Modeling and Correcting Bias in Sequential Evaluation

JINGYAN WANG and ASHWIN PANANJADY, Georgia Institute of Technology, USA

We consider the problem of sequential evaluation, in which an evaluator observes candidates in a sequence and assigns scores to these candidates in an online, irrevocable fashion. Sequential bias refers to dependencies between the evaluation outcome and the order in which the candidates appear, and extensive empirical studies have established its existence in many applications. Motivated by the psychology literature, we propose a natural model for the evaluator's rating process that captures the lack of calibration inherent to such a task. We conduct crowdsourcing experiments to demonstrate various facets of our model, propose a near-linear time, online algorithm for bias correction, and show that it is near-optimal in two canonical ranking metrics.

Formally we consider a set of n candidates. We let γ^* denote their true global ranking, so that candidate $i \in [n]$ is ranked $\gamma^*(i)$-th (the higher the better). Recall that the candidates arrive sequentially: we denote by $r_t^* := \{|\{i \in [t] : \gamma^*(i) \leq \gamma^*(t)\}|\}$ the rank of candidate t among the first t candidates that have arrived (the higher the better). We model the evaluator's rating of candidate t as $y_t := x(t, r_t^*) + \epsilon_t$, where ϵ_t is i.i.d. noise and x is a bivariate function. The dependence of x on r_t^* captures the fact that evaluation is performed by explicit or implicit comparison of candidates to one another, and the dependence on t allows for the nature of this comparison to change with time as more candidates arrive. A natural assumption on x is a monotonicity constraint, namely $x(t,r) < x(t,r')$ for every $r < r'$. We demonstrate various facets of this model using crowdsourcing experiments using an online evaluation task. We then specialize our model to a parametric model $x(t, r_t^*) = \frac{r_t^*}{t+1}$, and show that this model theoretically aligns with the optimal behavior of the evaluator if the evaluator aims to achieve "perfect calibration" for their ratings in an online fashion.

We demonstrate that the de-facto natural baseline — which ranks candidates naively based on their ratings — is inconsistent even when the ratings are generated noiselessly based on the parametric model. This observation motivates the need for bias correction, which we pose as the statistical inference problem of estimating the true ranking γ^* from noisy ratings $\{y_t\}_{t=1}^n$ generated according to the parametric model. We propose a near-linear time algorithm that "reconstructs" an estimate of the global ranking in an online fashion, and prove guarantees on its performance in two natural ranking metrics — the Spearmans's footrule $\|\gamma - \gamma^*\|_1$, and the worst error incurred across all positions $\|\gamma - \gamma^*\|_\infty$. We also show that our algorithm is information theoretically near-optimal, by showing nearly matching lower bounds in both metrics. Our simulations show that our algorithm significantly outperforms the de-facto baseline of ranking based on the evaluator's ratings themselves.

Our proofs control the subtle temporal evolution of ranking errors and make connections to inversion vectors from discrete mathematics; these may be of independent interest in other online ranking problems.

A full version of this paper can be found at https://arxiv.org/abs/2205.01607.

CCS Concepts: • **Mathematics of computing → Probability and statistics**; • **Computing methodologies → Machine learning**; • **Theory of computation → Design and analysis of algorithms**.

Additional Key Words and Phrases: sequential bias, ranking estimation, crowdsourcing, fairness

ACM Reference Format:

Jingyan Wang and Ashwin Pananjady. 2023. Modeling and Correcting Bias in Sequential Evaluation. In *Proceedings of the 24th ACM Conference on Economics and Computation (EC '23), July 9–12, 2023, London, United Kingdom.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3580507.3597747

EC '23, July 9–12, 2023, London, United Kingdom

ACM ISBN 979-8-4007-0104-7/23/07.
https://doi.org/10.1145/3580507.3597747