Online learning to rank for information retrieval: SIGIR 2016 tutorial
Grotov, A.; de Rijke, M.

Published in:
SIGIR'16

DOI:
10.1145/2911451.2914798

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Online Learning to Rank for Information Retrieval

SIGIR 2016 Tutorial

Artem Grotov  
University of Amsterdam  
Amsterdam, The Netherlands  
a.grotov@uva.nl

Maarten de Rijke  
University of Amsterdam  
Amsterdam, The Netherlands  
derijke@uva.nl

ABSTRACT
During the past 10–15 years offline learning to rank has had a tremendous influence on information retrieval, both scientifically and in practice. Recently, as the limitations of offline learning to rank for information retrieval have become apparent, there is increased attention for online learning to rank methods for information retrieval in the community. Such methods learn from user interactions rather than from a set of labeled data that is fully available for training up front.

Below we describe why we believe that the time is right for an intermediate-level tutorial on online learning to rank, the objectives of the proposed tutorial, its relevance, as well as more practical details, such as format, schedule and support materials.

Keywords
Online learning to rank; Bandit algorithms; Exploration vs. exploitation

1. INTRODUCTION
Today’s search engines have developed into complex systems that combines hundreds of ranking criteria with the aim of producing the optimal result list in response to users’ queries. For automatically tuning optimal combinations of large numbers of ranking criteria, learning to rank [22, LTR] has proved invaluable. For a given query, each document is represented by a feature vector. The features may be query dependent, document dependent or capture the relationship between the query and documents. The task of the learner is to find a model that combines these features such that, when this model is used to produce a ranking for an unseen query, user satisfaction is maximized.

Traditionally, learning to rank algorithms are trained in batch mode, on a complete dataset of query and document pairs with their associated manually created relevance labels. This setting has a number of disadvantages and is impractical in many cases. First, creating such datasets is expensive and therefore infeasible for smaller search engines, such as small web-store search engines [24]. Second, it may be impossible for experts to annotate documents, as in the case of personalized search [18]. Third, the relevance of documents to queries can change over time, like in a news search engine [7].

Online learning to rank addresses all of these issues by incrementally learning from user feedback in real time [34]. Online learning is closely related to active learning, incremental learning, and counterfactual learning. However, online learning is more difficult because the agent has to balance exploration and exploitation: actions with unknown performance have to be explored to learn better solutions [11].

There is a growing body of established methods for online learning to rank for information retrieval (see the schedule below for a broad range of examples). The time is right to organize and present this material to a broad audience of interested information retrieval researchers, whether junior or senior, whether academic or industrial. The online learning to rank methods available today have been proposed by different communities, in machine learning and information retrieval. A key aim of the tutorial is to bring these together and offer a unified perspective. To achieve this we illustrate the core and state of the art methods in online learning to rank, their theoretical foundations and real-world applications, as well as existing online learning algorithms that have not been used by information retrieval community so far.

We expect the tutorial to be useful for both academic and industrial researchers who either want to develop new online learning to rank methods, use them in their own research, or apply the methods described in the tutorial to improve search and recommendation systems.

2. OBJECTIVES
Online learning to rank from user interactions is fundamentally different from currently dominant supervised learning to rank approaches for information retrieval, where training data is assumed to be randomly sampled from some underlying distribution, and where absolute and reliable labels are provided by professional annotators [15]. When learning from user interactions, a system has no control over which queries it receives, it only receives feedback on the result lists it presents to users, and it has to present high quality result lists while learning, to satisfy user expectations.

Following Hofmann et al. [11], in this tutorial we formulate online learning to rank as a reinforcement learning problem, in which an agent, the search engine, learns from interactions with an environment, the user and their interactions, by trying out actions (e.g., returning a ranked list of items) and observing rewards (e.g., interpreting user feedback as absolute or relative feedback) in multiple rounds or discrete time steps; see Figure 1.

Particularly relevant to this tutorial are methods for tackling so-called contextual bandit problems (also known as bandits with side information) [1]. A contextual bandit problem is a special case of
The objectives of the tutorial are as follows:

- To describe existing online LTR algorithms in a unified way, i.e., using common notation and terminology, so that different models can easily be related to each other.
- Explain the importance of balancing exploration and exploitation in an online LTR setting.
- To explain how to analyse the performance of online LTR algorithms and why it is worth the effort.
- To present appropriate experimental and evaluation methodologies for online LTR in both synthetic and real world settings.
- To describe how to deploy online LTR algorithms in an industrial setting.
- To present online learning algorithms that have not been used yet for learning to rank and indicate how IR researchers might go about using them.
- To discuss future directions of research in online LTR.

These objectives are meant to give participants a thorough understanding of existing learning to rank for information retrieval methods and to present online learning methods that have so far not been applied to learning to rank, let alone to learning to rank for information retrieval.

3. DETAILED SCHEDULE

The tutorial will be organized in two halves of 90 minutes each, each mixing theory and experiment, with formal analyses of online learning to rank methods interleaved with discussions of code and of experimental outcomes. Part I is aimed at familiarizing participants with the key concepts and algorithms. In Part II we select a small number of topics to provide a more in-depth technical treatment.

**Part I**

**[10 minutes]** Introduction, aims and historical notes

Here we discuss the context in which online LTR is applied and the most important historical milestones in its development.

**[10 minutes]** LTR in IR.

- The task of ranking documents, its importance, relationship to other machine learning tasks, and the unique challenges of LTR [22].
- Current approaches to LTR and argue that they all have issues that have to be addressed, such as the cost of producing labelled data and the mismatch between manually curated labels and user intent [34].

**[15 minutes]** Online LTR: balancing exploration and exploitation.

- How does online LTR addresses the shortcomings of offline LTR [11, 34]?
- How does online LTR relate to, and differ from, other tasks such as learning from labelled data, active learning and learning from logged interactions [30]?
- We explain the importance of balancing exploration and exploitation [13].

**[5 minutes]** Introduction to bandits and reinforcement learning.

- An important formalism behind many online LTR methods: bandit algorithms [1, 20]
- Connection to reinforcement learning [11, 29]
- Illustrate the importance of formal analysis and present k-armed bandits [1], contextual bandits [20] and cascading bandits [19] as ways to formalize the online LTR setting.

**[10 minutes]** Online signals to learn from.

- Close connection with online and logged based IR Evaluation [9, 14], because in both settings one needs to make the connection between observed user feedback and the hidden quality of the system.
- How to use observed user feedback to train the ranking models? The observed user feedback includes clicks, absence of clicks, dwell times, abandonment and many other signals [17, 21].
- Signals can be interpreted in a number of ways, for example, clicks can be interpreted in the form of absolute click through rates, or as relative preferences between documents or retrieval systems or using click models [16].

**[20 minutes]** Dueling bandit gradient descent.

- Dueling bandit gradient descent [34, DBGD], one of the core methods used in online LTR. We present the theory behind this method and discuss under which conditions it is guaranteed to work.
- More advanced methods that build on DBGD such as Probabilistic Multileave Gradient Decent [23, 27] and DBGD with Candidate Preselection [12].
4. **TYPE OF SUPPORT MATERIALS TO BE SUPPLIED TO ATTENDEES**

- Slides
- Draft survey on online learning to rank for information retrieval [8]
- Code and data samples to follow experimental segments of the tutorial
- Lerot – experimental environment for online learning to rank [25]

**Acknowledgements**

This research was supported by Ahold, Amsterdam Data Science, the Bloomberg Research Grant program, the Dutch national program COMMIT, Elsevier, the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the ESF Research Network Program ELIAS, the Royal Dutch Academy of Sciences (KNAW) under the Elite Network Shifts project, the Microsoft Research Ph.D. program, the Netherlands Center for Genomics under project number 027.012.105, the Netherlands Institute for Sound and Vision, the Netherlands Organisation for Scientific Research (NWO) under project nrs 727.011.005, 612.001.116, HOR-11-10, 640.006.013, 612.066.930, CI-14-25, SH-322-15, 652.002.001, 612.001.551, the Yahoo Faculty Research and Engagement Program, and Yandex. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

**REFERENCES**

12. Katja Hofmann, Anne Schuth, Shimon Whiteson, and Maarten de Rijke. Reusing historical interaction data for faster online learning to rank for information retrieval. In


