assigned genres.

Our results for this experiment are shown in the recall/precision plot of **Figure 7**. Here, we observe very similar results to the ones of our LowMS, MedMS, and HighMS settings (see **Figure 6**). Thus, again $BLL_u$ provides the best accuracy results followed by $TIME_u$, $POP$, $CF_i$, and $CF_u$. As expected, the non-personalized $TOP$ approach provides the worst results in this setting. These results show that $BLL_u$ is also capable of effectively predicting music genre preferences in cold-start settings where users only have a few listening events available for training.

**Implications for personalized music recommendation.** In this section, so far, we have shown that $BLL_u$ outperforms the baseline approaches concerning prediction accuracy in different settings (i.e., LowMS, MedMS, HighMS, and cold-start). When looking at **Figure 6**, this is especially true for the LowMS group, in which users do not follow the preferences of the mainstream, and thus, a personalization technique, as given by the BLL equation, is critical. If we relate this to music recommender systems, which exploit the listening histories of users to suggest

other music that they might also like, our findings lead to interesting implications. Schedl and Hauger (2015) have shown that standard recommendation algorithms such as collaborative filtering cannot provide suitable music recommendations for users with low mainstreaminess. The results presented in this section support this. In other words, such users need different music recommendation algorithms that account for their highly individual listening preferences.

One way to achieve this could be to combine state-of-the-art music recommendation algorithms (see Section 2) with our music genre preference modeling approach based on the BLL equation presented in this paper. We could use the calculated $B'_{u,g}$ values given by our approach as an input for these algorithms or to rerank recommendation results based on the importance of a genre for a user. We elaborate on these ideas as well as other plans for future work in Section 5.

## 5. Conclusion and Future Work

In this paper, we presented $BLL_u$, an approach that utilizes the features of popularity and temporal drifts to model and predict music genre preferences via fine-grained genres. We leveraged the LFM-1b dataset of more than one billion music listening events, created by approximately 120,000 users of the online music service Last.fm. We divided the users into three groups based on the proximity of their music genre preferences to the mainstream: (i) LowMS, i.e., listeners of niche music, (ii) HighMS, i.e., listeners of mainstream music, and (iii) MedMS, i.e., listeners of music that lies in-between. To take into account the popularity and temporal drift of music genre preferences, we proposed to use the Base-Level Learning (BLL) equation from the cognitive architecture ACT-R, which quantifies the importance of information in human memory (e.g., a music genre) by considering how frequently (i.e., popularity) and recently (i.e., temporal drift) it was used in the past. A comparison between $BLL_u$ and a group-based baseline (i.e., $TOP$), a user-based collaborative filtering baseline (i.e., $CF_u$), an item-based collaborative filtering baseline (i.e., $CF_i$), a frequency-based baseline (i.e., $POP_u$) as well as a recency-based baseline (i.e., $TIME_u$) showed that $BLL_u$ outperforms all other approaches for all three user groups in terms of prediction accuracy.

Furthermore, our results indicate that $BLL_u$ is especially useful to predict the music genre preferences of users with interest in low-mainstream music (i.e., the LowMS user group), which opens up interesting possibilities for future work in the research area of personalized music recommender systems.

**Limitations and future work.** So far, we limited our approach to the BLL equation of the declarative memory module of ACT-R. Since the BLL equation is only a part of the more exhaustive ACT-R framework that does not consider contextual information, one needs to consider this limitation when utilizing our approach. For example, when we model music genre preferences exclusively via past listening behavior, phenomena such as over-personalization or filter-bubble effects could occur

(Nguyen et al., 2014). To overcome this, we plan to extend our model to the full activation equation of ACT-R, which also considers contextual information via its associative activation (Anderson et al., 2004). Moreover, we plan to extend our model by other components of ACT-R, for example, to investigate further context dimensions such as the mood or the current activity of the user (see, e.g., Ferwerda et al. (2015)). We could achieve this by defining and implementing so-called production rules from ACT-R's procedural memory module as, for instance, done in the SNIF-ACT model (Pirolli and Fu, 2003; Fu and Pirolli, 2007). Another limitation of our work is that we employed a rather simple definition for the mainstreaminess of a user. We, therefore, plan to extend our analysis to include more sophisticated mainstreaminess measures, e.g., based on rank-order correlation or Kullback-Leibler divergence (Schedl and Bauer, 2018). As part of future work, we plan to integrate our findings into music recommendation algorithms, with particular attention to addressing the low mainstreaminess group, since standard collaborative filtering approaches tend to fail to provide suitable music recommendations for this user group (Schedl and Hauger, 2015). For example, we plan to integrate the preference values we obtain for a specific user and a particular genre via our approach as a context dimension into a matrix factorization-based approach (Mnih and Salakhutdinov, 2008; Koenigstein et al., 2011) or a deep learning-based approach (Lin et al., 2018; Sachdeva et al., 2018).

Furthermore, we aim to apply our approach to the problem of music playlist continuation, which was also the task of the ACM RecSys Challenge 2018.[8] We believe that our findings concerning the temporal relistening patterns of music genres (see Section 3.1) could help identify genres that users commonly listened to consecutively. We could then, for example, incorporate such genre sequences into the two-stage convolutional neural network (CNN) model for automatic playlist continuation that was proposed by Volkovs et al. (2018). Finally, we would like to highlight that our approach could be easily leveraged by researchers and practitioners also for other related tasks (e.g., recommending music artists) and not only for genre prediction. Thus, we hope that future work in the areas of user modeling and music recommendation will be attracted by our insights.

## Reproducibility
To foster the reproducibility of our research, we use the publicly available LFM-1b Last.fm dataset (see Section 3.1). Furthermore, we provide our evaluation framework *TagRec* (see Section 4.1) freely for academic purposes. We hope that the approach presented in this paper and its implementation in *TagRec*, as well as the dataset, will attract further research on music preference modeling and recommender systems.

## Notes
[1] https://www.last.fm/.
[2] https://www.pandora.com/.
[3] https://www.spotify.com/.
[4] http://www.cp.jku.at/datasets/LFM-1b/.
[5] https://developers.google.com/freebase/ (no longer maintained).
[6] Here, we could also use $G$ instead of $G_u$, which would lead to the same results, but to reduce the computational effort, we only need to consider the genres that the target user $u$ has listened to in the past.
[7] https://github.com/learning-layers/TagRec.
[8] http://www.recsyschallenge.com/2018/.

## Competing Interests
The authors have no competing interests to declare.

## Author Contributions
Elisabeth Lex and Dominik Kowald contributed equally to this work.

## References
**Abdollahpouri, H., Mansoury, M., Burke, R.,** & **Mobasher, B.** (2019). The unfairness of popularity bias in recommendation. *arXiv preprint arXiv:1907.13286*.

**Aizenberg, N., Koren, Y.,** & **Somekh, O.** (2012). Build your own music recommender by modeling internet radio streams. In *Proceedings of the International World Wide Web Conference*, pages 1–10. ACM. DOI: https://doi.org/10.1145/2187836.2187838

**Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C.,** & **Qin, Y.** (2004). An integrated theory of the mind. *Psychological Review, 111*(4). DOI: https://doi.org/10.1037/0033-295X.111.4.1036

**Anderson, J. R.,** & **Schooler, L. J.** (1991). Reflections of the environment in memory. *Psychological Science, 2*(6), 396–408. DOI: https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

**Arnett, J.** (1992). The soundtrack of recklessness: Musical preferences and reckless behavior among adolescents. *Journal of Adolescent Research, 7*(3), 313–331. DOI: https://doi.org/10.1177/074355489273003

**Baeza-Yates, R.,** & **Ribeiro-Neto, B.** (2011). *Modern Information Retrieval*. ACM Press. DOI: https://doi.org/10.1145/2009916.2010172

**Bauer, C.,** & **Schedl, M.** (2019). Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems. *PLoS ONE, 14*(6), 1–36. DOI: https://doi.org/10.1371/journal.pone.0217389

**Cantor, J. R.,** & **Zillmann, D.** (1973). The effect of affective state and emotional arousal on music appreciation. *The Journal of General Psychology, 89*(1), 97–108. DOI: https://doi.org/10.1080/00221309.1973.9710822

**Cattell, R. B.,** & **Anderson, J. C.** (1953). The measurement of personality and behavior disorders by the IPAT music preference test. *Journal of Applied Psychology*, *37*(6), 446. DOI: https://doi.org/10.1037/h0056224

**Celma, O.** (2010). *Music Recommendation and Discovery – The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer. DOI: https://doi.org/10.1007/978-3-642-13287-2

**Celma, Ò.,** & **Cano, P.** (2008). From hits to niches?: Or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*. ACM. DOI: https://doi.org/10.1145/1722149.1722154

**Cremonesi, P., Turrin, R., Lentini, E.,** & **Matteucci, M.** (2008). An evaluation methodology for collaborative recommender systems. In *Proceedings of International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution*, pages 224–231. IEEE Computer Society. DOI: https://doi.org/10.1109/AXMEDIS.2008.13

**Delsing, M. J., Ter Bogt, T. F., Engels, R. C.,** & **Meeus, W. H.** (2008). Adolescents' music preferences and personality characteristics. *European Journal of Personality: Published for the European Association of Personality Psychology*, *22*(2), 109–130. DOI: https://doi.org/10.1002/per.665

**Dollinger, S. J.** (1993). Research note: Personality and music preference: Extraversion and excitement seeking or openness to experience? *Psychology of Music*, *21*(1), 73–77. DOI: https://doi.org/10.1177/030573569302100105

**Dunn, P. G., de Ruyter, B.,** & **Bouwhuis, D. G.** (2012). Toward a better understanding of the relation between music preference, listening behavior, and personality. *Psychology of Music*, *40*(4), 411–428. DOI: https://doi.org/10.1177/0305735610388897

**Ferwerda, B., Yang, E., Schedl, M.,** & **Tkalcic, M.** (2015). Personality traits predict music taxonomy preferences. In *Proceedings of ACM CHI Conference on Human Factors in Computing Systems*, pages 2241–2246. ACM. DOI: https://doi.org/10.1145/2702613.2732754

**Fu, W.-T.,** & **Pirolli, P.** (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human-Computer Interaction*, *22*(4), 355–412. DOI: https://doi.org/10.21236/ADA462156

**George, D., Stickle, K., Rachid, F.,** & **Wopnford, A.** (2007). The association between types of music enjoyed and cognitive, behavioral, and personality factors of those who listen. *Psychomusicology: A Journal of Research in Music Cognition*, *19*(2). DOI: https://doi.org/10.1037/h0094035

**Greenberg, D. M., Baron-Cohen, S., Stillwell, D. J., Kosinski, M.,** & **Rentfrow, P. J.** (2015). Musical preferences are linked to cognitive styles. *PLoS ONE*, *10*(7), 1–22. DOI: https://doi.org/10.1371/journal.pone.0131151

**Hargreaves, D. J., North, A. C.,** & **Tarrant, M.** (2015). How and why do musical preferences change in childhood and adolescence. *The Child as Musician: A Handbook of Musical Development*, pages 303–322. DOI: https://doi.org/10.1093/acprof:oso/9780198744443.003.0016

**Järvelin, K., Price, S. L., Delcambre, L. M.,** & **Nielsen, M. L.** (2008). Discounted cumulated gain based evaluation of multiple-query IR sessions. In *Proceedings of the European Conference on Information Retrieval*, pages 4–15. Springer. DOI: https://doi.org/10.1007/978-3-540-78646-7_4

**Juslin, P. N.,** & **Sloboda, J. A.** (2001). *Music and Emotion: Theory and Research*. Oxford University Press.

**Kim, N., Chae, W.-Y.,** & **Lee, Y.-J.** (2018). Music recommendation with temporal dynamics in multiple types of user feedback. In *Proceedings of the 7th International Conference on Emerging Databases*, pages 319–328. Springer. DOI: https://doi.org/10.1007/978-981-10-6520-0_35

**Koenigstein, N., Dror, G.,** & **Koren, Y.** (2011). Yahoo! music recommendations: Modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of ACM Conference on Recommender Systems*, pages 165–172. ACM. DOI: https://doi.org/10.1145/2043932.2043964

**Kowald, D., Kopeinik, S.,** & **Lex, E.** (2017a). The TagRec framework as a toolkit for the development of tag-based recommender systems. In *Adjunct Publication of the ACM Conference on User Modeling, Adapation and Personalization*, pages 23–28. ACM. DOI: https://doi.org/10.1145/3099023.3099069

**Kowald, D.,** & **Lex, E.** (2016). The influence of frequency, recency and semantic context on the reuse of tags in social tagging systems. In *Proceedings of ACM Conference on Hypertext and Social Media*, pages 237–242. ACM. DOI: https://doi.org/10.1145/2914586.2914617

**Kowald, D., Pujari, S. C.,** & **Lex, E.** (2017b). Temporal effects on hashtag reuse in twitter: A cognitiveinspired hashtag recommendation approach. In *Proceedings of the International World Wide Web Conference*, pages 1401–1410. ACM. DOI: https://doi.org/10.1145/3038912.3052605

**Krause, A. E.,** & **North, A. C.** (2018). 'Tis the season: Music-playlist preferences for the seasons. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(1). DOI: https://doi.org/10.1037/aca0000104

**Leadbeater, R.** (2014). *Magpies and mirrors: identity as a mediator of music preferences across the lifespan*. PhD thesis, Lancaster University.

**Lin, Q., Niu, Y., Zhu, Y., Lu, H., Mushonga, K. Z.,** & **Niu, Z.** (2018). Heterogeneous knowledge-based attentive neural networks for short-term music recommendations. *IEEE Access*, *6*. DOI: https://doi.org/10.1109/ACCESS.2018.2874959

**Maanen, L. V.,** & **Marewski, J. N.** (2009). Recommender systems for literature selection: A competition between decision making and memory models. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

**Mnih, A.,** & **Salakhutdinov, R. R.** (2008). Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, pages 1257–1264.

**Moore, J. L., Chen, S., Turnbull, D.,** & **Joachims, T.** (2013). Taste over time: The temporal dynamics of user preferences. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 401–406.

**Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L.,** & **Konstan, J. A.** (2014). Exploring the filter bubble: The effect of using recommender systems on content diversity. In *Proceedings of the International World Wide Web Conference*, pages 677–686. ACM. DOI: https://doi.org/10.1145/2566486.2568012

**North, A.,** & **Hargreaves, D.** (2008). *The Social and Applied Psychology of Music*. OUP Oxford. DOI: https://doi.org/10.1093/acprof:oso/9780198567424.001.0001

**North, A. C.,** & **Hargreaves, D. J.** (1999). Music and adolescent identity. *Music Education Research*, *1*(1), 75–92. DOI: https://doi.org/10.1080/1461380990010107

**Park, C. H.,** & **Kahng, M.** (2010). Temporal dynamics in music listening behavior: A case study of online music service. In *Proceedings of the IEEE/ACIS International Conference on Computer and Information Science*, pages 573–578. IEEE. DOI: https://doi.org/10.1109/ICIS.2010.142

**Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L.,** & **Brattico, E.** (2011). Music and emotions in the brain: Familiarity matters. *PLoS ONE*, *6*(11). DOI: https://doi.org/10.1371/journal.pone.0027241

**Peretz, I., Gaudreau, D.,** & **Bonnel, A.-M.** (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26*(5), 884–902. DOI: https://doi.org/10.3758/BF03201171

**Pirolli, P.,** & **Fu, W.-T.** (2003). SNIF-ACT: A model of information foraging on the World Wide Web. In *International Conference on User Modeling*, pages 45–54. Springer. DOI: https://doi.org/10.1007/3-540-44963-9_8

**Rentfrow, P. J.,** & **Gosling, S. D.** (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6). DOI: https://doi.org/10.1037/0022-3514.84.6.1236

**Rodà, A., Canazza, S.,** & **Poli, G. D.** (2014). Clustering affective qualities of classical music: Beyond the valence-arousal plane. *IEEE Transactions on Affective Computing*, *5*(4), 364–376. DOI: https://doi.org/10.1109/TAFFC.2014.2343222

**Sachdeva, N., Gupta, K.,** & **Pudi, V.** (2018). Attentive neural architecture incorporating song features for music recommendation. In *Proceedings of the ACM Conference on Recommender Systems*, pages 417–421. ACM. DOI: https://doi.org/10.1145/3240323.3240397

**Schäfer, T.,** & **Sedlmeier, P.** (2010). What makes us like music? Determinants of music preference. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4). DOI: https://doi.org/10.1037/a0018374

**Schedl, M.** (2016). The lfm-1b dataset for music retrieval and recommendation. In *Proceedings of the Conference on Multimedia Retrieval*, pages 103–110. ACM. DOI: https://doi.org/10.1145/2911996.2912004

**Schedl, M.,** & **Bauer, C.** (2018). An analysis of global and regional mainstreaminess for personalized music recommender systems. *Journal of Mobile Multimedia*, *14*, 95–112.

**Schedl, M.,** & **Ferwerda, B.** (2017). Large-scale analysis of group-specific music genre taste from collaborative tags. In *Proceedings of the IEEE International Symposium on Multimedia*, pages 479–482. IEEE. DOI: https://doi.org/10.1109/ISM.2017.95

**Schedl, M., Gómez, E., Trent, E., Tkalčič, M., Eghbal-Zadeh, H.,** & **Martorell, A.** (2018a). On the interrelation between listener characteristics and the perception of emotions in classical orchestra music. *IEEE Transactions on Affective Computing*, *9*, 507–525. DOI: https://doi.org/10.1109/TAFFC.2017.2663421

**Schedl, M.,** & **Hauger, D.** (2015). Tailoring music recommendations to users by considering diversity, mainstreaminess, and novelty. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 947–950. ACM. DOI: https://doi.org/10.1145/2766462.2767763

**Schedl, M., Knees, P., McFee, B., Bogdanov, D.,** & **Kaminskas, M.** (2015). Music recommender systems. In *Recommender Systems Handbook*, pages 453–492. Springer. DOI: https://doi.org/10.1007/978-1-4899-7637-6_13

**Schedl, M., Zamani, H., Chen, C.-W., Deldjoo, Y.,** & **Elahi, M.** (2018b). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, *7*(2), 95–116. DOI: https://doi.org/10.1007/s13735-018-0154-2

**Schein, A. I., Popescul, A., Ungar, L. H.,** & **Pennock, D. M.** (2002). Methods and metrics for coldstart recommendations. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 253–260. ACM. DOI: https://doi.org/10.1145/564376.564421

**Schubert, E.** (2007). The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music*, *35*(3), 499–515. DOI: https://doi.org/10.1177/0305735607072657

**Seitlinger, P., Kowald, D., Kopeinik, S., Hasani-Mavriqi, I., Lex, E.,** & **Ley, T.** (2015). Attention please! A hybrid resource recommender mimicking attention-interpretation dynamics. In *Companion Proceedings of International World Wide Web Conference*, pages 339–345. ACM. DOI: https://doi.org/10.1145/2740908.2743057

**Selvi, C.,** & **Sivasankar, E.** (2019). An efficient context-aware music recommendation based on emotion and time context. In Mishra, D. K., Yang, X.-S., & Unal, A., Editors, *Data Science and Big Data Analytics*, pages 215–228. Springer. DOI: https://doi.org/10.1007/978-981-10-7641-1_18

**Shi, Y., Larson, M.,** & **Hanjalic, A.** (2014). Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys*, *47*(1), 3:1–3:45. DOI: https://doi.org/10.1145/2556270

**Vall, A., Quadrana, M., Schedl, M.,** & **Widmer, G.** (2019). Order, context and popularity bias in next-song recommendations. *International Journal of Multimedia Information Retrieval, 8*(2), 101–113. DOI: https://doi.org/10.1007/s13735-019-00169-8

**van den Oord, A., Dieleman, S.,** & **Schrauwen, B.** (2013). Deep content-based music recommendation. In *Proceedings of Neural Information Processing Systems Conference*, pages 2643–2651. Curran Associates Inc.

**Volkovs, M., Rai, H., Cheng, Z., Wu, G., Lu, Y.,** & **Sanner, S.** (2018). Two-stage model for automatic playlist continuation at scale. In *Proceedings of ACM Conference on Recommender Systems*, page 9. ACM. DOI: https://doi.org/10.1145/3267471.3267480

**Zheng, E., Kondo, G. Y., Zilora, S.,** & **Yu, Q.** (2018). Tag-aware dynamic music recommendation. *Expert Systems with Applications, 106*, 244–251. DOI: https://doi.org/10.1016/j.eswa.2018.04.014