The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study

Dominik Kowald, Markus Schedl, Elisabeth Lex

ECIR 2020 - Reproducibility Track
14 - 17 April 2020

Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration

Motivation

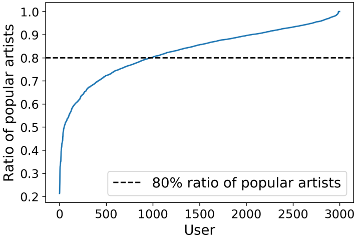
- Recommender systems support users in finding relevant information in large information spaces [RRS11]
- Popularity bias $\rightarrow$ underrepresentation of unpopular items in recommendation lists [BHS06]
- [AMBM19] has shown that this also leads to unfair treatment of users with less interest in popular items
- We reproduce this study from the movie domain in the music domain $\rightarrow$ vast amount of items [SZC$^+$18]
- Research questions [AMBM19]
  - **RQ1**: To what extent are users interested in popular items (i.e., music artists)?
  - **RQ2**: To what extent is the recommendation quality affected by this popularity bias?

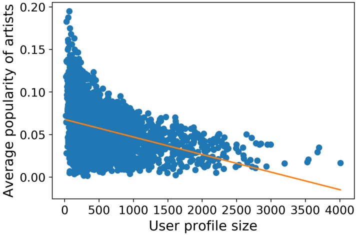
Dataset

- LFM-1b dataset [Sch16]
  - 120k users, 3.1M artists, 1.1B listening events
  - Metadata, e.g., mainstreaminess score for users [BS19]
- LFM-1b user groups
  - 1k users with lowest (LowMS), with medium mainstreaminess (MedMS) and with highest mainstreaminess (HighMS)
  - 3k users, 352k artists, 1.7M listening events $\rightarrow$ in MovieLens dataset only 3.9k movies
  - Available via Zenodo: https://zenodo.org/record/3475975

RQ1: Interest of Users in (Un)Popular Music

- Definitions [AMBM19]
  - Popular artist → in top 20% of artists with the highest number of listeners
  - Artist popularity → ratio of users who have listened to this artist
- 1/3 of our users listen to at least 20% of unpopular artists → LowMS
- Users with larger profile sizes tend to listen to more unpopular artists



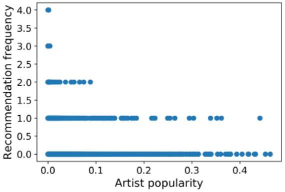


RQ2: Popularity Bias in Music Recommendations

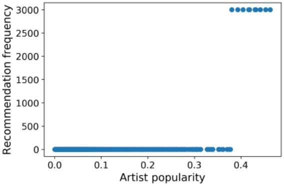
- Python-based open-source framework Surprise
- Rating prediction $\rightarrow$ number of listening events of user for artist
- Recommend top-10 artists with highest predicted preferences to user
- Evaluation protocol [AMBM19]
  - Random 80/20 train-test split
  - 3 baselines: Random, MostPopular, UserItemAvg [Kor10]
  - 2 knn-based approaches: UserKNN, UserKNNAvg [SFHS07]
  - 1 matrix factorization-based approach: NMF [LZXZ14]
- Available via Github: https://github.com/domkowald/LFM1b-analyses

surprise

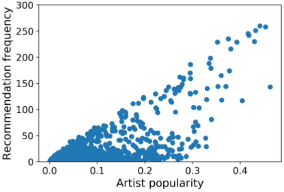
RQ2: Artist Popularity and Recommendation Frequency



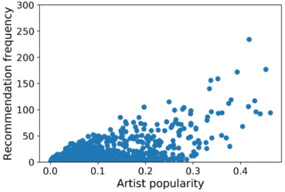
(a) Random.



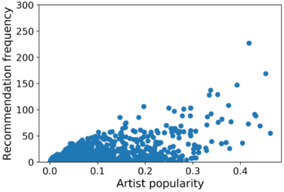
(b) MostPopular.



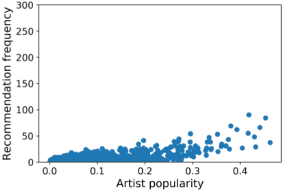
(c) UserItemAvg.



(d) UserKNN.



(e) UserKNNAvg.



(f) NMF.

RQ2: Recommendation Accuracy

- Mean Average Error (MAE) metric → the lower the better
- LowMS group receives worse recommendations than MedMS and HighMS for all algorithms
- Statistically significant according to a t-test with $p < .005$ as indicated by ***
- Also interesting:
  - NFM provides the best and fairest results
  - MedMS provides the best results → larger average profile size than LowMS and HighMS

User group	UserItemAvg	UserKNN	UserKNNAvg	NMF
LowMS	42.991***	49.813***	46.631***	**38.515*****
MedMS	33.934	42.527	37.623	**30.555**
HighMS	40.727	46.036	43.284	**37.305**
All	38.599	45.678	41.927	**34.895**

Conclusion and Future Work

- We reproduced the study of [AMBM19] on the unfairness of popularity bias in recommender systems in the music domain
- We get the same results:
  - RQ1: Users have interest in unpopular items and these users also have large profile sizes
  - RQ2: Users with interest in unpopular items receive worst recommendations

Future RQ

What are the special characteristics of these low-mainstream users and how can we provide better recommendations for them?

Thank you for your attention! Questions?

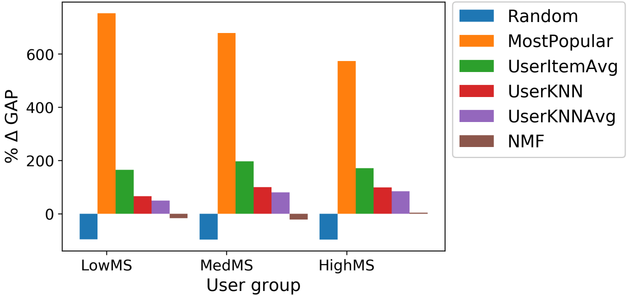
Contacts:
dkowald [AT] know-center [DOT] at
markus.schedl [AT] jku [DOT] at
elisabeth.lex [AT] tugraz [DOT] at

Data:
https://zenodo.org/record/3475975

Code:
https://github.com/domkowald/LFM1b-analyses

Appendix: Popularity Bias for User Groups

- Group Average Precision (GAP) metric [AMBM19]
- $GAP(g)_p \rightarrow$ average artist popularity in the user profiles p of group g
- $GAP(g)_r \rightarrow$ average artist popularity in the recommendation lists r of group g
- $\Delta GAP = \frac{GAP(g)_r - GAP(g)_p}{GAP(g)_p}$
- No clear difference between the groups except for MostPopular $\rightarrow$ large number of items (352k artists vs. 3.9k movies)



References I

Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher, *The unfairness of popularity bias in recommendation*, Workshop on Recommendation in Multi-stakeholder Environments (RMSE'19), in conjunction with the 13th ACM Conference on Recommender Systems, RecSys'19, 2019.

Erik Brynjolfsson, Yu Jeffrey Hu, and Michael D Smith, *From niches to riches: Anatomy of the long tail*, Sloan Management Review **47** (2006), no. 4, 67–71.

Christine Bauer and Markus Schedl, *Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems*, PloS one **14** (2019), no. 6, e0217389.

References II

- Yehuda Koren, *Factor in the neighbors: Scalable and accurate collaborative filtering*, ACM Transactions on Knowledge Discovery from Data (TKDD) **4** (2010), no. 1, 1.
- Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu, *An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems*, IEEE Transactions on Industrial Informatics **10** (2014), no. 2, 1273–1284.
- Francesco Ricci, Lior Rokach, and Bracha Shapira, *Introduction to recommender systems handbook*, Recommender Systems Handbook, Springer, 2011, pp. 1–35.

References III

- Markus Schedl, *The LFM-1B Dataset for Music Retrieval and Recommendation*, Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval (New York, NY, USA), ICMR '16, ACM, 2016, pp. 103–110.
- J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen, *Collaborative filtering recommender systems*, The Adaptive Web, Springer, 2007, pp. 291–324.
- Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi, *Current challenges and visions in music recommender systems research*, International Journal of Multimedia Information Retrieval **7** (2018), no. 2, 95–116.