Integrating the ACT-R Framework with Collaborative Filtering for Explainable Sequential Music Recommendation

Marta Moscati[1], Christian Wallmann[2], Markus Reiter-Haas[3], Dominik Kowald[3,4], Elisabeth Lex[3], Markus Schedl[1,5]

[1]Institute of Computational Perception, Johannes Kepler University Linz (Austria)
[2]Welser Profile Gresten (Austria)
[3]Graz University of Technology (Austria)
[4]Know-Center GmbH Graz (Austria)
[5]Human-centered AI Group, AI Lab, Linz Institute of Technology (Austria)

The ACT-R Cognitive Architecture for Music Relistening

The Adaptive Control of Thought-Rational cognitive architecture (ACT-R) effectively models music relistening by modelling human memory. A track is assigned a probability of being relistened based on several *scores*:

Base-Level Learning

$$B_i = \sum_{j=1}^{n} (t_{\text{ref}} - t_{ij})^{-d}$$

favors tracks listened to frequently and recently.

Valuation

$$V_i^n = V_i^{n-1} + \alpha\left(R_i^n - V_i^{n-1}\right)$$

favors tracks listened to entirely.

Spreading

$$S_i = \frac{P(i \in C_k)}{P(i)}$$

favors tracks listened to in sessions containing the last track.

Partial Matching

$$P_i = \text{sim}(i, k)$$

favors tracks similar to the last track.

Since each score is designed to model an aspect of memory, the overall ACT-R score of a track is intrinsically explainable. However, ACT-R assigns positive scores only to tracks that the user already listened to. Therefore, it cannot recommend novel tracks.

Combining ACT-R with Collaborative Filtering

We propose four algorithms that integrate collaborative filtering with ACT-R to design sequential music recommender systems that

- Leverage past collective behaviors for novel recommendations,
- Leverage ACT-R for explainable recommendations.

Rolling Session Completion

We conduct experiments on a one-month extract of the LFM-2b dataset. The performance of the algorithms is evaluated on the task of rolling session completion:

- For a session of length N, we consider all possible initial segment lengths, $l = 1, \ldots, N-1$.
- The listening events of a 7-days window and of the session's initial segment are assumed to be known.
- We recommend tracks for the remainder of the session.

Evaluation

The quality of recommendations is evaluated in terms of accuracy and beyond-accuracy metrics:

Accuracy is measured in terms of F_1 score.

Diversity is measured as Shannon entropy of the distribution of recommended tracks over genres.

Novelty is measured as the fraction of recommended tracks that have not been listened to by the target user.

P-Novelty measures the precision of novel recommendations.

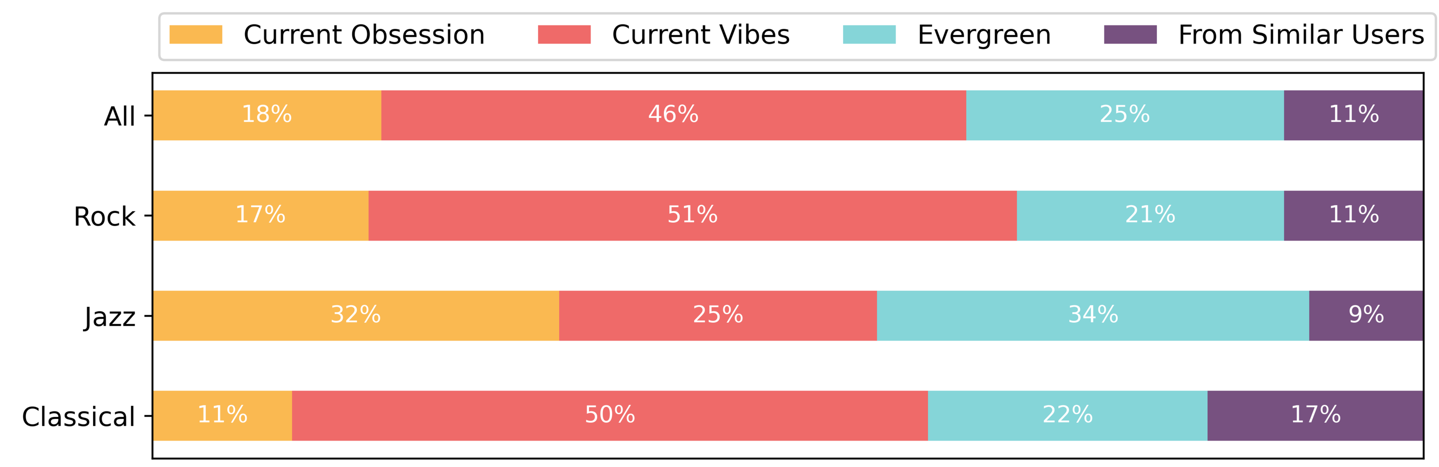
Model Performance

The proposed algorithms (in blue) achieve the best performance in terms of beyond-accuracy metrics, although not outperforming the accuracy of established algorithms for sequential recommendation.

	F_1	Diversity	Novelty	P-Novelty
GRU4Rec	**0.142**	0.929	0.716	0.126
Temporal UserkNN	0.122	0.786	0.631	0.146
Item BPR	0.114	0.658	**0.846**	0.130
Weighted MultVAE	0.111	0.941	0.554	0.136
ACT-R + BPR	0.104	0.923	0.056	**0.239**
Social ACT-R	0.101	**0.945**	0.056	0.155
MostRecent	0.094	0.891	0.000	0.000
ACT-R	0.093	0.889	0.000	0.000

Explainability

Leveraging ACT-R allows for explainable recommendations, which can be used to discover patterns in recommendations.



For instance, the relative importance of each component of the recommendation score varies depending on the genre of the track recommended. Since ACT-R is modeled on listening behavior, this is an indication that listening behaviors depend on genre.

This research was funded in whole, or in part, by the Austrian Science Funds (FWF): P33526 and DFH-23, and by the State of Upper Austria and the Federal Ministry of Education, Science, and Research, through grant LIT-2020-9-SEE-113. Memory icon created by Good Ware - Flaticon.

marta.moscati@jku.at
https://mmosc.github.io/

https://hcai.at/
https://github.com/hcai-mms/actr