LambdaFair: a Fair and Effective LambdaMART

Federico Marcuzzi[1,*], Claudio Lucchese[1] and Salvatore Orlando[1]

[1]Università Ca' Foscari Venezia, Via Torino, 155, 30170 Mestre, Venezia VE, Italy

Keywords
information retrieval, learning to rank, fairness

Traditional learning algorithms are known to introduce or exacerbate biases in data, leading to discrimination against individuals from protected groups (e.g., minorities or socially disadvantaged groups). This phenomenon extends to Information Retrieval (IR) systems, where biases in the data may translate into ranking systems that discriminate protected groups [1]. Ensuring fair treatment of protected individuals has become a pivotal challenge in IR to prevent discrimination; however, ranking effectiveness remains a crucial requirement for IR systems. As a consequence, providing fair ranking systems without significantly compromising their effectiveness poses a substantial challenge.

In this regard, our work introduces LambdaFair, a LambdaMART-based [2] in-processing method to train fairness-aware ranking models by simultaneously optimizing fairness and effectiveness. LambdaFair jointly optimizes fairness and effectiveness through a convex combination of NDCG [3] (effectiveness) and rND [4] (fairness).

NDCG (*Normalized Discounted Cumulative Gain*) is a well-known IR metric for evaluating ranking effectiveness. The optimum is achieved when documents with higher relevance to the query are ranked higher in the list. rND (*Normalized Discounted Difference*), instead, measures the fairness of a ranked list in terms of statistical parity. rND is optimal when each ranking prefix has the same proportion of protected items as the entire ranking.

To optimize these different-purpose metrics jointly, we designed three variants of LambdaFair, each balancing effectiveness and fairness at a different level. The first variant, named rND+, is fairness-oriented. When there is a conflict between optimizing rND and NDCG, i.e., when maximizing one metric forces a sub-optimal solution of the other, priority is given to fairness. The second variant, NDCG+, is effectiveness-oriented. NDCG+ is symmetric to rND+; in case of conflict, it favors NDCG over rND. The last variant, ΔrND, balances the two metrics by looking for a sub-optimal solution.

We compared LambdaFair with the state-of-the-art baseline PL-Rank-3 [5] and LambdaMART [2] on real-world publicly available datasets: MSLR-30K [6] and Statlog (German Credit Data) [7]. Our empirical results demonstrate that LambdaFair improves ranking fairness in terms of statistical party (rND) while maintaining competitive ranking effectiveness (NDCG).

IIR2024: 14th Italian Information Retrieval Workshop, 5th - 6th September 2024, Udine, Italy
*Corresponding author.
federico.marcuzzi@unive.it (F. Marcuzzi); claudio.lucchese@unive.it (C. Lucchese); orlando@unive.it (S. Orlando)
0000-0002-8141-8294 (F. Marcuzzi); 0000-0002-2545-0425 (C. Lucchese); 0000-0002-4155-9797 (S. Orlando)

Acknowledgement

This study was funded by the European Union - NextGenerationEU, in the framework of the iNEST - Interconnected Nord-Est Innovation Ecosystem (iNEST ECS_00000043 – CUP H43C22000540006). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

[1] M. Zehlike, K. Yang, J. Stoyanovich, Fairness in ranking, part II: learning-to-rank and recommender systems, ACM Comput. Surv. 55 (2023) 117:1–117:41. URL: https://doi.org/10.1145/3533380. doi:10.1145/3533380.

[2] C. J. Burges, From ranknet to lambdarank to lambdamart: An overview, Learning 11 (2010) 81.

[3] K. Järvelin, J. Kekäläinen, Cumulated gain-based evaluation of IR techniques, ACM Trans. Inf. Syst. 20 (2002) 422–446. URL: http://doi.acm.org/10.1145/582415.582418. doi:10.1145/582415.582418.

[4] K. Yang, J. Stoyanovich, Measuring fairness in ranked outputs, in: Proceedings of the 29th International Conference on Scientific and Statistical Database Management, Chicago, IL, USA, June 27-29, 2017, ACM, 2017, pp. 22:1–22:6. URL: https://doi.org/10.1145/3085504.3085526. doi:10.1145/3085504.3085526.

[5] H. Oosterhuis, Learning-to-rank at the speed of sampling: Plackett-luce gradient estimation with minimal computational complexity, in: E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, G. Kazai (Eds.), SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022, ACM, 2022, pp. 2266–2271. URL: https://doi.org/10.1145/3477495.3531842. doi:10.1145/3477495.3531842.

[6] T. Qin, T. Liu, Introducing LETOR 4.0 datasets, CoRR abs/1306.2597 (2013). URL: http://arxiv.org/abs/1306.2597.

[7] H. Hofmann, Statlog (German Credit Data), UCI Machine Learning Repository, 1994. DOI: https://doi.org/10.24432/C5NC77.