Optimizing Ranking Models in an Online Setting

Harrie Oosterhuis and Maarten de Rijke

April 16, 2019

University of Amsterdam

oosterhuis@uva.nl, derijke@uva.nl

https://staff.fnwi.uva.nl/h.r.oosterhuis

Introduction

Learning to Rank in Information Retrieval

Learning to rank enables the optimization of ranking systems.

Traditionally learning to rank uses **annotated datasets**:

- **Relevance annotations** for query-document pairs provided by **human judges**.

Learning to Rank in Information Retrieval

Learning to rank enables the optimization of ranking systems.

Traditionally learning to rank uses **annotated datasets**:

- **Relevance annotations** for query-document pairs provided by **human judges**.

There are **problems** with **annotated datasets**:

- **expensive** to make (Qin and Liu, 2013; Chapelle and Chang, 2011).
- **impossible** to create (Wang et al., 2016).
- **not necessarily aligned with actual user preferences** (Sanderson, 2010), i.e. annotators and users often disagree.

Online Learning to Rank

Online Learning to Rank

Online learning to Rank: **learn by interacting with users**.

Solves most of the problems of annotations:

- Interactions are **virtually free** if you have users.
- User **behaviour** gives **implicit feedback**.

Online Learning to Rank

Online learning to Rank: **learn by interacting with users**.

Solves most of the problems of annotations:

- Interactions are **virtually free** if you have users.
- User **behaviour** gives **implicit feedback**.

These methods have to handle:

- **Noise**: Users click for **unexpected reasons**.
- **Biases**: Interactions are affected by **position and selection bias**.

Dueling Bandit Gradient Descent

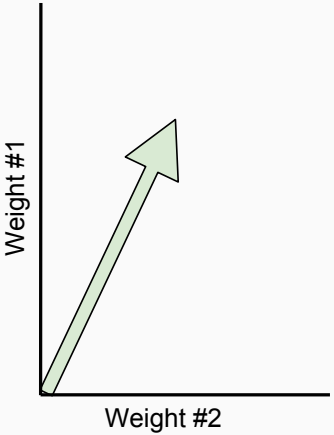
Dueling Bandit Gradient Descent: Introduction

Introduced by Yue and Joachims (2009) as the **first online learning to rank** method.

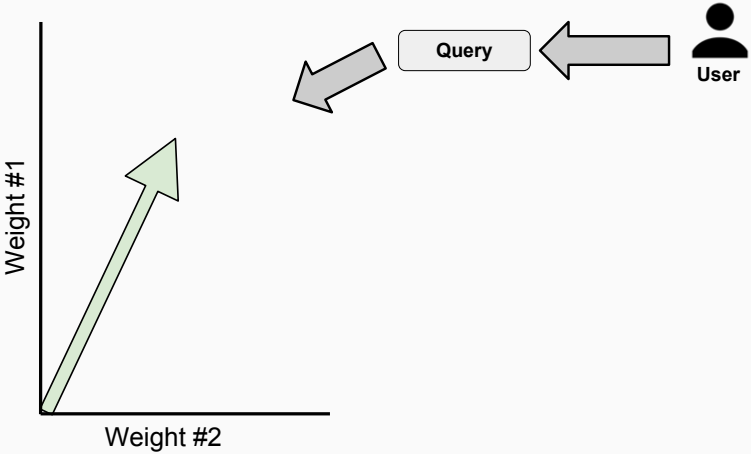
Intuition:

- **Interleaving** can compare rankers from **user interactions**.
- By **sampling model variants** and **comparing** them with **interleaving**, the *gradient* of a model w.r.t. user satisfaction can be **estimated**.

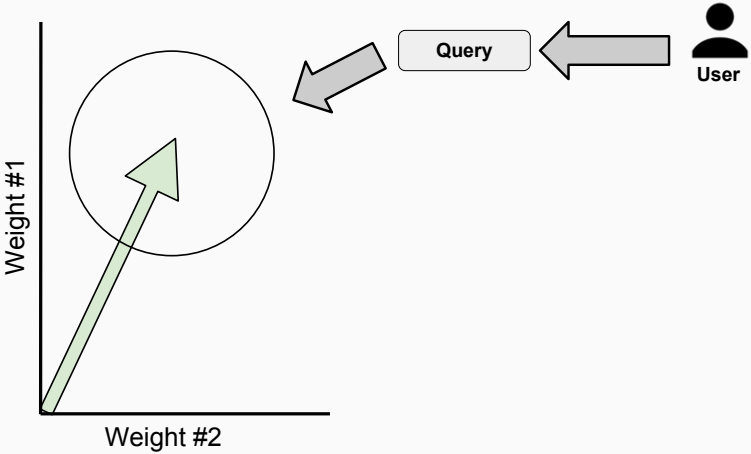
Dueling Bandit Gradient Descent: Visualization



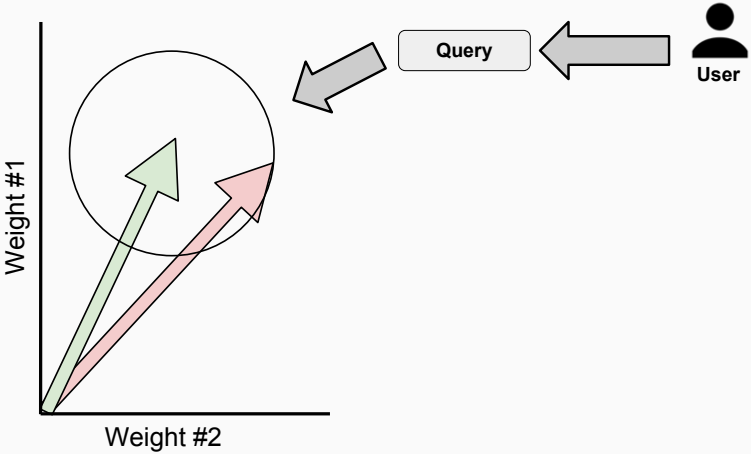
Dueling Bandit Gradient Descent: Visualization



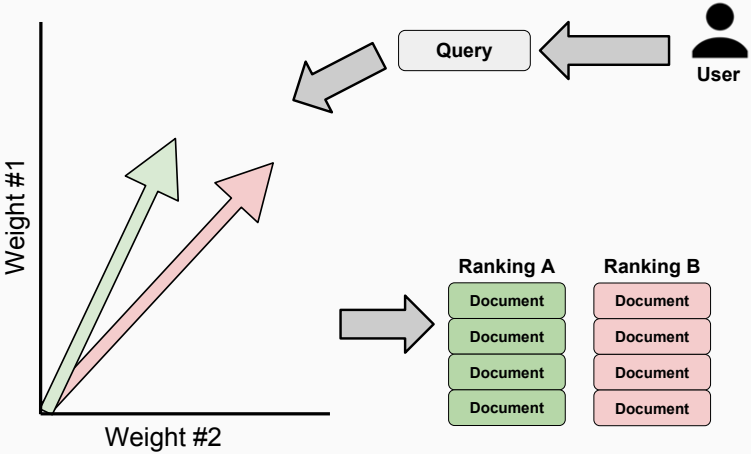
Dueling Bandit Gradient Descent: Visualization



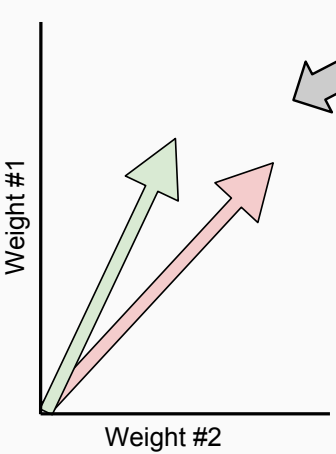
Dueling Bandit Gradient Descent: Visualization

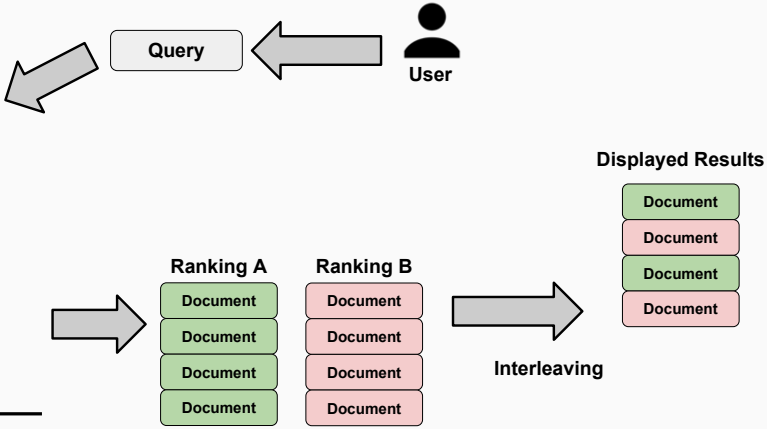


Dueling Bandit Gradient Descent: Visualization

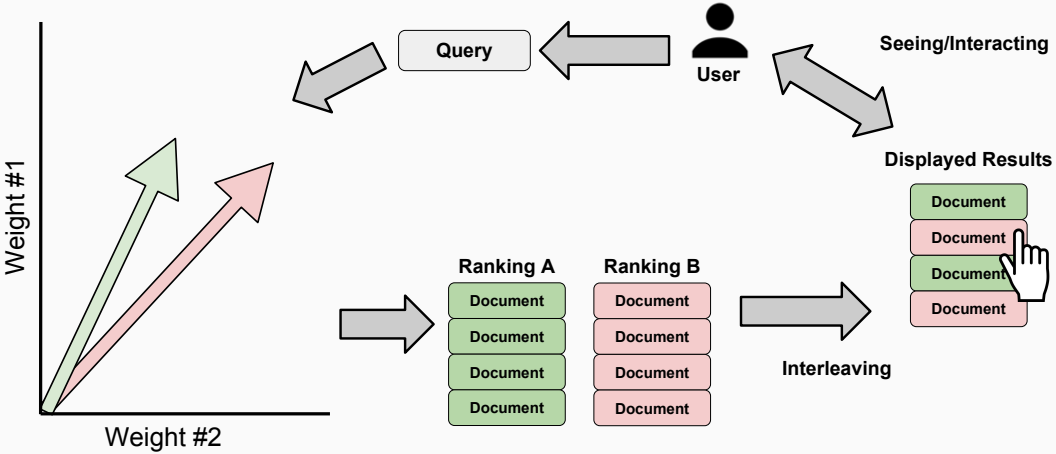


Dueling Bandit Gradient Descent: Visualization

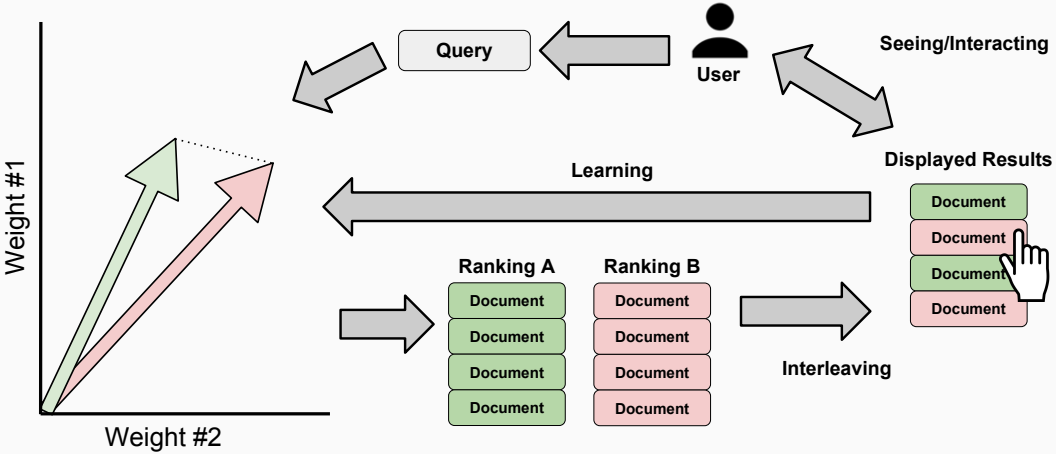




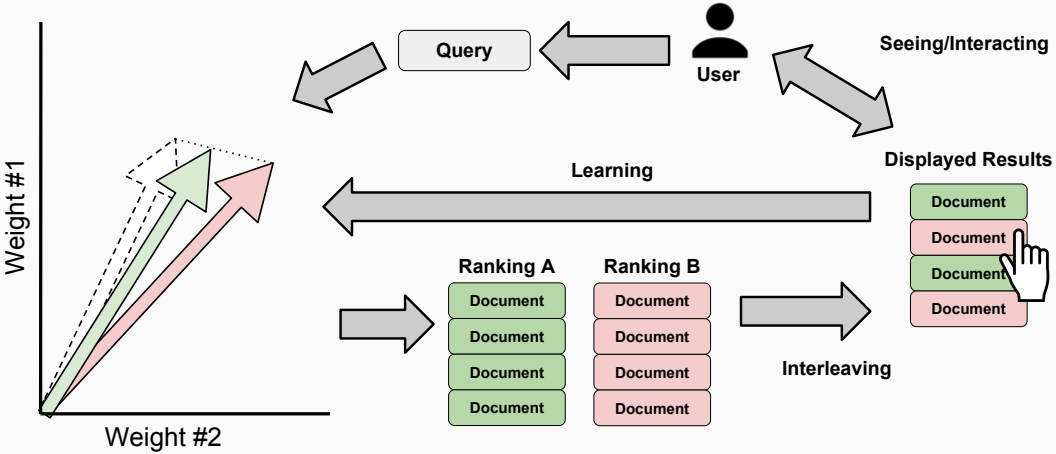
Dueling Bandit Gradient Descent: Visualization



Dueling Bandit Gradient Descent: Visualization



Dueling Bandit Gradient Descent: Visualization



Dueling Bandit Gradient Descent: Properties

Basis of the online learning to rank **field**,
virtually all existing methods are **extensions** of this algorithm
(Schuth et al., 2016; Hofmann et al., 2013; Zhao and King, 2016; Wang et al., 2018a).

Problems with Dueling Bandit Gradient Descent:

- A **considerable gap** between DBGD and *best possible* performance.
- Ineffective at optimizing **non-linear models**.

Dueling Bandit Gradient Descent: Properties

Basis of the online learning to rank **field**,
virtually all existing methods are **extensions** of this algorithm
(Schuth et al., 2016; Hofmann et al., 2013; Zhao and King, 2016; Wang et al., 2018a).

Problems with Dueling Bandit Gradient Descent:

- A **considerable gap** between DBGD and *best possible* performance.
- Ineffective at optimizing **non-linear models**.

Advantage of Dueling Bandit Gradient Descent:

- Has a **theoretical foundation** based on proven **regret bounds**.

Pairwise Differentiable Gradient Descent

Pairwise Differentiable Gradient Descent

We recently introduced **Pairwise Differentiable Gradient Descent** (Oosterhuis and de Rijke, 2018).

Intuition:

- A **pairwise** approach can be made **unbiased** while being **differentiable**.

Plackett Luce Model

Pairwise Differentiable Gradient Descent optimizes a **Plackett Luce** ranking model, this models a **probabilistic distribution over documents**.

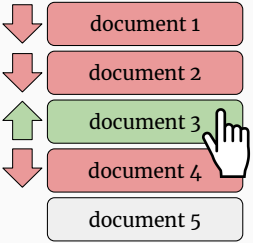
With the ranking scoring model $f(\mathbf{d})$ the distribution is:

$$P(d|f, D) = \frac{\exp^{f(\mathbf{d})}}{\sum_{d' \in D} \exp^{f(\mathbf{d}')}} \tag{1}$$

Unlike DBGD, **confidence is explicitly modelled**.

Bias in Pairwise Inference

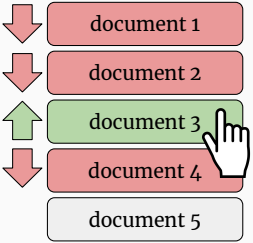
Similar to existing pairwise methods (Oosterhuis and de Rijke, 2017; Joachims, 2002), Pairwise Differentiable Gradient Descent infers **pairwise document preferences from user clicks**:



This **inference** is based on a **cascading assumption**.

Bias in Pairwise Inference

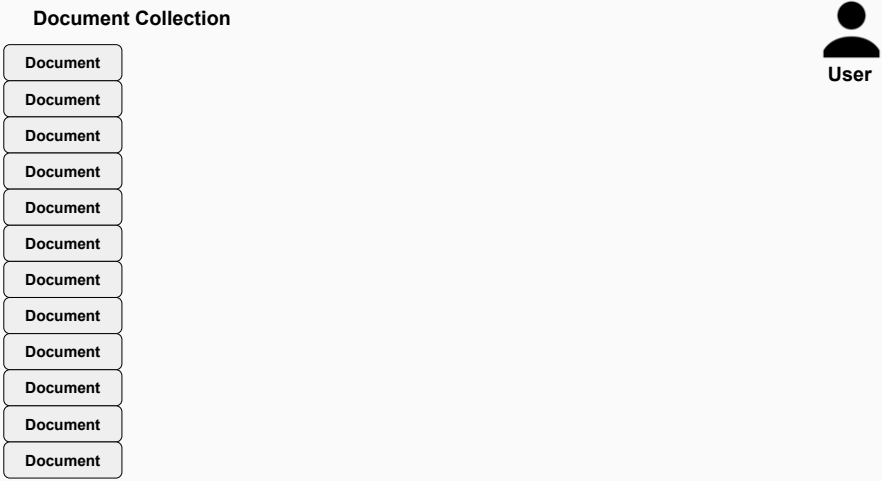
Similar to existing pairwise methods (Oosterhuis and de Rijke, 2017; Joachims, 2002), Pairwise Differentiable Gradient Descent infers **pairwise document preferences from user clicks**:



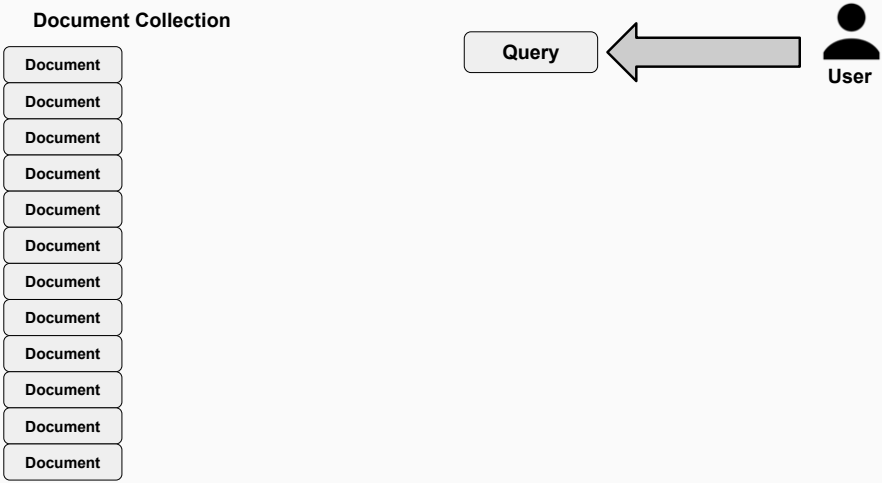
This **inference** is based on a **cascading assumption**.

PDGD weighs inferred preferences to account for **position/selection bias**.

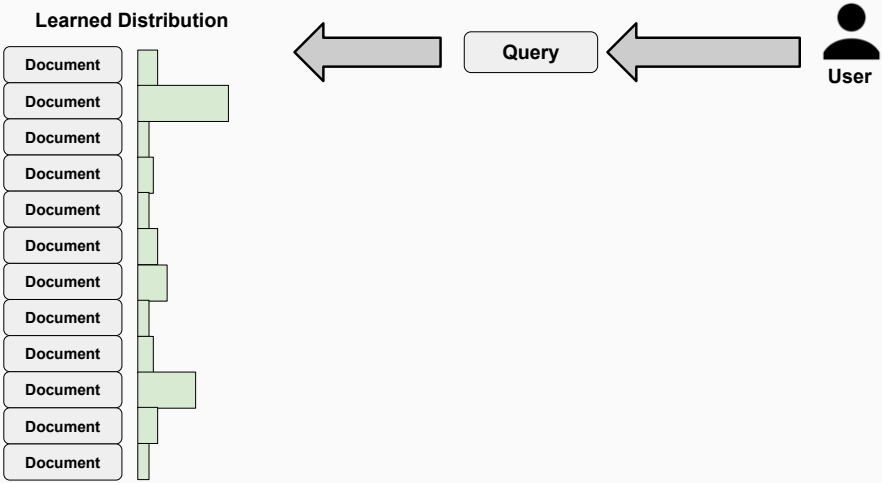
Pairwise Differentiable Gradient Descent: Visualization



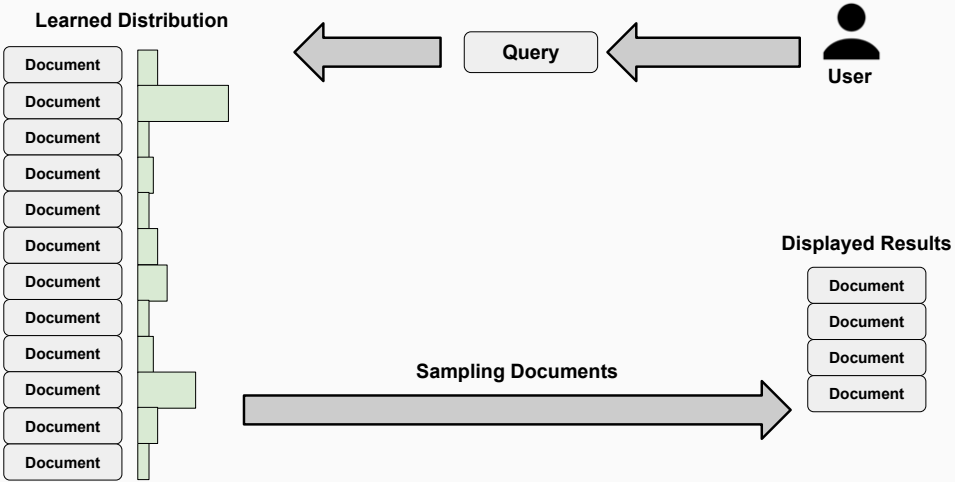
Pairwise Differentiable Gradient Descent: Visualization



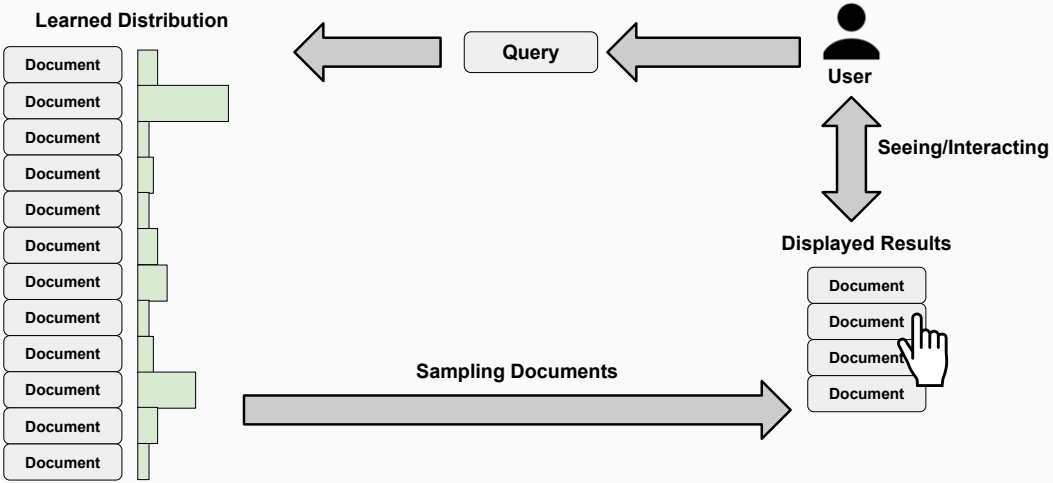
Pairwise Differentiable Gradient Descent: Visualization



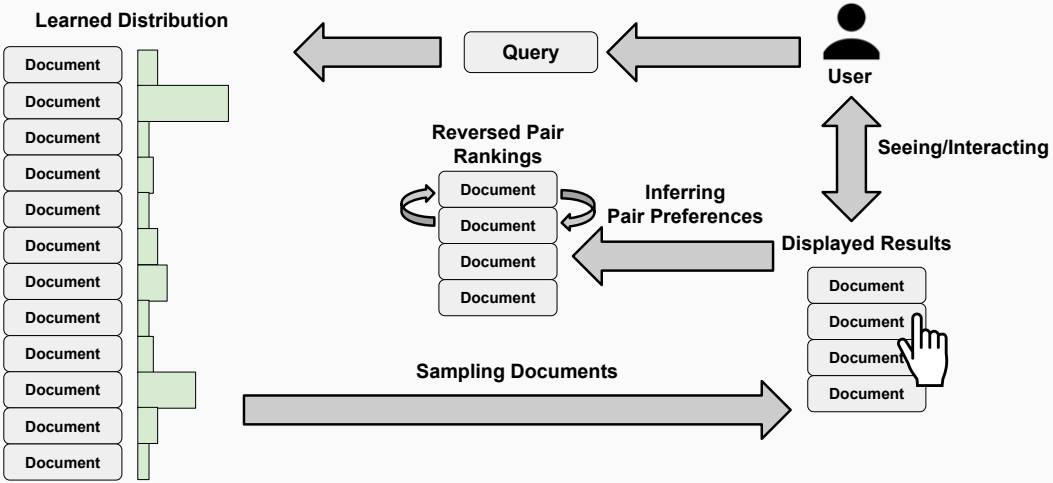
Pairwise Differentiable Gradient Descent: Visualization



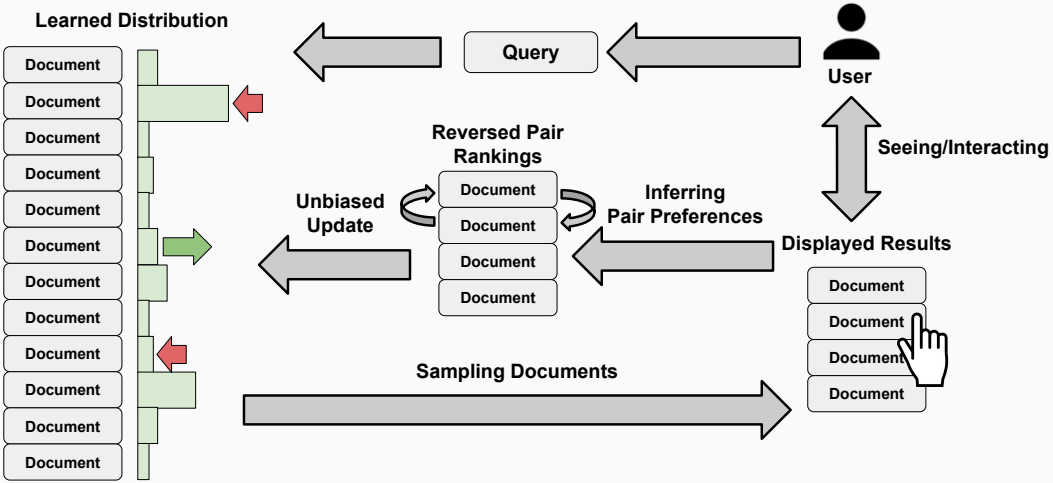
Pairwise Differentiable Gradient Descent: Visualization



Pairwise Differentiable Gradient Descent: Visualization



Pairwise Differentiable Gradient Descent: Visualization



Pairwise Differentiable Gradient Descent: Properties

Pairwise Differentiable Gradient Descent does **not rely** on **online evaluation** or the **sampling** of models.

Advantages claimed by previous work (Oosterhuis and de Rijke, 2018):

- Much better point of convergence i.e. **long-term performance**.
- Much faster learning i.e. **short term performance**.
- Computationally much faster.

Limitations of Existing Results

Limitations of Previous Existing Results

Comparison based on **previous work**:

- Pairwise Differentiable Gradient Descent **outperforms** DBGD.
- Dueling Bandit Gradient Descent has **proven sublinear regret bounds**.

Limitations of Previous Existing Results

Comparison based on **previous work**:

- Pairwise Differentiable Gradient Descent **outperforms** DBGD.
- Dueling Bandit Gradient Descent has **proven sublinear regret bounds**.

Problems with the current state of affairs:

- Past comparisons are based on **low-noise, cascading click models**,

Limitations of Previous Existing Results

Comparison based on **previous work**:

- Pairwise Differentiable Gradient Descent **outperforms** DBGD.
- Dueling Bandit Gradient Descent has **proven sublinear regret bounds**.

Problems with the current state of affairs:

- Past comparisons are based on **low-noise, cascading click models**, Pairwise Differentiable Gradient Descent **assumes cascading behaviour!**

Limitations of Previous Existing Results

Comparison based on **previous work**:

- Pairwise Differentiable Gradient Descent **outperforms** DBGD.
- Dueling Bandit Gradient Descent has **proven sublinear regret bounds**.

Problems with the current state of affairs:

- Past comparisons are based on **low-noise, cascading click models**, Pairwise Differentiable Gradient Descent **assumes cascading behaviour!**
- There is a **conflict** between the **proven regret bounds** of Dueling Bandit Gradient Descent and its **observed performance**.

Our contribution

First, we **critically look** at the regret bounds of Dueling Bandit Gradient Descent:

- We **prove** that its **assumptions cannot be true** for standard ranking models.

Our contribution

First, we **critically look** at the regret bounds of Dueling Bandit Gradient Descent:

- We **prove** that its **assumptions cannot be true** for standard ranking models.

Then we reproduce the comparison under different conditions:

- **Both cascading** and **non-cascading** click behaviour.
- Simulated conditions ranging from **ideal to extremely difficult**.
  We simulate *near-random* behaviour.

Experimental Results

Experimental Setup

Simulations based on **largest available industry datasets**:

- MSLR-Web10k, Yahoo Webscope, Istella.

Experimental Setup

Simulations based on **largest available industry datasets**:

- MSLR-Web10k, Yahoo Webscope, Istella.

User behaviour simulated using **cascading** and **non-cascading click models**.
Simulated behaviour ranging from:

- **ideal:** no noise, no position bias,
- **extremely difficult:** mostly noise, very high position bias.

Experimental Setup

Simulations based on **largest available industry datasets**:

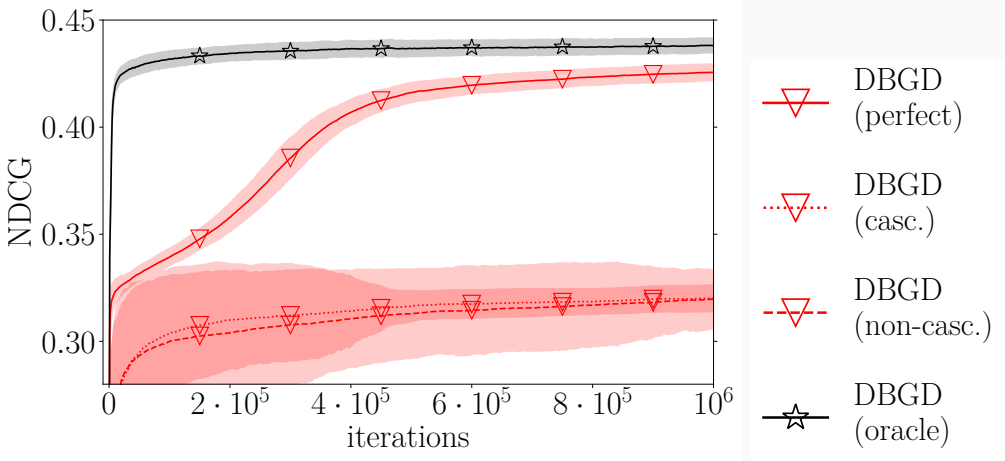
- MSLR-Web10k, Yahoo Webscope, Istella.

User behaviour simulated using **cascading** and **non-cascading click models**.
Simulated behaviour ranging from:

- **ideal:** no noise, no position bias,
- **extremely difficult:** mostly noise, very high position bias.

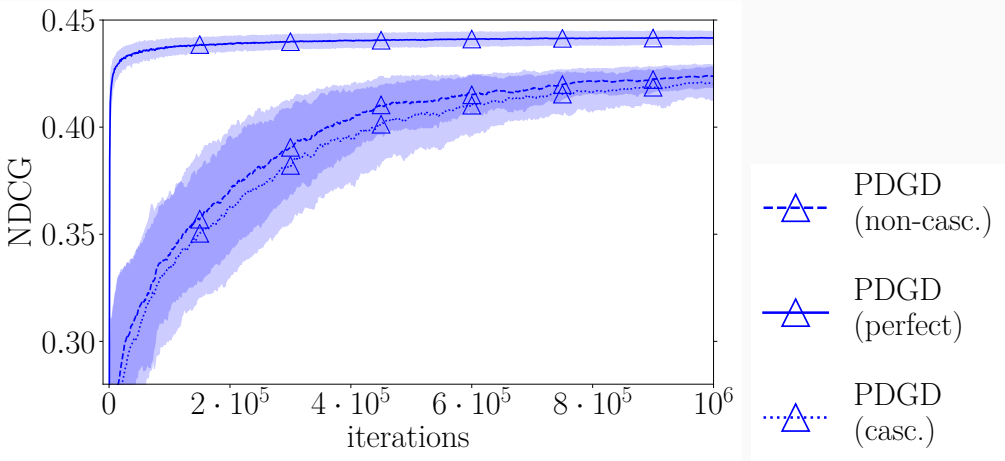
Dueling Bandit Gradient Descent with an **oracle instead of interleaving**, to see the **maximum potential** of better interleaving methods.

Comparison: Dueling Bandit Gradient Descent



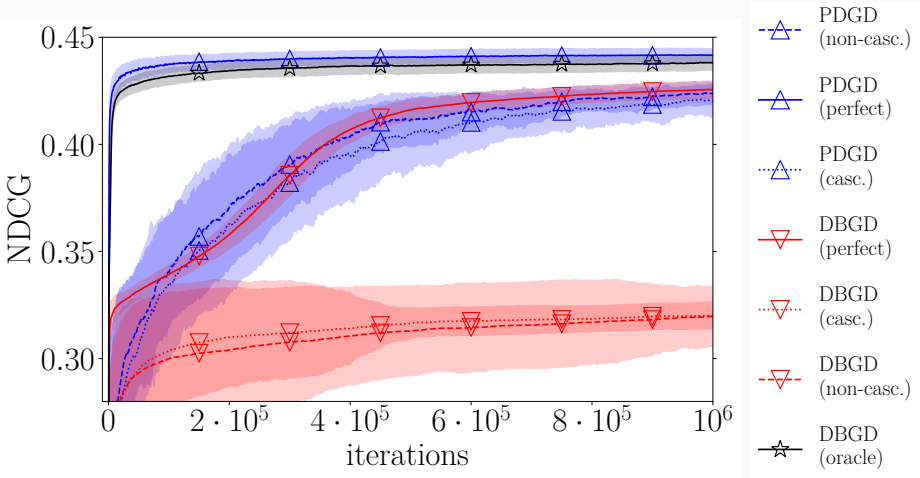
Results from simulations on the MSLR-WEB10k dataset.

Comparison: Pairwise Differentiable Gradient Descent



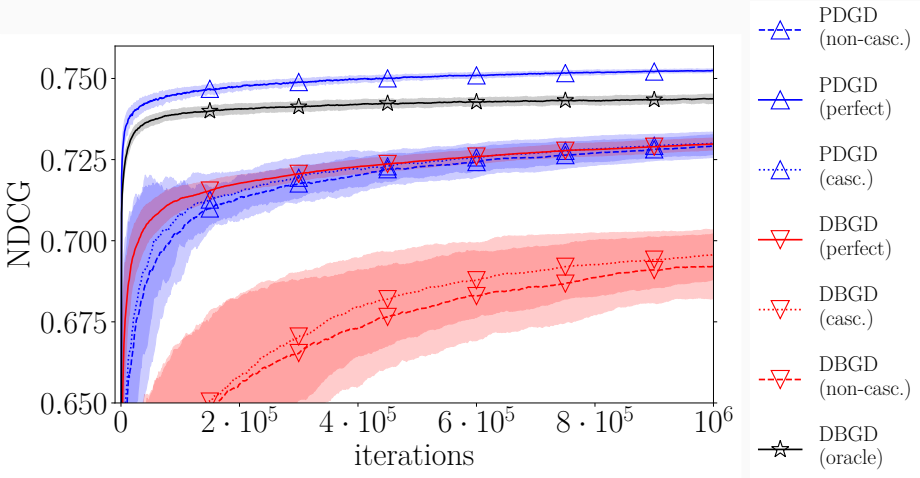
Results from simulations on the MSLR-WEB10k dataset.

Comparison: Complete



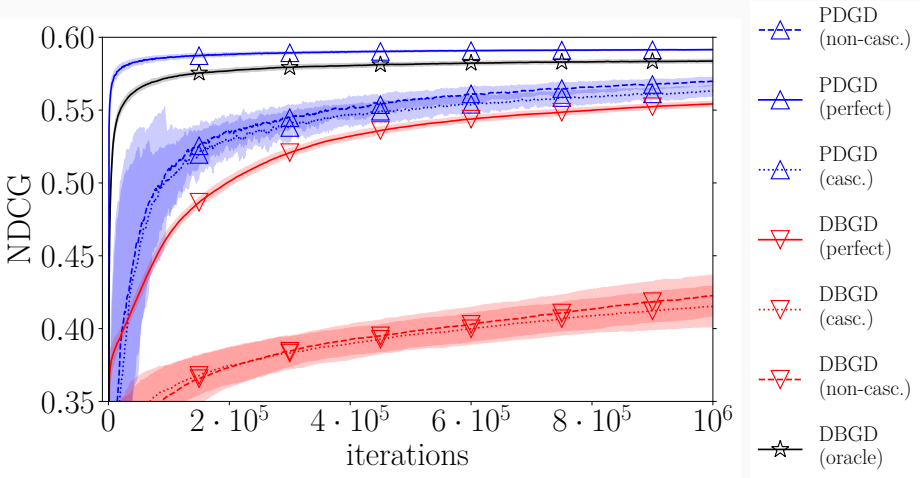
Results from simulations on the MSLR-WEB10k dataset.

Comparison: Complete



Results from simulations on the Yahoo Webscope dataset.

Comparison: Complete



Results from simulations on the Istella dataset.

Conclusion

Conclusion

In this reproducibility paper we have:

Conclusion

In this reproducibility paper we have:

- shown that an **existing proof** for regret bounds is **unsound**.

Conclusion

In this reproducibility paper we have:

- shown that an **existing proof** for regret bounds is **unsound**.
- **Reproduced a comparison** between Pairwise Differentiable Gradient Descent and Dueling Bandit Gradient Descent,

Conclusion

In this reproducibility paper we have:

- shown that an **existing proof** for regret bounds is **unsound**.
- **Reproduced a comparison** between Pairwise Differentiable Gradient Descent and Dueling Bandit Gradient Descent, and **generalized their conclusions** from ideal circumstances to extremely difficult circumstances.

Conclusion

In this reproducibility paper we have:

- shown that an **existing proof** for regret bounds is **unsound**.
- **Reproduced a comparison** between Pairwise Differentiable Gradient Descent and Dueling Bandit Gradient Descent, and **generalized their conclusions** from ideal circumstances to extremely difficult circumstances.
- Shown that **under all experimental conditions** we could simulate, Pairwise Differentiable Gradient Descent **outperforms** previous methods by **large margins**.

Conclusion

In this reproducibility paper we have:

- shown that an **existing proof** for regret bounds is **unsound**.
- **Reproduced a comparison** between Pairwise Differentiable Gradient Descent and Dueling Bandit Gradient Descent, and **generalized their conclusions** from ideal circumstances to extremely difficult circumstances.
- Shown that **under all experimental conditions** we could simulate, Pairwise Differentiable Gradient Descent **outperforms** previous methods by **large margins**.

Please continue our work: `https://github.com/HarrieO/OnlineLearningToRank`

References i

Q. Ai, K. Bi, C. Luo, J. Guo, and W. B. Croft. Unbiased learning to rank with unbiased propensity estimation. *arXiv preprint arXiv:1804.05938*, 2018.

O. Chapelle and Y. Chang. Yahoo! Learning to Rank Challenge Overview. *Journal of Machine Learning Research*, 14:1–24, 2011.

K. Hofmann, A. Schuth, S. Whiteson, and M. de Rijke. Reusing historical interaction data for faster online learning to rank for ir. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 183–192. ACM, 2013.

T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.

T. Joachims, A. Swaminathan, and T. Schnabel. Unbiased learning-to-rank with biased feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 781–789. ACM, 2017.

References ii

D. Lefortier, P. Serdyukov, and M. de Rijke. Online exploration for detecting shifts in fresh intent. In *CIKM 2014: 23rd ACM Conference on Information and Knowledge Management*. ACM, November 2014.

H. Oosterhuis and M. de Rijke. Sensitive and scalable online evaluation with theoretical guarantees. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 77–86. ACM, 2017.

H. Oosterhuis and M. de Rijke. Differentiable unbiased online learning to rank. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1293–1302. ACM, 2018.

H. Oosterhuis and M. de Rijke. Optimizing ranking models in the online setting. In *ECIR*. ACM, 2019.

T. Qin and T.-Y. Liu. Introducing letor 4.0 datasets. *arXiv preprint arXiv:1306.2597*, 2013.

M. Sanderson. Test collection based evaluation of information retrieval systems. *Foundations and Trends in Information Retrieval*, 4(4):247–375, 2010.

References iii

A. Schuth, H. Oosterhuis, S. Whiteson, and M. de Rijke. Multileave gradient descent for fast online learning to rank. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 457–466. ACM, 2016.

H. Wang, R. Langley, S. Kim, E. McCord-Snook, and H. Wang. Efficient exploration of gradient space for online learning to rank. *arXiv preprint arXiv:1805.07317*, 2018a.

X. Wang, M. Bendersky, D. Metzler, and M. Najork. Learning to rank with selection bias in personal search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 115–124. ACM, 2016.

X. Wang, N. Golbandi, M. Bendersky, D. Metzler, and M. Najork. Position bias estimation for unbiased learning to rank in personal search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 610–618. ACM, 2018b.

Y. Yue and T. Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1201–1208. ACM, 2009.

References iv

T. Zhao and I. King. Constructing reliable gradient exploration for online learning to rank. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1643–1652. ACM, 2016.

Acknowledgments

All content represents the opinion of the author(s), which is not necessarily shared or endorsed by their employers and/or sponsors.