

Transparent Music Preference Modeling and Recommendation with a Model of Human Memory Theory

Dominik Kowald^{1,2(✉)}, Markus Reiter-Haas², Simone Kopeinik¹, Markus Schedl³, and Elisabeth Lex²

¹ Know-Center GmbH, Graz, Austria
dkowald,skopeinik@know-center.at

² Graz University of Technology, Graz, Austria
reiter-haas,elisabeth.lex@tugraz.at

³ Johannes Kepler University Linz and Linz Institute of Technology, Austria
markus.schedl@jku.at

Abstract. In this chapter, we discuss how to utilize human memory models for the task of modeling music preferences for recommender systems. Therefore, we discuss the theoretical underpinnings of using cognitive models for user modeling and recommender systems in order to introduce a model based on the cognitive architecture ACT-R to predict the music genre preferences of users in the Last.fm platform. By implementing the declarative memory module of ACT-R, comprising past usage frequency and recency, as well as the current semantic context, we model the music relistening behavior of users. We evaluate our approach using three user groups that we identify in Last.fm, namely (i) low-mainstream music listeners, (ii) medium-mainstream music listeners, and (iii) high-mainstream music listeners. We find that our approach provides significantly higher prediction accuracy than various baseline algorithms for all three user groups, and especially for the low-mainstream user group. Since our approach is based on a well-established human memory model, we also discuss how this contributes to the transparency of the calculated predictions.

Keywords: User Modeling, Human Memory Theory, ACT-R, Music Recommendation, Recommender Systems, Transparency

1 Introduction

Computational models of user preferences are important elements of music recommender systems [52] to tailor recommendations to the preferences of the user. Such user models are typically derived from the listening behavior of the users, i.e., their interactions with music artifacts, content features of music [64], or hybrid combinations of both. Research in music psychology [38] has shown that a wide range of factors impact music preferences [52], such as users' emotional

state [9, 15], a user’s current context [44], or a user’s personality [44, 50]. Several aspects make the modeling of music preferences challenging, such as, e.g., that music consumption is context-dependent [48] and serves various purposes for listeners [53]. Also, related research [20, 24, 26] has verified that traditional music recommendation approaches suffer from popularity bias, i.e., they are biased to the mainstream that is prevalent in a music community. As a result, listeners of non-mainstream music receive less relevant recommendations compared to listeners of popular, mainstream music [6, 40, 46, 47].

In our own previous research [23], we introduced a psychology-inspired approach to model and predict the music genre preferences of users. We based our approach on findings from music psychology that show that music liking is positively influenced by prior exposure to the music [41, 54]. This has been attributed to the *mere exposure effect* or *familiarity principle* [63], i.e., users tend to establish positive preferences for items to which they are frequently and consistently exposed. Our idea was to computationally model prior exposure to music genres using the activation equation of human memory from the cognitive architecture *Adaptive Control of Thought–Rational* (ACT-R) [2, 4]. The activation equation determines the usefulness of a memory unit (i.e., its *activation*) for a user in the current context, based on how frequently and recently a user accessed it in the past as well as how important this unit is in the current context. In particular, we utilized the activation equation of ACT-R for music genre predictions. The equation enabled us to tune the predictions to the current context of the user. As the current context, we utilized the set of genres that are assigned to the most recently listened artist of a user.

On a publicly available dataset of Last.fm music listening histories, we modeled the genre preferences of users from three different groups, which we extracted using behavioral data in the form of music listening events: (i) LowMS, i.e., listeners of niche music (low mainstreamness), (ii) MedMS, i.e., the middle tier of listeners (medium mainstreamness), and (iii) HighMS, i.e., listeners of mainstream music (high mainstreamness). We introduced the $ACT_{u,a}$ approach that employs the activation equation to take into account the current context of the user, which we defined as the user’s current genre preference. We compared the efficacy of $ACT_{u,a}$ to a variant, i.e., BLL_u , that uses only a part of the activation equation (the base-level learning (BLL) component) to model the past usage frequency (i.e., popularity) and recency (i.e., time). Furthermore, we compared both approaches to five baselines, including two collaborative filtering variants, mainstream-aware genre modeling, popularity-aware genre modeling, and time-based genre modeling. Here, we found that both BLL_u and $ACT_{u,a}$ outperform the five baseline methods in all three groups, with $ACT_{u,a}$ achieving the significantly highest performance. Our results also showed that with both BLL_u and $ACT_{u,a}$, we could specifically improve the prediction performance for the users in the LowMS group, i.e., the music consumers whose prediction quality typically suffers the most from popularity bias.

In this chapter, we extend our previous work [23] with the goal of discussing how to utilize human memory models for a transparent modeling process of

music preferences for recommender systems. Therefore, in Section 2, we provide a general description of the ACT-R cognitive architecture. This section provides the theoretical underpinnings for using human memory theory to model music preferences of users as presented in [23], and in Sections 3 and 4. Furthermore, in Section 5, we discuss potential extensions of our prediction model by reviewing additional components of ACT-R. Finally, in Section 6, we conclude this chapter and discuss possibilities for future research.

2 Theoretical Underpinnings

Cognitive Science evolved as a research field that combines knowledge of different disciplines, namely Psychology, Philosophy, Linguistics, Anthropology, Neuroscience and Computer Science, in a multi- or interdisciplinary manner. Among the core hypotheses of this field is the belief that processes and states of the human mind can be emulated via computer models [39]. On the basis of this fundamental assumption, cognitive modeling describes the development of executable computer models that approximate cognitive processes, mechanisms, and representations [57]. Cognitive models are divided into three main categories: (i) computational, (ii) mathematical, and (iii) verbal-conceptual models [13].

In principle, the three types of models differ with respect to their detail of formalization: Computational models are algorithmic descriptions that use processes to emulate tasks of human cognition. Mathematical models are considered a subset of computational models. They consist of mathematical equations that formalize relationships between entities that interact in human cognition tasks. Despite their typical lack of process details, they can mostly be implemented as computer models. Verbal-conceptual models, on the other hand, describe such cognitive processes, entities and their relationships in a relatively natural language [57]. In this work, we focus on computational models, and more precisely, the cognitive architecture ACT-R, which stands short for “Adaptive Control of Thought – Rational” and has been previously suggested to model human cognition in Human-Computer Interaction (HCI) tasks [7].

ACT-R is a cognitive architecture developed by John Robert Anderson [2]. ACT-R defines and formalizes the basic cognitive operations of the human mind (e.g., access to information in human memory). It is grounded on the assumption that all components in a human mind act in concert to generate behavior. To that end, the ACT-R theory describes how different parts of humans’ minds work together and, based on this, proposes an architecture built from a number of collaborating modules. The so-called production system coordinates the information flow between different modules in the center of the model. Each module is provided with a buffer containing its most important information to reduce the working load on the production system. This is the data the production system is aware of and reacts to. Figure 1 schematically illustrates the main components of ACT-R that explain a memory perspective. In general, ACT-R differs between short-term memory modules, such as the working memory module, and long-term memory modules, such as the declarative and procedural

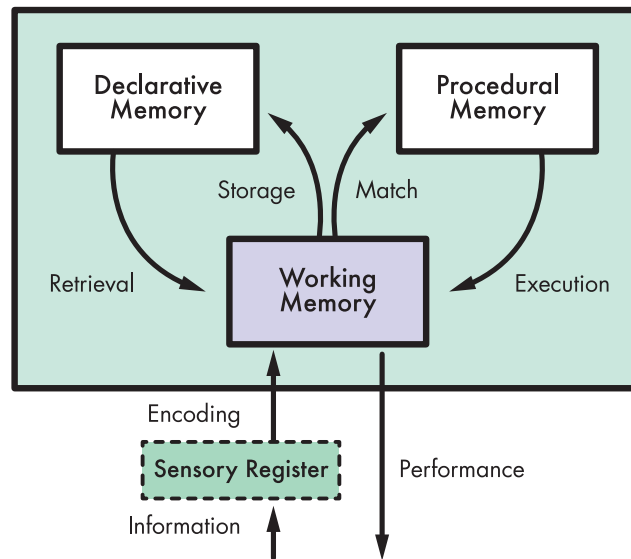


Fig. 1. Schematic illustration of ACT-R. In our work, we focus on the activation equation of the declarative memory module.

memory modules. Using a sensory register (i.e., the ultra-short-term memory), the encoded information is passed to the short-term working memory module, which interacts with the long-term memory modules. In the declarative memory, the encoded information can be stored, and already stored information can be retrieved. In procedural memory, the information can be matched against stored rules, which may lead to actions [62]. Thus, declarative memory holds factual knowledge (e.g., what something is), and procedural memory consists of sequences of actions (e.g., how to do something).

In this work, we present a transparent music genre modeling and prediction approach based on the declarative memory module of ACT-R. In particular, we built on the activation equation that formulates the availability of elements in a person’s declarative memory, which is described as part of the ACT-R theory [2]. The activation equation is commonly used to model memory recall tasks [37], and has been proposed in the context of tag recommendations [21, 27], item recommendations [29, 43], and hybrid recommendations [31, 36], as well as for social semantic technologies [18]. A thorough theoretical survey and derivation of the activation equation is presented in [4].

3 Data and Approach

In this section, we describe the Last.fm dataset as well as our transparent music genre modeling and prediction approaches. This section is mainly based on our previous work [23].

3.1 Dataset

For our experiments, we use the publicly available *LFM-1b* dataset⁴ of music listening information shared by users of the online music platform Last.fm. *LFM-1b* contains listening histories of more than 120,000 users, which sums up to over 1.1 billion listening events (LEs) collected between January 2005 and August 2014. Each LE contains a user identifier, the artist, the album, the track name, and a timestamp [45]. Furthermore, the *LFM-1b* dataset contains demographic data of the users such as country, age, gender, and a mainstreamness score, which is defined as the overlap between a user’s personal listening history and the aggregated listening history of all Last.fm users in the dataset. This overlap is measured using a symmetric variant of Kullback-Leibler divergence. Thus, the mainstreamness score reflects a user’s inclination to music listened to by the Last.fm mainstream listeners (i.e., the “average” Last.fm listener) [51].

User groups. In order to study different types of users, we use this mainstreamness score to split the *LFM-1b* dataset into three equally sized user groups based on their mainstreamness (i.e., low, medium, and high). Specifically, we sort all users based on their mainstreamness score and assign the 1,000 users with the lowest scores to the low-mainstream group (i.e., *LowMS*), the 1,000 users with scores around the median mainstreamness (= .379) to the medium-mainstream group (i.e., *MedMS*), and the 1,000 users with the highest scores to the high-mainstream group (i.e., *HighMS*).

We consider only users with a minimum of 6,000 and a maximum of 12,000 LEs. We choose these thresholds based on the average number of LEs per user in the dataset, which is 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 the proposed models (i.e., users with less than 6,000 LEs). On the other hand, we exclude so-called power listeners (i.e., users with more than 12,000 LEs) that might distort our results. Table 1 summarizes the statistics and characteristics of our three user groups. We see that, even if we only consider 1,000 users per group, we have a sufficient amount of LEs, i.e., between 6.9 to 8.3 million, to train and test our music genre modeling and prediction approaches. Further characteristics of our user groups are as follows:

(i) **LowMS.** The LowMS group represents the $|U| = 1,000$ users with the smallest mainstreamness scores. These users have an average mainstreamness value of $Avg.MS = .125$. LowMS 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 consists of the $|U| = 1,000$ users with mainstreamness scores around the median and, thus, lying between the ones of the LowMS and HighMS groups. This group has an average mainstreamness value of $Avg.MS = .379$. The majority of dataset statistics of this group lies between the ones of the LowMS and HighMS users.

⁴ <http://www.cp.jku.at/datasets/LFM-1b/>

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 average number of genre assignments per LE, $|G|/|U|$ is the average number of genres a user has listened to, and *Avg.MS* is the average mainstreamness value.

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

(iii) **HighMS.** The HighMS group represents the $|U| = 1,000$ users in the *LFM-1b* dataset with the highest mainstreamness scores (*Avg.MS* = .688). These users listen to the highest number of distinct genres on average (i.e., $|G|/|U| = 186.010$), indicating that music which is considered mainstream is quite diverse on Last.fm. Also, this user group exhibits the highest number of distinct genres ($|G| = 973$).

Additionally, we investigate the most frequent countries of the users. Here, for all three groups, the United States (US) is the dominating country. The share of US users increases with the mainstreamness, i.e., while this share is only 14% for LowMS and 18% for MedMS, it is already 22% for HighMS. Interestingly, Russia (RU, 13%), Poland (PL, 9%), and Japan (JP, 8%) are frequent in the LowMS group, while the United Kingdom (UK) contributes a substantial share in the other two groups (9% for MedMS and 14% for HighMS). Germany (DE) is among the most frequently occurring countries in all three groups (10% for LowMS and HighMS, 8% for MedMS); Brazil (BR) can only be found among the most popular countries in the MedMS group (8%); and the Netherlands (NL, 5%) as well as Spain (ES, 4%) can only be found in the HighMS group.

Genre mapping. For mapping music genres to artists, we use an extension of the *LFM-1b* dataset, namely the *LFM-1b UGP* dataset [49], which describes the genres of an artist by leveraging social tags assigned by Last.fm users. Specifically, *LFM-1b UGP* contains a weighted mapping of 1,998 music genres available in the online database Freebase⁵ to Last.fm artists. This database includes a fine-grained representation of musical styles, including genres such as “Progressive Psytrance” or “Pagan Black Metal”.

The genre weightings for any given artist correspond to the relative frequency of tags assigned to that artist 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). 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,

⁵ <https://developers.google.com/freebase/> (no longer maintained)

for “Metallica”, we end up with three genres, i.e., “thrash metal”, “metal” and “heavy metal”.

3.2 Approach

Our approach focuses on the declarative part of the cognitive architecture ACT-R [2], which contains the activation equation of human memory. The activation equation determines the usefulness, i.e., the activation level A_i , of a memory unit i (e.g., a music genre in our case) for a user u in the current context. It is given by:

$$A_i = B_i + \sum_j W_j \cdot S_{j,i} \quad (1)$$

Here, the B_i component represents the *base-level* activation and quantifies the general usefulness of the unit i by considering how frequently and recently it has been used in the past. It is given by the base-level learning (BLL) equation:

$$B_i = \ln \left(\sum_{j=1}^n t_j^{-d} \right) \quad (2)$$

where n is the frequency of i ’s occurrences and t_j is the time since the j^{th} occurrence of i . The exponent d accounts for the power-law of forgetting, which means that each unit’s activation level caused by the j^{th} occurrence decreases in time according to a power function [2].

The second component (right addend) of Equation 1 represents the *associative activation* that tunes the base-level activation of the unit i to the current context. The context is given by any contextual element j that is relevant for the current situation. In the case of a music recommender system, that could be a music genre that the user prefers in the current situation. Through learned associations, the contextual elements are connected with i and can increase i ’s activation depending on the weight W_j and the strength of association $S_{j,i}$.

Modeling and Predicting Music Genre Preferences For a transparent modeling and prediction approach, we investigate two algorithms: BLL_u based on the BLL equation to model the past usage frequency (i.e., popularity) and recency (i.e., time), and $ACT_{u,a}$ based on the full activation equation to also take the current context into account.

We start with BLL_u and thus, with defining the base-level activation $B(g, u)$ for genre g and user u by utilizing the previously defined BLL equation:

$$B(g, u) = \ln \left(\sum_{j=1}^n t_{u,g,j}^{-d} \right) \quad (3)$$

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 in seconds since the j^{th} LE

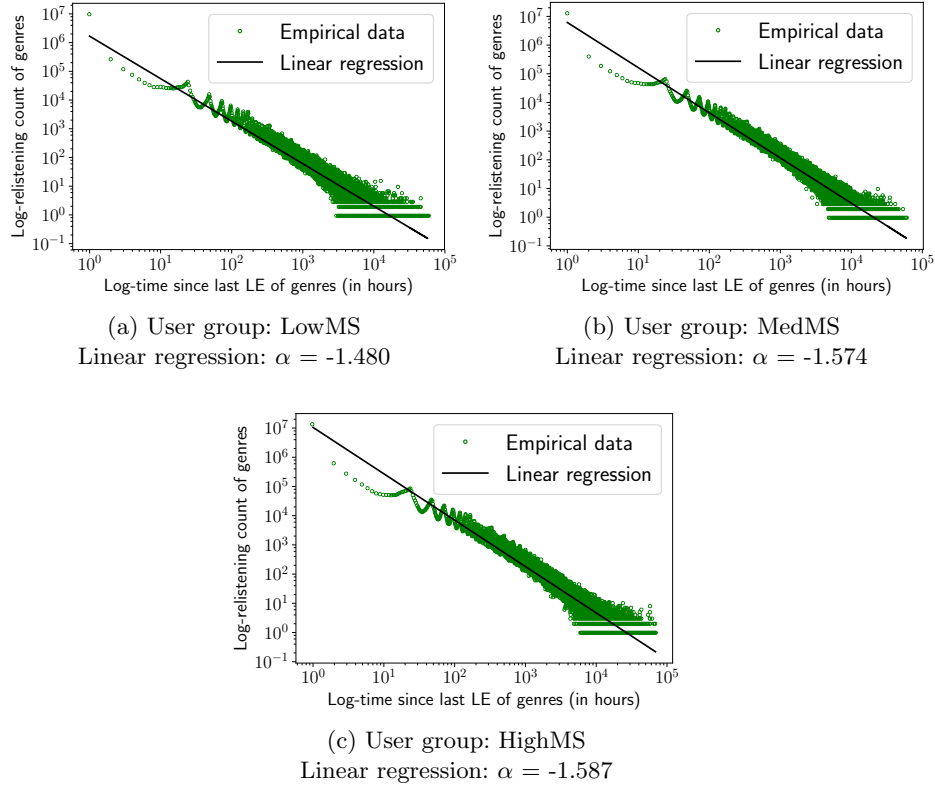


Fig. 2. Calculation of the BLL equation’s d parameter. On a log-log scale, we plot the relistening count of the genres over the time since their last LEs. We set d to the slopes α of the corresponding linear regression lines.

of g by u , and d is the power-law decay factor, which we identify using a similar method as used in [25]. Thus, in Figure 2, for all LEs and genres in our dataset, we plot the relistening count of a genre g over the time since the last LE of g . Then, we set d to the slope α of the linear regression lines of this data, which leads to 1.480 for LowMS, 1.574 for MedMS, and 1.587 for HighMS.

The resulting base-level activation values $B(g, u)$ are then normalized using a simple softmax function in order to map them onto a range of $[0, 1]$ that sums up to 1 [22, 25]:

$$B'(g, u) = \frac{\exp(B(g, u))}{\sum_{g' \in G_u} \exp(B(g', u))} \quad (4)$$

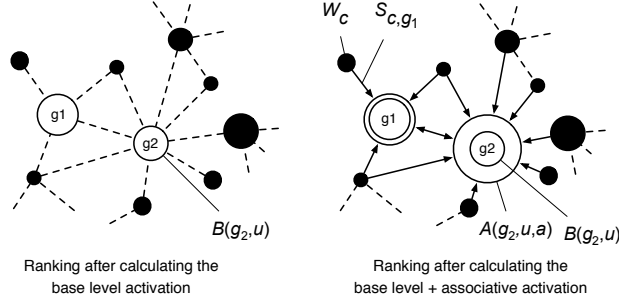


Fig. 3. Example illustrating the difference between BLL_u (left panel) and $ACT_{u,a}$ (right panel) based on [59]. Here, unfilled nodes represent target genres g_1 and g_2 , and black nodes represent genres of the last artist listened to by the target user (i.e., contextual genres). For g_1 and g_2 , the node sizes represent the activation levels and for the contextual genres, the node sizes represent the attentional weights W_c . The association strength $S_{c,g}$ is represented by the edge lengths. While BLL_u determines a higher activation level for g_1 than for g_2 , $ACT_{u,a}$ gives a higher activation level to g_2 than to g_1 by also considering the associative association based on the current context.

Here, 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 $B'(g, u)$ values to u :

$$\underbrace{\widetilde{G}_u^k = \arg \max_{g \in G_u}^k (B'(u, g))}_{BLL_u} \quad (5)$$

To investigate not only the factors of frequency and time but also the current context by means of an associative activation, we implement the full activation equation (see Equation 1) in the form of:

$$A(g, u, a) = B'(g, u) + \sum_{c \in G_a} W_c \cdot S_{c,g} \quad (6)$$

where the first part represents the base-level activation by means of the BLL equation and the second part represents the associative activation.

To calculate the associative activation and, thus, to model a user's current context, we incorporate the set of genres G_a assigned to the most recently listened to artist a by the user u . When applying this equation in the context of recommender systems, related literature [60] suggests using a measure of normalized co-occurrence to represent the strength of an association $S_{c,g}$. Accordingly, we define the co-occurrence between two genres as the number of artists to which both genres are assigned. We normalize this co-occurrence value according to the Jaccard coefficient:

$$S_{c,g} = \frac{|A_c \cap A_g|}{|A_c \cup A_g|} \quad (7)$$

where A_c is the set of artists to which any genre c of the current context (i.e., any genre of the artist most recently listened to) is assigned, and A_g is the set of artists to which genre g is assigned. Thus, we set the number of times two genres co-occur into relation with the number of times in which at least one of the two genres appears. In this work, we set the attentional weight W_c of context-genre c to 1. By doing so, we give equal weights to all genres assigned to an artist, which avoids down-ranking of less popular, but perhaps more specific, and hence more valuable, genres.

Finally, we normalize the $A(g, u, a)$ values using the aforementioned softmax function and predict the top- k genres \widetilde{G}_u^k with the highest $A'(g, u, a)$ values for a given user u and the genres of the user’s most recently listened artist a (i.e., the current context):

$$\widetilde{G}_u^k = \underbrace{\arg \max_{g \in G_u}^k (A'(u, g, a))}_{ACT_{u,a}} \quad (8)$$

We further illustrate the difference between BLL_u and $ACT_{u,a}$ in the example of Figure 3 [59] by showing the additional impact of the associative activation defined by the second component of the activation equation. As defined, this associative activation is evoked by the current context (i.e., the genres of the last artist the target user has listened to).

The left panel of Figure 3 shows two genres, g_1 and g_2 , with different base-level activation levels (illustrated by the circle size). Thus, according to BLL_u , g_1 reaches a higher base-level activation, which means a better rank, than g_2 . This relationship changes in the right panel of Figure 3, where we consider the influence of the genres in the current context (illustrated by the black nodes). Specifically, depending on the weights W_c (represented by the size of the black nodes) and strength of association $S_{c,g}$ (represented by the length and direction of the edges), the genres in the current context spread additional associative activation to the genres g_1 and g_2 . Now, according to $ACT_{u,a}$, g_2 receives stronger associative activation than g_1 , which also leads to a better rank.

4 Experiments and Results

In this section, we describe our experimental setup, i.e., the baseline algorithms, the evaluation protocol and metrics, as well as the results of our experiments. This section is mainly based on our previous work [23].

4.1 Baseline Algorithms

We compare the BLL_u and $ACT_{u,a}$ approaches to five baseline algorithms:

Mainstream-based baseline: *TOP*. The *TOP* approach models a user u ’s music genre preferences using the overall top- k genres of all users (i.e., the mainstream) in u ’s user group (i.e., LowMS, MedMS, HighMS). This is given by:

$$\widetilde{G}_u^k = \arg \max_{g \in G}^k (|GA_g|) \quad (9)$$

Here \widetilde{G}_u^k denotes the set of k predicted genres, G the set of all genres, and $|GA_g|$ corresponds to the number of times g occurs in all genre assignments GA of u 's user group.

User-based collaborative filtering baseline: CF_u . User-based collaborative filtering-based approaches aim to find similar users for the target user u (i.e., the set of neighbors N_u) and predict the genres these similar users have listened to in the past [56]. CF_u is given by:

$$\widetilde{G}_u^k = \arg \max_{g \in G(N_u)}^k \left(\sum_{v \in N_u} sim(G_u, G_v) \cdot |GA_{g,v}| \right) \quad (10)$$

where \widetilde{G}_u^k denotes the set of k predicted genres for user u , $G(N_u)$ are the genres listened to by the set of neighbors N_u ,⁶ $sim(G_u, G_v)$ is the cosine similarity between the genre distributions of user u and neighbor v . Finally, $|GA_{g,v}|$ indicates how often v has listened to genre g in the past. This approach is similar to the category recommender algorithm introduced in [28].

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 S_{A_u} , for the artists A_u that user u has listened to in the past. Then, it predicts the genres that are assigned to these similar artists as given by:

$$\widetilde{G}_u^k = \arg \max_{g \in G(S_{A_u})}^k \left(\sum_{a \in A_u} \sum_{s \in S_a} sim(G_a, G_s) \right) \quad (11)$$

where $G(S_{A_u})$ are the genres assigned to the similar artists S_{A_u} , S_a is the set of similar artists for an artist $a \in A_u$,⁷ and $sim(G_a, G_s)$ is the cosine similarity between the genre distributions assigned to a and the genres assigned to a similar artist $s \in S_a$.

Popularity-based baseline: POP_u . POP_u is a personalized music genre modeling technique, which predicts the k most frequently listened genres in the listening history of user u . POP_u is given by the following equation:

$$\widetilde{G}_u^k = \arg \max_{g \in G_u}^k (|GA_{g,u}|) \quad (12)$$

Here, G_u is the set of genres u has listened to in the past and $|GA_{g,u}|$ denotes the number of times u has listened to g . Thus, it ranks the genres u has listened to in the past by popularity.

Time-based baseline: $TIME_u$. The time-based baseline $TIME_u$ predicts the k genres that user u has most recently listened to. It is given by:

$$\widetilde{G}_u^k = \arg \min_{g \in G_u}^k (t_{u,g,n}) \quad (13)$$

⁶ We set the neighborhood size for CF_u and CF_i to 20.

⁷ For A_u , we consider the set of the 20 artists that u has listened to most frequently.

where $t_{u,g,n}$ is the time since the last (i.e., the n^{th}) LE of g by u .

4.2 Evaluation Protocol and Metrics

We split the datasets into train and test sets [10]. While doing so, we ensure that our evaluation protocol preserves the temporal order of the LEs, which simulates a real-world scenario in which we predict genres of future LEs based on past ones and not the other way round [25]. This also means that a classic k -fold cross-validation evaluation protocol is not useful in our setting.

Specifically, we put the most recent 1% of the LEs of each user into the test set (i.e., LE_{test}) and keep the remaining LEs for the train set (i.e., LE_{train}). We do not use a classic 80/20 split as the number of LEs per user is large (i.e., on average, 7,689 LEs per user). 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. To finally measure the prediction quality of the approaches, we use the following six well-established performance metrics [5]:

Recall: $R@k$. Recall is calculated as the number of correctly predicted genres divided by the number of relevant genres taken from the LEs in the test set LE_{test} . It is a measure for the completeness of the predictions and is formally given by:

$$R@k = \frac{1}{|LE_{test}|} \sum_{u,a \in LE_{test}} \frac{|\widetilde{G}_u^k \cap G_{u,a}|}{|G_{u,a}|} \quad (14)$$

where \widetilde{G}_u^k denotes the k predicted genres and $G_{u,a}$ the set of relevant genres of an artist a in user u 's LEs in the test set.

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. It is given by:

$$P@k = \frac{1}{|LE_{test}|} \sum_{u,a \in LE_{test}} \frac{|\widetilde{G}_u^k \cap G_{u,a}|}{k} \quad (15)$$

We report recall and precision for $k = 1 \dots 10$ predicted genres in form of recall/precision plots.

F1-score: $F1@k$. F1-score is the harmonic mean of recall and precision:

$$F1@k = 2 \cdot \frac{P@k \cdot R@k}{P@k + R@k} \quad (16)$$

We report the F1-score for $k = 5$, where it typically reaches its highest value if 10 genres are predicted.

Mean Reciprocal Rank: $MRR@k$. MRR is the average of reciprocal ranks $r(g)$ of all relevant genres in the list of predicted genres:

$$MRR@k = \frac{1}{|LE_{test}|} \sum_{u,a \in LE_{test}} \frac{1}{|G_{u,a}|} \sum_{g \in G_{u,a}} \frac{1}{r(g)} \quad (17)$$

This means that a high MRR is achieved if relevant genres occur at the beginning of the predicted genre list.

Mean Average Precision: $MAP@k$. MAP is an extension of the precision metric by also taking the ranking of the correctly predicted genres into account and is given by:

$$MAP@k = \frac{1}{|LE_{test}|} \sum_{u,a \in LE_{test}} \frac{1}{|G_{u,a}|} \sum_{i=1}^k Rel_i \cdot P@i \quad (18)$$

Here, Rel_i is 1 if the predicted genre at position i is among the relevant genres (0 otherwise) and $P@i$ is the precision calculated at position i according to Equation 15.

Normalized Discounted Cumulative Gain: $nDCG@k$. nDCG is another ranking-dependent metric. It is based on the Discounted Cumulative Gain ($DCG@k$) measure [14], which is defined as:

$$DCG@k = \sum_{i=1}^k \left(\frac{2^{Rel_i} - 1}{\log_2(1 + i)} \right) \quad (19)$$

where Rel_i is 1 if the genre predicted for the i^{th} item is relevant (0 otherwise). $nDCG@k$ is given as $DCG@k$ divided by $iDCG@k$, which is the highest possible DCG value that can be achieved if all relevant genres are predicted in the correct order:

$$nDCG@k = \frac{1}{|LE_{test}|} \sum_{u,a \in LE_{test}} \left(\frac{DCG@k}{iDCG@k} \right) \quad (20)$$

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

4.3 Results and Discussion

In this section, we present and discuss our evaluation results. The accuracy results according to $F1@5$, $MRR@10$, $MAP@10$, and $nDCG@10$ are shown in Table 2 for the five baseline approaches as well as the proposed BLL_u and $ACT_{u,a}$ algorithms. Furthermore, we provide recall/precision plots for $k = 1 \dots 10$ predicted genres in Figure 4.

Accuracy of baseline approaches. When analyzing the performance of the baseline approaches TOP , CF_u , CF_i , POP_u , and $TIME_u$, we see a clear difference between the non-personalized and the personalized algorithms. While

Table 2. Genre prediction accuracy results comparing our BLL_u and $ACT_{u,a}$ approaches with a mainstream-based baseline (TOP), a user-based collaborative filtering baseline (CF_u), an item-based collaborative filtering baseline (CF_i), a popularity-based baseline (POP_u) and a time-based baseline ($TIME_u$). For all three user groups (i.e., LowMS, MedMS, and HighMS), $ACT_{u,a}$ outperforms all other approaches. According to a t-test with $\alpha = .001$, “***” indicates statistically significant differences between $ACT_{u,a}$ and all other approaches.

| User group | Evaluation metric | TOP | CF_u | CF_i | POP_u | $TIME_u$ | BLL_u | $ACT_{u,a}$ |
|------------|-------------------|-------|--------|--------|---------|----------|---------|----------------|
| LowMS | $F1@5$ | .108 | .311 | .341 | .356 | .368 | .397 | .485*** |
| | $MRR@10$ | .101 | .389 | .425 | .443 | .445 | .492 | .626*** |
| | $MAP@10$ | .112 | .461 | .505 | .533 | .550 | .601 | .785*** |
| | $nDCG@10$ | .180 | .541 | .590 | .618 | .625 | .679 | .824*** |
| MedMS | $F1@5$ | .196 | .271 | .284 | .292 | .293 | .338 | .502*** |
| | $MRR@10$ | .146 | .248 | .264 | .274 | .272 | .320 | .511*** |
| | $MAP@10$ | .187 | .319 | .336 | .351 | .365 | .419 | .705*** |
| | $nDCG@10$ | .277 | .419 | .441 | .460 | .452 | .523 | .753*** |
| HighMS | $F1@5$ | .247 | .273 | .266 | .282 | .228 | .304 | .427*** |
| | $MRR@10$ | .188 | .232 | .229 | .242 | .201 | .266 | .412*** |
| | $MAP@10$ | .246 | .304 | .298 | .314 | .267 | .348 | .569*** |
| | $nDCG@10$ | .354 | .413 | .402 | .429 | .357 | .462 | .642*** |

the non-personalized TOP approach, which predicts the top- k genres of the mainstream, provides better accuracy results in the HighMS setting than in the LowMS setting, the personalized CF_u , CF_i , POP_u , and $TIME_u$ algorithms provide better results in the LowMS setting than in the HighMS setting. Hence, personalized genre modeling approaches provide better results, the lower the mainstreamness of the users. Non-personalized genre modeling approaches, however, have higher performance, the higher the mainstreamness of the users.

Next, we compare the accuracy of the two collaborative filtering-based methods, CF_u and CF_i . Here, the item-based CF variant CF_i reaches higher accuracy estimates in the LowMS and MedMS settings, while the user-based CF variant CF_u provides better performance in the HighMS setting. To better understand this pattern of results, we provide the average pairwise user similarity in the form of boxplots in Figure 5. Here, for all three user groups, we calculate the pairwise similarity between the users via the cosine similarity metric based on the users’ genre distribution vectors. We see that users in the HighMS setting are very similar to each other, which explains the good performance of an algorithm that is based on user similarities, such as CF_u .

POP_u and $TIME_u$ reach the highest accuracy estimates among the five baseline approaches. Interestingly, the popularity-based POP_u algorithm provides the best results for the HighMS user group, while the time-based $TIME_u$ algorithm provides the best results for the LowMS user group. For the MedMS user group, however, both algorithms reach a comparable accuracy performance, which shows the importance of both factors, frequency (i.e., popularity) and recency (i.e., time).

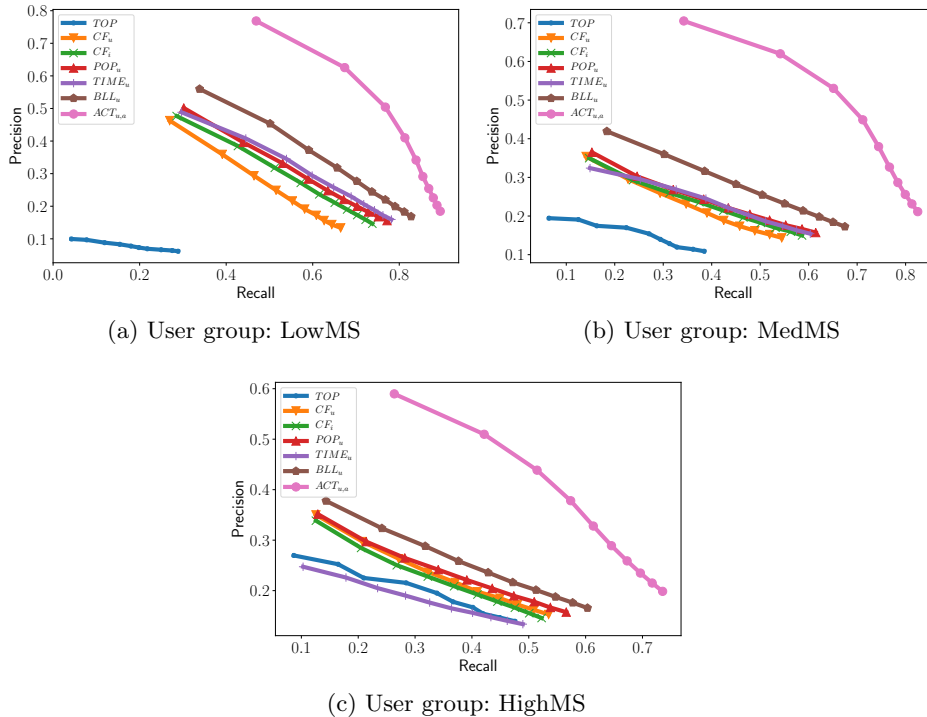


Fig. 4. Recall/precision plots for $k = 1 \dots 10$ predicted genres of the baselines and our BLL_u and $ACT_{u,a}$ approaches for the three user groups LowMS, MedMS, and HighMS. $ACT_{u,a}$ achieves the best results in all settings.

Accuracy of BLL_u and $ACT_{u,a}$. We discuss the results of the BLL_u and $ACT_{u,a}$ approaches, which utilize human memory processes as defined by the cognitive architecture ACT-R in order to model and predict music genre preferences. Specifically, BLL_u combines the factors of past usage frequency and recency via the BLL equation (see Equation 3) and $ACT_{u,a}$ extends BLL_u by also considering the current context via the activation equation (see Equation 6). In this work, we define the current context by the genres assigned to the artist that the target user u has listened to most recently.

As expected, when combining the factors of past usage frequency and recency in the form of BLL_u , we can outperform the best performing baseline approaches POP_u and $TIME_u$ in all three settings (i.e., LowMS, MedMS, and HighMS). We can further improve the accuracy performance when we additionally consider the current context in the form of $ACT_{u,a}$. Here, we reach a statistically significant improvement⁸ over all other approaches across all evaluation metrics and user groups. Furthermore, in Figure 6, we present a recall/precision plot showing the

⁸ According to a t-test with $\alpha = .001$.

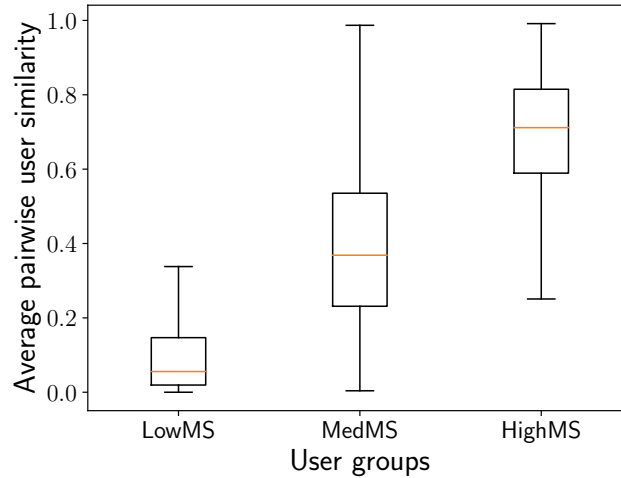


Fig. 5. Average pairwise user similarity for LowMS, MedMS, and HighMS. We calculate the user similarity using the cosine similarity metric based on the users’ genre distributions. While users in the LowMS group show a very individual listening behavior, users in the HighMS group tend to listen to similar music genres.

accuracy of $ACT_{u,a}$ for $k = 1 \dots 10$ predicted genres for LowMS, MedMS, and HighMS. We observe good results for all three user groups, but especially in the LowMS setting, in which we are faced with users with a low interest in mainstream music.

This shows that the proposed $ACT_{u,a}$ algorithm can provide accurate predictions of music genres listened to in the future for all user groups. Moreover, since our approach utilizes human memory processes, it is based on psychological principles of human intelligence rather than artificial intelligence. We believe that this theoretical underpinning contributes to the explanation effectiveness of our approach, as we can fully understand why a specific genre was predicted for a target user in a given context. To further illustrate this with an example, we would like to refer back to Figure 3.

In this figure, we have shown the differences between BLL_u and $ACT_{u,a}$ for two predicted genres g_1 and g_2 . Let us assume that these are the top-2 predicted genres for a target user u . According to BLL_u , we know that these genres got the highest activation levels because u has listened to them very frequently and recently. When looking at the activation levels calculated by $ACT_{u,a}$, we also take the current context into account and, thus, get an indication for the similarity of g_1 and g_2 to the genres assigned to the most recently listened artist a of user u . In our example, genre g_2 is strongly related to the current context, while genre g_1 only has a weak relation to it. Taken together, with our $ACT_{u,a}$ approach, we can easily explain genre prediction results according to three simple factors that are relevant for human memory processes according to the cognitive

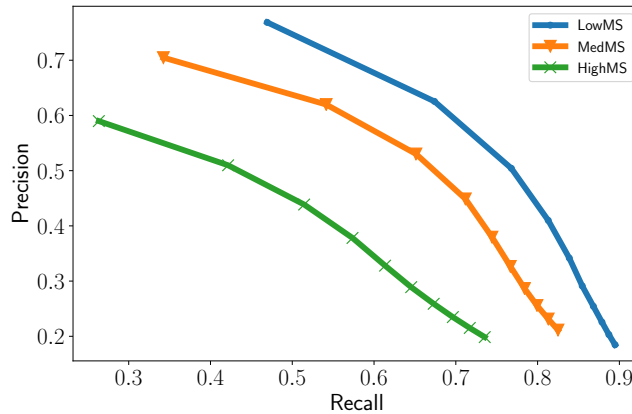


Fig. 6. Recall/precision plot of our $ACT_{u,a}$ approach for $k = 1 \dots 10$ predicted genres for the three user groups LowMS, MedMS, and HighMS. We observe good prediction accuracy results for $ACT_{u,a}$ in all settings, but especially for LowMS. This shows that our approach based on human memory processes is especially useful for predicting the music genre preferences of users with low interest in mainstream music.

architecture ACT-R: past usage frequency, past usage recency, and similarity to current context.

5 Discussion of Model Extensions

Besides the previously discussed three main factors for modeling music genre preferences, we now present potential model extensions to further enhance the transparency of our approach based on related music recommendation literature [43]. We first present the individual components comprising feature similarity, associated rewards, and randomness in behavior. We then provide the adapted activation equation and discuss further alterations of components.

Partial matching component. The partial matching component [3] is another core component of ACT-R’s activation equation [8]. The basic idea of partial matching revolves around retrieval based on *similarity*. Concerning music preferences, consider the case of two genres with mostly distinct artists but similar sounds (e.g., symphonic vs power metal), and a user with a strong preference to one genre but so far almost no listening events to the other. We can reasonably hypothesize that such a user would also show a (weaker but still noticeable) preference to the so far unexplored genre. Hence, genres can be retrieved even when they do not fully match the user’s historical preferences, i.e., only partially matches. We can predict the top- k \widetilde{G}_u^k based on partial matching of a user’s preference by:

$$\widetilde{G}_u^k = \arg \max_{g \in G}^k \sum_{f \in F_u} P \cdot M_{f,g} \quad (21)$$

where F_u is the specification for the retrieval, i.e., a set of preferred features of the user u . Example features are acoustics or lyrics that are associated with tracks, artists, and genres. A user can either explicitly set F_u , or it can automatically be extracted from the user’s historical LEs. $M_{f,g}$ represents the match similarities between a particular feature f and genre g :

$$M_{f,g} = \text{sim}(f, g) \quad (22)$$

where $\text{sim}(\cdot, \cdot)$ is an arbitrary similarity function. The factor P represents the match scale and is by default set to a constant value of 1 [8]. If $\text{sim}(\cdot, \cdot)$ is also modeled as a multiplication of values associated with f and g , then partial matching is equivalent to the dot product between the specification f and genre g . Moreover, by dynamically adapting P based on f and g , other well-established functions, such as the cosine similarity can be used.

In comparison to base-level and associative activation, partial matching can be seen as content-based retrieval and allows unexplored genres to be retrieved (e.g., in a cold-start setting [30]). Herein, it could use both user-to-item and item-to-item based recommendation.

Valuation component. Besides the core components, several extensions have been proposed, such as aggregate retrieval [32] or hybrid approaches [61]. In the following, we discuss one particular extension, i.e., the *valuation component*, that we deem relevant for modeling music preferences [12]. We predict the top- k genres \widetilde{G}_u^k based on their valuation $V_{u,g}(n)$ at the n^{th} LE of genre g by user u :

$$\widetilde{G}_u^k = \arg \max_{g \in G_u}^k V_{u,g}(n) \quad (23)$$

where valuations are learned according to the following equation:

$$V_{u,g}(n) = V_{u,g}(n-1) + \alpha(R_{u,g}(n) - V_{u,g}(n-1)) \quad (24)$$

The valuation $V_{u,g}(n)$ is based on the valuation $V_{u,g}(n-1)$ of the previous LE (i.e., $n-1$) and updated with the associated reward $R_{u,g}(n)$ weighted by the learning rate α . The initial valuation is determined by $V_{u,g}(0)$ and can be set to $V_{u,g}(0) = 0$ to specify that users do not have prior preferences. $R_{u,g}(n)$ is the *reward* that user u associates with the genre g at n^{th} LE. The reward can be, for instance, modeled according to the listening time of LEs (either total time or ratio of track length). Thus, longer LEs would result in greater valuations (i.e., a positive signal). Moreover, very short LEs could even be assigned a negative reward, as such events could indicate skipping (i.e., a negative signal). Alternatively, we could set $R_{u,g}(n) = 1$ to learn the familiarity with a given genre. Hence, the valuation component would retrieve equivalent genres as POP_u if used exclusively (but the scores would differ depending on the learning rate). Furthermore, the reward could also depend on explicit signals, such as ratings or up- and downvotes, if such data is available.

Noise. To account for *randomness* in behavior, a noise value ϵ_g can be considered for the activation, which is a (typically small) random number. Hence, the

activation level for each genre $g \in G$ deviates slightly by chance, therefore, a random genre is predicted:

$$\widetilde{G}_u^k = \arg \max_{g \in G}^k \epsilon_g \quad (25)$$

Adapted activation equation. If all components are taken together, the adapted activation equation becomes:

$$A(g, u, a, n) = B'(g, u) + \sum_{c \in G_a} W_c \cdot S_{c,g} + \sum_{f \in F_u} P \cdot M_{f,g} + V_{u,g}(n) + \epsilon_g \quad (26)$$

Hence, the activation depends on several additive components, where the genres with the highest overall activation is retrieved:

$$\widetilde{G}_u^k = \arg \max_{g \in G}^k A(g, u, a, n) \quad (27)$$

Finally, we want to emphasize that alternative implementations of the components are possible. For instance, the base level component can be simplified as $B(g, u) = \ln \left(\frac{n}{\sqrt{t_{u,g,0}}} \right)$ [1]. Thus, the adapted equation only considers the frequency (i.e., the total number of retrievals n), normalized by recency of initial retrieval (i.e., the time since the first retrieval $t_{u,g,0}$). Similarly, the associative component can also be modeled regarding probabilities of certain outcomes [12], e.g., whether a particular genre is likely listened to in a session. We, therefore, see potential for additional model extensions in future work.

6 Conclusion and Future Work

In this chapter, we extended our previous work [23], and discussed the use of cognitive models for context-aware prediction of users' music genre preferences. Based on relevant literature, we derived the theoretical underpinnings of BLL_u and $ACT_{u,a}$, two music genre preference modeling, and prediction approaches based on the human memory module of the cognitive architecture ACT-R. While BLL_u utilizes the BLL equation of ACT-R in order to model the factors of past usage frequency (i.e., popularity) and recency (i.e., time), $ACT_{u,a}$ integrates the activation equation of ACT-R to also incorporate the current context. We defined this context as the genres assigned to the most recently listened artist of the target user.

Using a dataset gathered from the music platform Last.fm, we evaluated BLL_u and $ACT_{u,a}$ against a mainstream-based approach TOP , a user-based CF approach CF_u , an item-based CF approach CF_i , a popularity-based approach POP_u as well as a time-based approach $TIME_u$. We used six evaluation metrics (i.e., recall, precision, F1-score, MRR, MAP, and nDCG) in three evaluation settings in which the evaluated users differed in terms of their inclination

to mainstream music (i.e., LowMS, MedMS, and HighMS user groups). Our evaluation results show that both BLL_u and $ACT_{u,a}$ outperform the five baseline methods in all three settings; $ACT_{u,a}$ even does so in a statistically significant manner. Furthermore, we find that especially the current context is critical when aiming for accurate genre predictions. Finally, in this chapter, we also discussed potential model extensions by surveying additional components of ACT-R.

Summed up, we have shown that human memory processes in the form of ACT-R’s activation equation can be effectively used for context-aware genre preference modeling and prediction. In addition, we also reviewed the literature in the field of cognitive-inspired recommender systems, and discussed potential model extensions of additional ACT-R components. By following such a psychology-inspired approach, we also believe that we can model a user’s preferences transparently, in contrast to, e.g., deep learning-based approaches based on latent user representations. Therefore, our approach could be useful to produce more transparent and explainable music recommender systems.

Future work. In addition to the ACT-R model extensions presented in Section 5, we plan to utilize the procedural memory processes of ACT-R. As, for instance, done in the SNIF-ACT model [11, 42], we will define so-called production rules in order to transfer the user’s preferences into actual music recommendation strategies. By making these rules transparent to the user, we aim to contribute to research on transparent recommender systems that create explainable recommendations based on psychological models [34].

As another research strand, we want to investigate fairness in the form of gender bias [33], confirmation bias [17], or popularity bias [26] in the field of cognitive- and psychology-informed recommender systems [34]. For example, we plan to study if recommendations generated using ACT-R are less prone to biased results than alternative, purely data-driven algorithms.

Finally, we will explore the effectiveness of other cognitive models in the domain of music recommender systems. For example, we plan to leverage a cognitive model of human category learning [35] to recommend music that fits a user’s current focus, similar to [16, 55], who used that model to tailor learning resources to a learner’s current task.

Reproducibility. To foster the reproducibility of our research, we use the publicly available LFM-1b dataset (see Section 3). Furthermore, we provide the source code of our approach as part of our TagRec framework [19, 58].

Acknowledgements. This work is supported by the “DDAI” COMET Module within the COMET - Competence Centers for Excellent Technologies Programme, funded by the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit), the Austrian Federal Ministry for Digital and Economic Affairs (bmdw), the Austrian Research Promotion Agency (FFG), the province of Styria (SFG) and partners from industry and academia. The COMET Programme is managed by FFG. In addition, the work received funding from the Austrian Science Fund (FWF): DFH-23 and P33526.

References

1. Altmann, E.M., Trafton, J.G.: Memory for goals: An activation-based model. *Cognitive science* **26**(1), 39–83 (2002)
2. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. *Psychological review* **111**(4), 25 pages (2004)
3. Anderson, J.R., Reder, L.M., Lebiere, C.: Working memory: Activation limitations on retrieval. *Cognitive psychology* **30**(3), 221–256 (1996)
4. Anderson, J.R., Schooler, L.J.: Reflections of the environment in memory. *Psychological science* **2**(6), 396–408 (1991)
5. Baeza-Yates, R., Ribeiro, B.d.A.N., et al.: *Modern Information Retrieval*. New York: ACM Press; Harlow, England: Addison-Wesley (2011)
6. Bauer, C., Schedl, M.: Global and country-specific mainstreamness measures: Definitions, analysis, and usage for improving personalized music recommendation systems. *PloS one* **14**(6), e0217389 (2019)
7. Biswas, P., Springett, M.: User modeling. *The Wiley Handbook of Human Computer Interaction* **1**, 143–169 (2018)
8. Bothell, D.: Act-r 7.21+ reference manual. Tech. rep., Technical Report. Technical Report, Carnegie Mellon University (2020)
9. Cantor, J.R., Zillmann, D.: The effect of affective state and emotional arousal on music appreciation. *The Journal of General Psychology* **89**(1), 97–108 (1973)
10. Cremonesi, P., Turrin, R., Lentini, E., Matteucci, M.: An evaluation methodology for collaborative recommender systems. In: *Proceedings of AXMEDIS’2008*. pp. 224–231. IEEE Computer Society, Washington, DC, USA (2008). <https://doi.org/10.1109/AXMEDIS.2008.13>, <https://doi.org/10.1109/AXMEDIS.2008.13>
11. Fu, W.T., Pirolli, P.: Snif-act: A cognitive model of user navigation on the world wide web. *Human–Computer Interaction* **22**(4), 355–412 (2007)
12. Fum, D., Stocco, A.: Memory, emotion, and rationality: An act-r interpretation for gambling task results. In: *ICCM*. pp. 106–111. Citeseer (2004)
13. Graham, G., Bechtel, W.: *A companion to cognitive science*. Blackwell (1998)
14. Järvelin, K., Price, S.L., Delcambre, L.M., Nielsen, M.L.: Discounted cumulated gain based evaluation of multiple-query ir sessions. In: *Proceedings of ECIR’2008*. pp. 4–15. Springer (2008)
15. Juslin, P.N., Sloboda, J.A.: *Music and emotion: Theory and research*. Oxford University Press (2001)
16. Kopeinik, S., Kowald, D., Hasani-Mavriqi, I., Lex, E.: Improving collaborative filtering using a cognitive model of human category learning. *The Journal of Web Science (JWS)* **2**(4), 45 – 61 (2017)
17. Kopeinik, S., Lex, E., Kowald, D., Albert, D., Seitlinger, P.: A real-life school study of confirmation bias and polarisation in information behaviour. In: *Transforming Learning with Meaningful Technologies: 14th European Conference on Technology Enhanced Learning, EC-TEL 2019, Delft, The Netherlands, September 16–19, 2019, Proceedings 14*. pp. 409–422. Springer (2019)
18. Kowald, D., Dennerlein, S., Theiler, D., Walk, S., Trattner, C.: The social semantic server: A framework to provide services on social semantic network data. In: *9th International Conference on Semantic Systems, I-SEMANTICS 2013*. pp. 50–54. CEUR (2013)
19. Kowald, D., Kopeinik, S., Lex, E.: The tagrec framework as a toolkit for the development of tag-based recommender systems. In: *Adjunct Publication of UMAP’2017*. pp. 23–28. ACM (2017)

20. Kowald, D., Lacic, E.: Popularity bias in collaborative filtering-based multimedia recommender systems. In: Boratto, L., Faralli, S., Marras, M., Stilo, G. (eds.) *Advances in Bias and Fairness in Information Retrieval*. pp. 1–11. Springer International Publishing, Cham (2022)
21. Kowald, D., Lex, E.: Evaluating tag recommender algorithms in real-world folksonomies: A comparative study. In: *Proceedings of RecSys’2015*. pp. 265–268. ACM (2015)
22. Kowald, D., Lex, E.: The influence of frequency, recency and semantic context on the reuse of tags in social tagging systems. In: *Proceedings of Hypertext’2016*. pp. 237–242. ACM, New York, NY, USA (2016)
23. Kowald, D., Lex, E., Schedl, M.: Utilizing human memory processes to model genre preferences for personalized music recommendations. In: *4th Workshop on Transparency and Explainability in Adaptive Systems through User Modeling Grounded in Psychological Theory co-located with ACM IUI’2020*. Association of Computing Machinery (2020)
24. Kowald, D., Muellner, P., Zangerle, E., Bauer, C., Schedl, M., Lex, E.: Support the underground: characteristics of beyond-mainstream music listeners. *EPJ Data Science* **10**(1), 14 (2021)
25. Kowald, D., Pujari, S.C., Lex, E.: Temporal effects on hashtag reuse in twitter: A cognitive-inspired hashtag recommendation approach. In: *Proceedings of WWW’2017*. p. 10 pages. ACM (2017)
26. Kowald, D., Schedl, M., Lex, E.: The unfairness of popularity bias in music recommendation: A reproducibility study. In: *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II* 42. pp. 35–42. Springer (2020)
27. Kowald, D., Seitlinger, P., Trattner, C., Ley, T.: Long time no see: The probability of reusing tags as a function of frequency and recency. In: *Proceedings of the 23rd International Conference on World Wide Web*. pp. 463–468 (2014)
28. Lacic, E., Kowald, D., Eberhard, L., Trattner, C., Parra, D., Marinho, L.B.: Utilizing online social network and location-based data to recommend products and categories in online marketplaces. In: *Mining, Modeling, and Recommending ‘Things’ in Social Media: 4th International Workshops, MUSE 2013, Prague, Czech Republic, September 23, 2013, and MSM 2013, Paris, France, May 1, 2013, Revised Selected Papers*. pp. 96–115. Springer (2015)
29. Lacic, E., Kowald, D., Seitlinger, P., Trattner, C., Parra, D.: Recommending items in social tagging systems using tag and time information. *Proceedings of the 1st International Workshop on Social Personalisation (SP’2014) co-located with Hypertext’2014* (2014)
30. Lacic, E., Kowald, D., Traub, M., Luzhnica, G., Simon, J.P., Lex, E.: Tackling cold-start users in recommender systems with indoor positioning systems. In: *Poster Proceedings of the 9th {ACM} Conference on Recommender Systems*. Association of Computing Machinery (2015)
31. Lacic, E., Reiter-Haas, M., Duricic, T., Slawicek, V., Lex, E.: Should we embed? a study on the online performance of utilizing embeddings for real-time job recommendations. In: *Proceedings of the 13th ACM Conference on Recommender Systems*. pp. 496–500 (2019)
32. Lebiere, C.: Blending: An act-r mechanism for aggregate retrievals. In: *Proceedings of the sixth annual act-r workshop, george mason university, fairfax, va, usa* (1999)
33. Lesota, O., Melchiorre, A., Rekabsaz, N., Brandl, S., Kowald, D., Lex, E., Schedl, M.: Analyzing item popularity bias of music recommender systems: are different

- genders equally affected? In: Proceedings of the 15th ACM Conference on Recommender Systems. pp. 601–606 (2021)
34. Lex, E., Kowald, D., Seitlinger, P., Tran, T.N.T., Felfernig, A., Schedl, M.: Psychology-informed recommender systems. *Foundations and Trends® in Information Retrieval* **15**(2), 134–242 (2021). <https://doi.org/10.1561/15000000090>, <http://dx.doi.org/10.1561/15000000090>
 35. Love, B.C., Medin, D.L., Gureckis, T.M.: Sustain: a network model of category learning. *Psychological review* **111**(2), 309 (2004)
 36. Moscati, M., Wallmann, C., Reiter-Haas, M., Kowald, D., Lex, E., Schedl, M.: Integrating the act-r framework with collaborative filtering for explainable sequential music recommendation. In: Proceedings of the 17th ACM Conference on Recommender Systems. pp. 840–847 (2023)
 37. Mozer, M.C., Lindsey, R.V.: Predicting and improving memory retention: Psychological theory matters in the big data era. In: *Big data in cognitive science*, pp. 43–73. Psychology Press (2016)
 38. North, A., Hargreaves, D.: *The social and applied psychology of music*. OUP Oxford (2008)
 39. Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., Semenuk, A.: What happened to cognitive science? *Nature human behaviour* **3**(8), 782–791 (2019)
 40. Oord, A.v.d., Dieleman, S., Schrauwen, B.: Deep content-based music recommendation. In: Proceedings of NIPS’2013. pp. 2643–2651. Curran Associates Inc., USA (2013)
 41. Pereira, C.S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S.L., Brattico, E.: Music and emotions in the brain: familiarity matters. *PloS one* **6**(11), e27241 (2011)
 42. Pirolli, P., Fu, W.T.: Snif-act: A model of information foraging on the world wide web. In: *International Conference on User Modeling*. pp. 45–54. Springer (2003)
 43. Reiter-Haas, M., Parada-Cabaleiro, E., Schedl, M., Motamedi, E., Tkalcic, M., Lex, E.: Predicting music relistening behavior using the act-r framework. In: Proceedings of the 15th ACM Conference on Recommender Systems. pp. 702–707 (2021)
 44. Rentfrow, P.J., Gosling, S.D.: The do re mi’s of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology* **84**(6), 21 pages (2003)
 45. Schedl, M.: The LFM-1b Dataset for Music Retrieval and Recommendation. In: Proceedings of the 2016 Conference on Multimedia Retrieval. pp. 103–110. ACM (2016)
 46. Schedl, M., Bauer, C.: Distance-and rank-based music mainstreamness measurement. In: *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. pp. 364–367. ACM (2017)
 47. Schedl, M., Bauer, C.: An Analysis of Global and Regional Mainstreamness for Personalized Music Recommender Systems. *Journal of Mobile Multimedia* **14**, 95–112 (2018)
 48. Schedl, M., Bauer, C., Reisinger, W., Kowald, D., Lex, E.: Listener modeling and context-aware music recommendation based on country archetypes. *Frontiers in Artificial Intelligence* **3**, 508725 (2021)
 49. Schedl, M., Ferwerda, B.: Large-scale Analysis of Group-specific Music Genre Taste From Collaborative Tags. In: Proceedings of ISM’2017. pp. 479–482. IEEE (2017)
 50. Schedl, M., Gómez, E., Trent, E., Tkalčić, M., Eghbal-Zadeh, H., Martorell, A.: On the Interrelation between Listener Characteristics and the Perception of Emotions in Classical Orchestra Music. *IEEE Transactions on Affective Computing* **9**, 507–525 (2018)

51. Schedl, M., Hauger, D.: Tailoring music recommendations to users by considering diversity, mainstreaminess, and novelty. In: Proceedings of SIGIR'2015. pp. 947–950. ACM (2015)
52. Schedl, M., Knees, P., McFee, B., Bogdanov, D., Kaminskas, M.: Music recommender systems. In: Recommender systems handbook, pp. 453–492. Springer (2015)
53. Schedl, M., Zamani, H., Chen, C.W., Deldjoo, Y., Elahi, M.: Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* **7**(2), 95–116 (Jun 2018)
54. Schubert, E.: The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music* **35**(3), 499–515 (2007)
55. Seitlinger, P., Kowald, D., Kopeinik, S., Hasani-Mavriqi, I., Lex, E., Ley, T.: Attention please! a hybrid resource recommender mimicking attention-interpretation dynamics. In: Proceedings of WWW'2015 companion. pp. 339–345. ACM (2015)
56. Shi, Y., Larson, M., Hanjalic, A.: 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 (May 2014)
57. Sun, R.: Introduction to computational cognitive modeling. Cambridge handbook of computational psychology pp. 3–19 (2008)
58. Trattner, C., Kowald, D., Lacic, E.: Tagrec: towards a toolkit for reproducible evaluation and development of tag-based recommender algorithms. *ACM SIGWEB Newsletter* pp. 1–10 (2015)
59. Trattner, C., Kowald, D., Seitlinger, P., Kopeinik, S., Ley, T.: Modeling activation processes in human memory to predict the reuse of tags. *The Journal of Web Science* **2** (2016)
60. Van Maanen, L., Marewski, J.N.: Recommender systems for literature selection: A competition between decision making and memory models. In: Proceedings of the 31st Annual Conference of the Cognitive Science Society. pp. 2914–2919 (2009)
61. Vinokurov, Y., Lebiere, C., Herd, S., O'Reilly, R.: A metacognitive classifier using a hybrid act-r/leabra architecture. In: Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence (2011)
62. Wheeler, S.: Learning theories: Adaptive control of thought (2014), [Online under <http://www.teachthought.com/learning/theory-cognitive-architecture/>; accessed 19-December-2019]
63. Zajonc, R.B.: Attitudinal effects of mere exposure. *Journal of personality and social psychology* **9**(2p2), 1 (1968)
64. Zangerle, E., Pichl, M.: Content-based user models: Modeling the many faces of musical preference. In: 19th International Society for Music Information Retrieval Conference (2018)