On the Longitudinal Impact of Exposure Bias in Recommender Systems

On the Longitudinal Impact of Exposure Bias in Recommender Systems

Andrea Pisani[1,2][0009−0001−9736−522X]

[1] Politecnico di Milano, Italy
[2] Politecnico di Torino, Italy
andrea.pisani@polito.it

Abstract. Recommender Systems are a cornerstone of digital interaction, particularly on platforms such as social media, where personalised content is pivotal to user engagement. One of the main challenges related to such Systems is exposure bias: since users are unconsciously subject to what the System shows them, they engage with a biased sample of items. This proposal introduces the concept of *compound* exposure bias — i.e., the progressive accumulation and intensification of exposure bias over time, due to Recommender models continuously adapting upon the outcomes of their previous recommendations. Such a recursive cycle may lead to skewed content exposure, thus hindering diverse information access. Additionally, compound bias might contribute to problematic user behaviours, including addiction, since the reliance of many Recommenders on implicit feedback signals inherently equates the time spent on a platform with the degree of user satisfaction with it.

We propose to study the dynamics of compound exposure bias through a multi-phase approach, exploring its real-world impact on recommendation accuracy and beyond. Firstly, we will establish formal methodologies to detect and quantify compound bias within the context of recommendation. This involves both theoretical modelling and empirical, longitudinal evaluation. Secondly, we will propose novel strategies to mitigate compound bias while retaining recommendation effectiveness as much as possible, leveraging techniques such as bandits, agentic models, re-ranking, and fairness-constrained learning.

This research seeks to enhance the long-term fairness, robustness, and user-centricity of Recommender Systems. Our goal is to offer insights for improving user experience and reducing detrimental effects associated with personalised content consumption.

Keywords: Agentic models · Bandits · Exposure bias · Implicit feedback · Longitudinal analysis · Multi-objective optimisation · Recommender systems · Re-ranking.

1 Introduction and Related Works

Recommender Systems (RS) are an integral part of modern digital ecosystems: they extract personalized shortlists starting from vaste online catalogues, in order to ease user experience and improve the quality of user interactions. In social media feeds, for example, they serve as key drivers of engagement. However, current RS are affected by several types of bias, which hinder their effective utility [3]. Chen et al. [5] identify three sources of bias within the life cycle of RS, namely *data bias*, *inductive bias* and *recommendation bias*. *Data bias* is created by users while interacting with the items they are shown. Examples of data bias include the tendency of users to interact only with the topmost items within a ranked list (*position bias*), and the incorrect assumption that all unobserved interactions are negative, while, in reality, users are much more likely to interact just with a small portion of what is recommended to them. Overlooking unobserved interactions takes the name of *exposure bias* when considering items that have not been recommended, and, conversely, the name of *selection bias* when considering recommended items that the users choose to ignore. *Inductive bias* represents the assumptions a Machine Learning-based RS takes in order to learn a generalised model of user behaviour, while *recommendation bias* describes the possible disproportions in the recommendations a RS provides, such as the notorious popularity bias. Numerous works [13, 12, 10, 11] have shown that such biases significantly decrease the effectiveness of many RS. Moreover, real-world RS are periodically re-trained, thus cycling back and forth from the collection of biased data to the production of biased recommendations, in a self-reinforcing loop which increases the impact of biases over time. However, making a longitudinal study to assess such impact requires long-term access to large user bases, in order to track the progressive evolution of their interactions, which is not a realistic prospect for the majority of researchers. Consequently, some studies resorted to simulation techniques to approximate a longitudinal analysis. For instance, Jannach et al. [9] report an extensive heuristic-based simulation investigating popularity bias reinforcement; Ferraro et al. [6] use the same methodology to study longitudinal effects on session-based RS; Adomavicius et al. and Zhou et al. [1, 19] both employ agent-based modelling and simulation [4] to investigate the impact of various types of bias on user preferences. All the aforementioned works find confirmations of progressive bias reinforcement over time. Adding to the significance of this problem is the fact that many online platforms adopt implicit user feedback as the main driver to train their RS, inherently correlating the time users spend on their platform to their degree of satisfaction with it, an assumption that can be interpreted as induction bias. In contexts such as social media, where exposure bias is predominant, as a sizeable portion of the interactions is influenced by RS, the combination and progressive reinforcement of exposure and inductive bias leads to less and less diverse recommendations that are also selected to maximise the time users spend on the platform, exposing them to the risk of developing harmful behaviours such as addiction [17, 14].

We propose to structure a two-phase investigation on the longitudinal impact of bias in social media RS. In Section 2, we provide a formulation of *compound bias*, i.e., the self-reinforcement of bias over time. In Section 3 and 4, we outline the main research questions which will drive our investigation; finally, we draw several conclusive statements in Section 5.

2 Problem Formulation

As described in Chen et al. [5], the task of effective recommendation can be summarised as the minimisation of a *true risk* function, i.e., the expected error between the modelled and true ratings, averaged over all users and items within the considered platform. As the ideal probability distribution of ratings is not known a priori, the task translates, in practice, to the minimisation of an *empirical risk* function, i.e., the average error between the modelled and true ratings over all the observed ratings within the dataset. Such translation can be considered an unbiased estimation assuming that the train and test set data are independent and identically distributed (i.i.d), and that the dataset is sufficiently large, following PAC learning theory [8]. In real-world scenarios, the i.i.d assumption falls: the empirical risk estimation is biased. Moreover, from the moment in which a (biased) RS is employed, newly generated interaction data (those deriving from the user "trusting" the recommendations) will reflect such bias: thus, the interaction distribution observed in training will further deviate from the ideal distribution. Upon re-training, the estimator's bias is bound to grow.

We define as *compound bias* at n the bias value relative to a RS which has been re-trained n times. Since the true risk is not observable, its exact value cannot be quantified. Nevertheless, the compound *rate* can be estimated by measuring the decay in recommendation effectiveness (considering accuracy, beyond accuracy or both) at each re-training, i.e., by means of a longitudinal analysis. As mentioned in Section 1, due to lack of continuous access to industrial grade platforms, most researchers resort to simulating such analysis.

3 Detection of *Compound Bias* and Impact Assessment

To the best of our knowledge, the existing simulation frameworks [18, 19] use simple algorithms and broad assumptions to model user behaviour. In the first phase of our investigation, we propose to structure a simulation using Large Language Model (LLM) agents, aligned on ground-truth user interactions, as recent literature has shown their accuracy in numerous fields [7, 2]. We will specifically focus on the effects of exposure bias on social media interactions, studying both the rate of decay in accuracy and beyond accuracy metrics and the effects on user consumption in terms of time spent on the platform, estimated by the total amount of posts viewed per each time-frame in between RS re-training steps. We will analyse the long-term behaviour of a number of recommendation models that have already been described in literature, from the classical content-based and collaborative filtering approaches to matrix factorisation to neural network-based RS. Our goal will be to answer the following research questions:

RQ1. To what extent does compound exposure bias negatively affect the long-term accuracy of existing RS?

RQ2. To what extent does compound exposure bias negatively affect the long-term diversity, novelty and serendipity of the recommendations provided by existing RS?

RQ3. Does compound exposure bias lead to an increase in the time users spend on social media?

4 Compound Bias Mitigation

After an initial analysis of the effects of compound exposure bias over existing RS in the context of social media, we propose to evaluate with the same simulation framework the effectiveness of several mitigation methods, including debiasing techniques applied at re-training time (e.g., propensity-based debiasing [16]), re-ranking techniques at inference time [6], and fairness-constrained learning techniques [15] to be applied at each re-training step. Our goal will be to find one or more viable strategies to successfully reduce the impact of compound exposure bias, to improve recommendation quality and user well-being and satisfaction. The following research questions will be targeted:

RQ4. Which bias mitigation technique is most effective, in terms of robustness to compound exposure bias and recommendation effectiveness retention?
RQ5. Is combining multiple bias mitigation techniques beneficial to RS robustness to compound exposure bias?
RQ6. What is the effect of compound exposure bias mitigation on the time users spend on social media?

5 Conclusions

In conclusion, this proposal seeks to address the critical and pervasive issue of compound exposure bias within RS, particularly as it affects social media platforms. By introducing and rigorously studying the self-reinforcing nature of exposure bias, we aim to quantify its long-term impact on recommendation accuracy, diversity, and user engagement. Our proposed two-phase approach will first assess the impact of compound exposure bias on existing RS models, by means of a simulation framework based on LLM agents, and then evaluate possible mitigation strategies, such as debiasing, re-ranking, and fairness-constrained learning. The ultimate goal is to improve the robustness and fairness of RS, while mitigating harmful outcomes such as reduced content diversity and problematic user behaviours. By offering a deeper understanding and practical solutions to compound exposure bias, this research is expected to provide a meaningful contribution to the development of ethically sound, sustainable recommendation practices in social media platforms. The proposal's scientific soundness, comprehensiveness and feasibility are to be discussed during the Doctoral Consortium, together with potential additional user studies to better assess the impact of compound exposure bias on consumption habits.

References

1. Adomavicius, G., Jannach, D., Leitner, S., Zhang, J.: Understanding Longitudinal Dynamics of Recommender Systems with Agent-Based Modeling and Simulation (Aug 2021). https://doi.org/10.48550/arXiv.2108.11068

2. Aher, G.V., Arriaga, R.I., Kalai, A.T.: Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies. In: Proceedings of the 40th International Conference on Machine Learning. pp. 337–371. PMLR (Jul 2023), https://proceedings.mlr.press/v202/aher23a.html
3. Chaney, A.J.B., Stewart, B.M., Engelhardt, B.E.: How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 224–232. RecSys '18, Association for Computing Machinery, New York, NY, USA (Sep 2018). https://doi.org/10.1145/3240323.3240370
4. Chang, M.H., Harrington, J.E.: Chapter 26 Agent-Based Models of Organizations. In: Tesfatsion, L., Judd, K.L. (eds.) Handbook of Computational Economics, vol. 2, pp. 1273–1337. Elsevier (Jan 2006). https://doi.org/10.1016/S1574-0021(05)02026-5
5. Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., He, X.: Bias and Debias in Recommender System: A Survey and Future Directions. ACM Trans. Inf. Syst. **41**(3), 67:1–67:39 (Feb 2023). https://doi.org/10.1145/3564284
6. Ferraro, A., Jannach, D., Serra, X.: Exploring Longitudinal Effects of Session-based Recommendations. In: Proceedings of the 14th ACM Conference on Recommender Systems. pp. 474–479. RecSys '20, Association for Computing Machinery, New York, NY, USA (Sep 2020). https://doi.org/10.1145/3383313.3412213
7. Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., Li, Y.: Large language models empowered agent-based modeling and simulation: A survey and perspectives. Humanities and Social Sciences Communications **11**(1), 1–24 (Sep 2024). https://doi.org/10.1057/s41599-024-03611-3
8. Haussler, D., Warmuth, M.: The Probably Approximately Correct (PAC) and Other Learning Models. In: The Mathematics Of Generalization. CRC Press (1995)
9. Jannach, D., Lerche, L., Kamehkhosh, I., Jugovac, M.: What recommenders recommend: An analysis of recommendation biases and possible countermeasures. User Modeling and User-Adapted Interaction **25**(5), 427–491 (Dec 2015). https://doi.org/10.1007/s11257-015-9165-3
10. Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately Interpreting Clickthrough Data as Implicit Feedback. SIGIR Forum **51**(1), 4–11 (Aug 2017). https://doi.org/10.1145/3130332.3130334
11. Klimashevskaia, A., Jannach, D., Elahi, M., Trattner, C.: A survey on popularity bias in recommender systems. User Modeling and User-Adapted Interaction (Jul 2024). https://doi.org/10.1007/s11257-024-09406-0
12. Liu, D., Cheng, P., Dong, Z., He, X., Pan, W., Ming, Z.: A General Knowledge Distillation Framework for Counterfactual Recommendation via Uniform Data. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 831–840. SIGIR '20, Association for Computing Machinery, New York, NY, USA (Jul 2020). https://doi.org/10.1145/3397271.3401083
13. Marlin, B., Zemel, R.S., Roweis, S., Slaney, M.: Collaborative Filtering and the Missing at Random Assumption (Jun 2012). https://doi.org/10.48550/arXiv.1206.5267
14. Mujica, A.L., Crowell, C.R., Villano, M.A., Uddin, K.M.: ADDICTION BY DESIGN: Some Dimensions and Challenges of Excessive Social Media Use. Medical Research Archives **10**(2) (Feb 2022). https://doi.org/10.18103/mra.v10i2.2677
15. Wang, Y., Ma, W., Zhang, M., Liu, Y., Ma, S.: A Survey on the Fairness of Recommender Systems. ACM Trans. Inf. Syst. **41**(3), 52:1–52:43 (Feb 2023). https://doi.org/10.1145/3547333

16. Yang, L., Cui, Y., Xuan, Y., Wang, C., Belongie, S., Estrin, D.: Unbiased offline recommender evaluation for missing-not-at-random implicit feedback. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 279–287. RecSys '18, Association for Computing Machinery, New York, NY, USA (Sep 2018). https://doi.org/10.1145/3240323.3240355
17. Zakon, A.: Optimized for Addiction: Extending Product Liability Concepts to Defectively Designed Social Media Algorithms and Overcoming the Communications Decency Act. Wisconsin Law Review **2020**, 1107 (2020). https://doi.org/10.2139/ssrn.3682048
18. Zhang, J., Adomavicius, G., Gupta, A., Ketter, W.: Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework. Information Systems Research **31**(1), 76–101 (Mar 2020). https://doi.org/10.1287/isre.2019.0876
19. Zhou, M., Zhang, J., Adomavicius, G.: Longitudinal Impact of Preference Biases on Recommender Systems' Performance. Information Systems Research (Nov 2023). https://doi.org/10.1287/isre.2021.0133