# Struggling with popularity bias in recommendation systems

The long-tail phenomenon is common in recommendation systems data: in most cases, a small fraction of popular items account for the majority of user interactions. When trained on such data, the model usually gives higher scores to popular items than their ideal values while simply predicting unpopular items as negative. As a result, popular items are recommended even more frequently than their original popularity exhibited in the dataset.

## 1 Introduction

Table 1 shows a list of solutions on how to struggle with popularity bias in data or how to measure it in existing systems.

Table 1: Comparative analysis of basic solutions: how to estimate and solve popularity bias

| Solution | Strengths | Weakness |
| --- | --- | --- |
| Agent-Based Modeling [1] | This type of modeling provides a playground for how to simulate user interactions on synthetic data and measure the bias without running AB on real users. | A model is limited by the setting rules of interaction; the real-world processes are more sophisticated. |
| Multifactorial inversed propensity score (reweightening technique) [2] | The method is straightforward and easy for result interpretation. | This method might significantly reduce model quality metrics. |
| Regularization/ Causal graph / Adversarial learning [3] | This is a long 40-page article with a detailed overview of different types of biases (not only popularity bias) and the ways to struggle with them. For popularity bias, the authors suggest regularization (straightforward and simple), causal graphs (explainable), and adversarial learning (balancing representation). | The authors provide only a helicopter view on popularity bias with classical methods. The problem of regularization and adversarial learning is that they hurt model accuracy; the problem of causal graph is that you need assumptions for data generation. |
| Crowd-based estimation [4] | The authors get real users feedback for different settings of RS. | Expensive and hard to scale. The knowledge that they participate in experiments creates some bias itself. |

## References


[1] Gediminas Adomavicius, Dietmar Jannach, Stephan Leitner, and Jingjing Zhang. Understanding longitudinal dynamics of recommender systems with agent-based modeling and simulation. *arXiv preprint arXiv:2108.11068*, 2021.

[2] Jin Huang, Harrie Oosterhuis, Masoud Mansoury, Herke Van Hoof, and Maarten de Rijke. Going beyond popularity and positivity bias: Correcting for multifactorial bias in recommender systems. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 416–426, 2024.

[3] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, 41(3):1–39, 2023.

[4] Oleg Lesota, Gustavo Escobedo, Yashar Deldjoo, Bruce Ferwerda, Simone Kopeinik, Elisabeth Lex, Navid Rekabsaz, and Markus Schedl. Computational versus perceived popularity miscalibration in recommender systems. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1889–1893, 2023.