Multileave Gradient Descent for Fast Online Learning to Rank (Abstract)

Anne Schuth† a.g.schuth@uva.nl

Harrie Oosterhuis† harrie.oosterhuis@student.uva.nl

Shimon Whiteson‡ shimon.whiteson@cs.ox.ac.uk

Maarten de Rijke† derijke@uva.nl

† University of Amsterdam, Amsterdam, The Netherlands
‡ University of Oxford, Oxford, United Kingdom

1. INTRODUCTION AND METHOD

We summarize the findings of Schuth et al. [6]. Modern search engines base their rankings on combinations of dozens or even hundreds of features. Online learning to rank methods optimize combinations of features while interacting with users of a search engine. Clicks have proven to be a valuable source of information when interpreted as a preference between either rankings [4] or documents [3]. In particular, when clicks are interpreted using interleaved comparison methods, they can reliably infer preferences between a pair of rankers [1, 3]. *Dueling bandit gradient descent* (DBGD) [7] is an online learning to rank algorithm that learns from these interleaved comparisons. It uses the inferred preferences to estimate a gradient, which is followed to find a locally optimal ranker. At every learning step, DBGD estimates this gradient with respect to a *single* exploratory ranker and updates its solution if the exploratory ranker seems better. Exploring *more than one* ranker before updating towards a promising one could lead to finding a better ranker using fewer updates. However, when using interleaved comparisons, this would be too costly, since it would require pairwise comparisons involving users between all exploratory rankers. Instead, we propose to learn from *multileaved comparison methods* [5] that allow for comparisons of multiple rankers at once, using a single user interaction. In this way, our proposed method, *multileave gradient descent* (MGD), aims to speed up online learning to rank. We propose two variants of MGD that differ in how they estimate the gradient. In MGD *winner takes all* (MGD-W), the gradient is estimated using one ranker randomly sampled from those who won the multileaved comparison. In MGD *mean winner* (MGD-M), the gradient is estimated using the mean of all winning rankers. Our contributions are: 1) two approaches, MGD-W and MGD-M, to using multileaved comparison outcomes in an online learning to rank method; and 2) extensive empirical validation of our new methods via experiments on nine learning to rank datasets, showing that MGD-W and MGD-M outperform the state of the art in online learning to rank.

2. RESULTS AND CONCLUSIONS

We run experiments on nine learning to rank datasets and use the setup described by Hofmann et al. [2] to simulate user interactions. Our empirical results, based on extensive experiments encompassing 86M user interactions, show that MGD dramatically improves over the DBGD baseline. In particular, when the noise in user feedback increases, we find that MGD is capable of learning better rankers much faster than the baseline does. Figure 1 shows the results over a larger number of queries. The graph shows that even after 100,000 queries DBGD has not converged, and MGD still performs better.

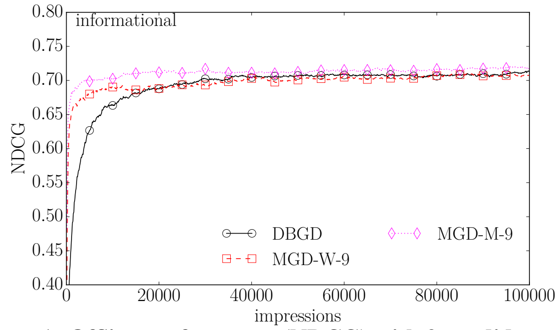

Figure 1: Offline performance (NDCG) with 9 candidates compared to DBGD on *NP2003* dataset.

Generally, after 1,000 query impressions with *noisy* feedback, MGD performs almost on par with DBGD trained on feedback *without any noise*.

An important implication of our results is that orders of magnitude less user interaction data is required to find good rankers when multileaved comparisons are used as feedback mechanism for online learning to rank. This results in far fewer users being exposed to inferior rankers and it allows search engines to adapt faster to changes in user preferences.

REFERENCES

[1] O. Chapelle, T. Joachims, F. Radlinski, and Y. Yue. Large-scale validation and analysis of interleaved search evaluation. *ACM Trans. Inf. Syst.*, 30(1), 2012.

[2] K. Hofmann, A. Schuth, S. Whiteson, and M. de Rijke. Reusing historical interaction data for faster online learning to rank for IR. In *WSDM '13*. ACM, 2013.

[3] T. Joachims. Optimizing search engines using clickthrough data. In *KDD '02*. ACM, 2002.

[4] F. Radlinski, M. Kurup, and T. Joachims. How does clickthrough data reflect retrieval quality? In *CIKM '08*. ACM, 2008.

[5] A. Schuth, F. Sietsma, S. Whiteson, D. Lefortier, and M. de Rijke. Multileaved comparisons for fast online evaluation. In *CIKM '14*, pages 71–80. ACM, Nov. 2014.

[6] A. Schuth, H. Oosterhuis, S. Whiteson, and M. de Rijke. Multileave gradient descent for fast online learning to rank. In *WSDM '16*, 2016.

[7] Y. Yue and T. Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *ICML '09*, 2009.