Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems

Dominik Kowald, Emanuel Lacic

ECIR 2022 - BIAS Workshop
10 - 14 April 2022

Motivation

- **Collaborative filtering (CF)** is vastly used in the field of multimedia recommender systems (MMRS) [SFHS07], e.g., **movies, music, digital books, animes**
- **CF** $\rightarrow$ popularity bias $\rightarrow$ **underrepresentation of unpopular items** in recommendation lists [BHS06]
- [AMBM19, KSL20] has shown that CF-based approaches lead to **unfair treatment of users with less interest in popular items** in the movie and music domains
- Four MMRS domains with common evaluation protocol $\rightarrow$ **two RQs**:
  - **RQ1**: To what extent does an item's popularity affect this item's recommendation frequency in MMRS? **(item level)**
  - **RQ2**: To what extent does a user's inclination to popular items affect the quality of MMRS? **(user level)**

MMRS Datasets

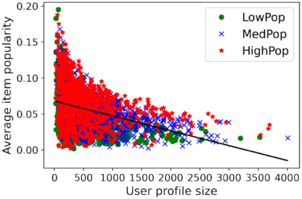
- Four datasets to represent the four domains
  - **Last.fm**: LFM-1b [Sch16] dataset provided by JKU Linz
  - **MovieLens**: Movielens 1M dataset provided by GroupLens
  - **BookCrossing**: provided by University Freiburg
  - **MyAnimeList**: provided by Kaggle
  - Remove users with < 50 or $> 2,000$ ratings
- User groups
  - 1k users with lowest (**LowPop**), with medium (**MedPop**) and with highest (**HighPop**) inclination to popularity
  - Available via Zenodo: https://zenodo.org/record/6123879

Dataset	$\|U\|$	$\|I\|$	$\|R\|$	$\|R\|/\|U\|$	$\|R\|/\|I\|$	Sparsity	R-range
Last.fm	3,000	352,805	1,755,361	585	5	0.998	[1-1,000]
MovieLens	3,000	3,667	675,610	225	184	0.938	[1-5]
BookCrossing	3,000	223,607	577,414	192	3	0.999	[1-10]
MyAnimeList	3,000	9,450	649,814	216	69	0.977	[1-10]

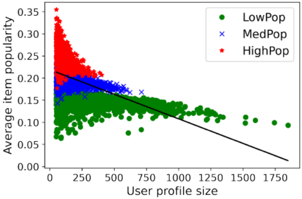
Defining Popularity

- Item level [AMBM19, KSL20]
  - **Item popularity**: $Pop_i = \frac{|U_i|}{|U|}$
    - U_i is the set of users who rated item i
  - **Average item popularity** in user profile: $Pop_{i,u} = \frac{1}{|I_u|} \sum_{i \in I_u} Pop_i$
    - I_u is the set of items in the user profile
  - **Popular item** $\rightarrow$ it falls within the top-20% of Pop_i scores
- User level [AMBM19, KSL20]
  - $I_{u,Pop}$ is the set of **popular items in the user profile**
  - **Ratio of popular items** in user profile: $Pop_u = \frac{|I_{u,Pop}|}{|I_u|}$
    - Create LowPop, MedPop, HighPop user groups in MovieLens, BookCrossing, MyAnimeList
- Mainstreaminess (**Last.fm** $\rightarrow$ repeated consumption) [BS19]
  - Compare artist-playclount (APC) dist. between u and average user
  - $M_{R,APC}^{global}(u) = \tau(ranks(APC), ranks(APC(u)))$

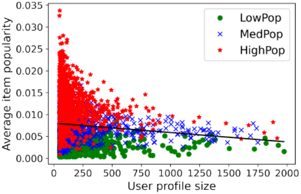
Relationship Between Popularity and Profile Size



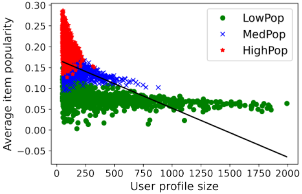
(a) Last.fm



(b) MovieLens



(c) BookCrossing



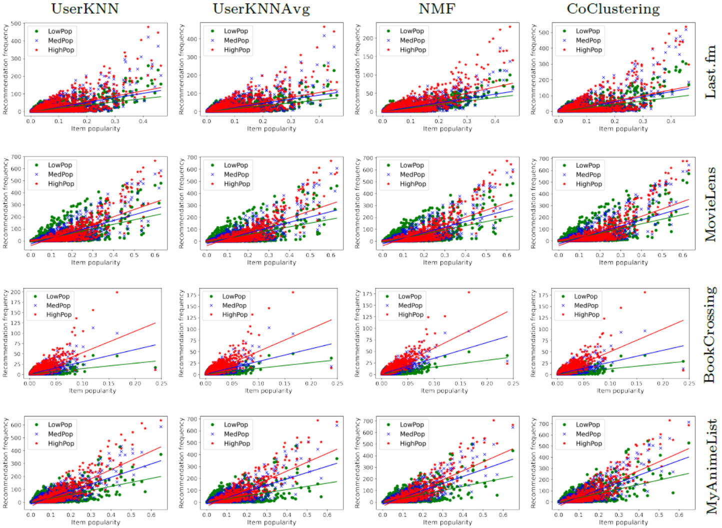
(d) MyAnimeList

- LowPop users have large profile sizes and **are important for MMRS!**

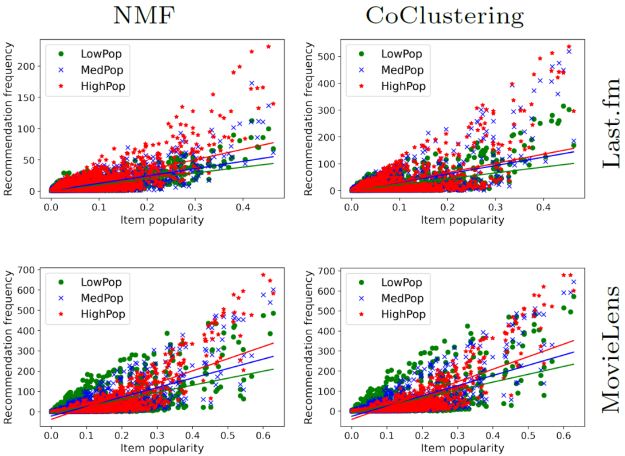
Recommendation Algorithms and Evaluation Protocol

- Python-based open-source framework **Surprise**
- **Rating prediction** → predict APC in case of Last.fm
- **Mean Average Error (MAE)** metric → the lower the better
- 4 CF-based recommendation algorithms
  - 2 **knn**-based approaches: UserKNN, UserKNNAvg [SFHS07]
  - 1 **matrix factorization**-based approach: NMF [LZXZ14]
  - 1 scalable **co-clustering**-based approach: CoClustering [GM05]
- Evaluation protocol
  - Random **80/20** train-test split
  - **Five-fold** cross validation
  - Pairwise **t-test** between LowPop and MedPop / LowPop and HighPop
- Available via Github: https://github.com/domkowald/FairRecSys

RQ1: Popularity Bias on Item Level



RQ1: Popularity Bias on Item Level (selection)



- Positive relationship between item popularity and recommendation frequency for all user groups! **(but weakest one for LowPop)**

RQ2: Popularity Bias on User Level

Dataset	User group	UserKNN	UserKNNAvg	NMF	CoClustering
Last.fm	LowPop	49.489***	46.483***	**39.641*****	47.304***
	MedPop	*42.899*	*37.940*	**32.405**	*37.918*
	HighPop	45.805	43.070	**38.580**	42.982
MovieLens	LowPop	0.801***	0.763***	0.753***	**0.738*****
	MedPop	0.748	0.727	0.722	**0.705**
	HighPop	*0.716*	*0.697*	*0.701*	**0.683**
BookCrossing	LowPop	1.403***	**1.372*****	1.424***	1.392***
	MedPop	*1.154*	**1.122**	*1.214*	*1.134*
	HighPop	1.206	**1.155**	1.274	1.162
MyAnimeList	LowPop	1.373***	**1.001*****	1.010***	1.001***
	MedPop	1.341	**0.952**	0.968	*0.956*
	HighPop	*1.311*	**0.948**	*0.951*	0.975

- **Statistically significant** according to t-test with $p < .005$ (***)

Conclusion and Future Work

- **RQ1:** Popular items have higher probability of getting recommended than unpopular items
- **RQ2:** Users with interest in unpopular items (i.e., LowPop) receive the worst recommendations
- **BUT** LowPop users have the largest user profile sizes
- **Future Work**
  - **Investigate differences in results** across user groups and algorithms (e.g., why does MedPop gets the best results in Last.fm?)
  - **Popularity bias mitigation** strategies, e.g.,
    - In-processing, e.g., regularization [BFM21]
    - Aggregate diversity (unique recommended items) [AR17]
    - Calibration based on popularity [AMB+21]
    - Personalized re-ranking [ABM19]
  - Investigate further **popularity bias evaluation metrics**, e.g., GAP (group average popularity) [AMBM19]

Thank you! Questions?

Contacts:

dkowald [AT] know-center [DOT] at
elacic [AT] know-center [DOT] at

Data:

https://zenodo.org/record/6123879

Code:

https://github.com/domkowald/FairRecSys

Paper:

https://arxiv.org/abs/2203.00376

Posters in main (Wednesday) and industry track (Thursday)

References I

Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher, *Managing popularity bias in recommender systems with personalized re-ranking*, The thirty-second international flairs conference, 2019.

Himan Abdollahpouri, Masoud Mansoury, Robin Burke, Bamshad Mobasher, and Edward Malthouse, *User-centered evaluation of popularity bias in recommender systems*, Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, 2021, pp. 119–129.

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.

References II

Arda Antikacioglu and R Ravi, *Post processing recommender systems for diversity*, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 707–716.

Ludovico Boratto, Gianni Fenu, and Mirko Marras, *Connecting user and item perspectives in popularity debiasing for collaborative recommendation*, Information Processing & Management **58** (2021), no. 1, 102387.

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.

References III

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.

Thomas George and Srujana Merugu, *A scalable collaborative filtering framework based on co-clustering*, Fifth IEEE International Conference on Data Mining (ICDM'05), IEEE, 2005, pp. 4–pp.

Dominik Kowald, Markus Schedl, and Elisabeth Lex, *The unfairness of popularity bias in music recommendation: A reproducibility study*, European Conference on Information Retrieval, Springer, 2020, pp. 35–42.

References IV

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.

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.