Learning to Rank for Information Retrieval from User Interactions

Katja Hofmann
Microsoft Research
and
Shimon Whiteson
Intelligent Systems Lab Amsterdam, University of Amsterdam
and
Anne Schuth
Intelligent Systems Lab Amsterdam, University of Amsterdam
and
Maarten de Rijke
Intelligent Systems Lab Amsterdam, University of Amsterdam

In this article we give an overview of our recent work on online learning to rank for information retrieval (IR). This work addresses IR from a reinforcement learning (RL) point of view, with the aim to enable systems that can learn directly from interactions with their users. Learning directly from user interactions is difficult for several reasons. First, user interactions are hard to interpret as feedback for learning because it is usually biased and noisy. Second, the system can only observe feedback on actions (e.g., rankers, documents) actually shown to users, which results in an exploration–exploitation challenge. Third, the amount of feedback and therefore the quality of learning is limited by the number of user interactions, so it is important to use the observed data as effectively as possible. Here, we discuss our work on interpreting user feedback using probabilistic interleaved comparisons, and on learning to rank from noisy, relative feedback.

DOI: 10.1145/2591453.2591458 http://doi.acm.org/10.1145/2591453.2591458

1. INTRODUCTION

Information retrieval (IR) systems, such as web search engines, provide easy access to a vast and constantly growing source of information. The user submits a query, receives a ranked list of results, and follows the link or links to the most promising ones. To address the flood of data available on the web, today's web search engines have developed into complex systems that combine hundreds of information sources (features) with the goal of creating the best possible result rankings for all their users at all times.

Making a result ranking as useful as possible may be easy for some search tasks, but in many cases it depends on context, such as users' background knowledge, age, or location, or their specific search goals. And even though there is an enormous variety in the tasks and goals encountered in web search, web search engines are only the tip of the iceberg. More specialized systems are everywhere: search engines for company's intranets, local and national libraries, and users' personal documents (e.g., photos, emails, and music) all provide access to different, more or less specialized, document collections, and cater to different users with different search goals and expectations. The more we learn about how much context influences peoples' search behavior and goals, the more it becomes clear that addressing each of these settings individually is not feasible. Instead, we need to look for scalable methods that can learn good rankings without expensive tuning.

In our recent work, we have focused on developing methods for "self-learning search engines" that learn online, i.e., directly from natural interactions with their users. Such systems promise to be able to continuously adapt and improve their rankings to the specific setting they are deployed in, and continue to learn for as long as they are being used.

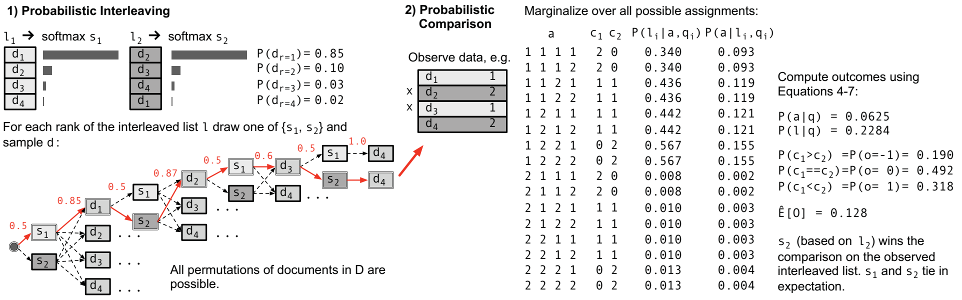
While data about user interactions with a search engine (e.g., clicks) are a natural by-product of normal search engine use, using them to gain insights in user preferences, and learning from these insights, poses several unique challenges. Search interactions are not consciously made to reflect users' preferences. They are at most noisy indicators that may correlate with preference. They are strongly affected by how results are presented (e.g., position on the result page). These effects may systematically influence (i.e., bias) which results are clicked and distinguishing these effects from those of true ranking quality may be hard.

In this article, we give an overview of the solutions we have developed to address these challenges, focusing on evaluation and learning of ranking functions for IR. Our solutions include (1) an interleaved comparison method that allows unbiased and fine-grained ranker comparison using noisy click data, and, most importantly, that allows reuse of such data for new comparisons, (2) an approach for modeling and compensating for click bias in user interactions with a web search engine, (3) an experimental framework that allows the assessment of online learning to rank methods for IR using annotated data sets and click models, (4) learning approaches that improve the online performance of self-learning search engines by balancing exploration and exploitation, and (5) methods for reducing the effects of click noise in learning to rank by reusing previously observed interaction data. We provide an overview of (1) and (2) in Section 2, and discuss (3)–(5) in Section 3.

2. EVALUATION: COMPARING RANKERS USING PROBABILISTIC INTERLEAVE

A key challenge in an online learning to rank for IR is the interpretation of user feedback, such as clicks. Previous work has shown that absolute interpretations of such feedback suffer from presentation bias, and do not reliably reflect system quality [Radlinski et al. 2008].

In response to the limitations of absolute interpretations, several approaches have been developed that interpret feedback instead as *relative preferences* [Joachims 2002]. One such approach is called *interleaving*, or *interleaved comparison* [Chapelle et al. 2012]. Interleaving methods compare pairs of rankers using user interactions by combining results from both rankers. In this way, interleaving can be seen as a controlled within-subject


1) Probabilistic Interleaving
l1 → softmax s1
l2 → softmax s2
d1
d2
d3
d4
d2
d3
d4
d1
P(d_{r=1}) = 0.85
P(d_{r=2}) = 0.10
P(d_{r=3}) = 0.03
P(d_{r=4}) = 0.02
For each rank of the interleaved list l draw one of {s1, s2} and sample d:
0.5
0.85
0.5
0.87
0.5
0.6
0.5
1.0
s1
s2
d1
d2
d3
d4
All permutations of documents in D are possible.
2) Probabilistic Comparison
Observe data, e.g.
d1 1
x d2 2
x d3 1
d4 2
Marginalize over all possible assignments:

a	c_1 c_2	$P(l_i\|a,q_i)$	$P(a\|l_i,q_i)$
1 1 1 1	2 0	0.340	0.093
1 1 1 2	2 0	0.340	0.093
1 1 2 1	1 1	0.436	0.119
1 1 2 2	1 1	0.436	0.119
1 2 1 1	1 1	0.442	0.121
1 2 1 2	1 1	0.442	0.121
1 2 2 1	0 2	0.567	0.155
1 2 2 2	0 2	0.567	0.155
2 1 1 1	2 0	0.008	0.002
2 1 1 2	2 0	0.008	0.002
2 1 2 1	1 1	0.010	0.003
2 1 2 2	1 1	0.010	0.003
2 2 1 1	1 1	0.010	0.003
2 2 1 2	1 1	0.010	0.003
2 2 2 1	0 2	0.013	0.004
2 2 2 2	0 2	0.013	0.004

Compute outcomes using Equations 4-7:

$P(a|q) = 0.0625$
$P(l|q) = 0.2284$

$P(c_1>c_2) = P(o=-1) = 0.190$
$P(c_1==c_2) = P(o=0) = 0.492$
$P(c_1<c_2) = P(o=1) = 0.318$

$\hat{E}[O] = 0.128$

s_2 (based on l_2) wins the comparison on the observed interleaved list. s_1 and s_2 tie in expectation.

Fig. 1. Example interleaving (1) and comparison (2) using Probabilistic Interleave.

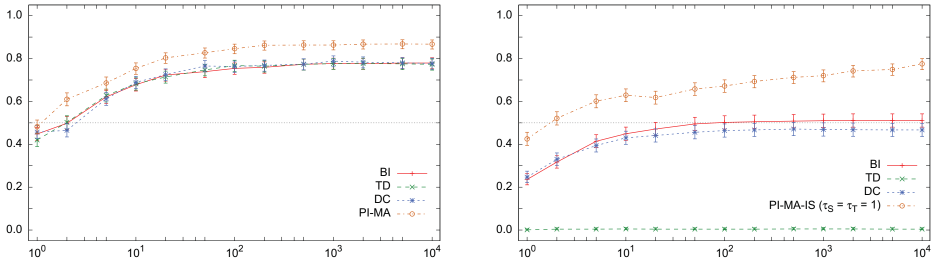
experiment.

Briefly, interleaved comparison methods work as follows. When a user submits a query, the system generates two original result lists, one for each candidate ranker, and creates an interleaved result list in such a way that each original list is equally likely to contribute its highest-ranked documents to the top of the interleaved result list (to control for position bias). This interleaved result list is presented to the user, and clicks on result documents are observed. The clicks are projected back to the original result lists to infer which is likely to be preferred by the user. The inferred comparison outcomes are aggregated over many queries to obtain a reliable estimate of the relative quality of the rankers.

Our main contribution to this area is a new interleaved comparison method, called Probabilistic Interleave [Hofmann et al. 2011b; 2013b]. One advantage of this method is that it is the first that can detect all differences between rankers in cases where one ranker Pareto-dominates (i.e., ranks all clicked documents at least as high as the competitor) the other in terms of how it ranks clicked documents. This is achieved by generating interleaved result lists probabilistically, in such a way that documents that are ranked highly by the original rankers have a high probability of being ranked highly in the interleaved list. After clicks are observed, knowledge of the probabilistic interleaving process is used to infer comparison outcomes. This results in comparisons that are more reliable and fine-grained than previously possible.

An overview and example of interleaved comparisons with Probabilistic Interleave is given in Figure 1. The first part of the figure (on the left), shows how two ranked lists, l_1 and l_2, are first transformed into probability distributions over documents (softmax, s_1 and s_2). Generating an interleaved result list is then implemented as repeated sampling without replacement from the resulting distribution. This means that all permutations of candidate documents have non-zero probability of being observed. After observing user clicks (here: on documents d_2 and d_3), we infer a comparison outcome by marginalizing over all possible assignments (ways in which either of the two rankers could have contributed documents to the observed interleaved list), as shown in part 2 of the figure.

The most important advantage of Probabilistic Interleave is that it enables the application of a statistical technique called importance sampling to the interleaving process and comparisons. In [Hofmann et al. 2012; 2013b] we showed that importance sampling can be leveraged for unbiased reuse of historical data (i.e., data observed in previous ranker com-



(a)

(b)

Fig. 2. Results: accuracy (portion of correctly detected ranker preferences) of Balanced Interleave (BI), Team Draft (TD), Document Constraints (DC), and Probabilistic Interleave (PI-MA (Probabilistic Interleave with marginalization) under live data, and PI-MA-IS (PI with marginalization and importance sampling) under historical data) under live (a) and historical (b) data.

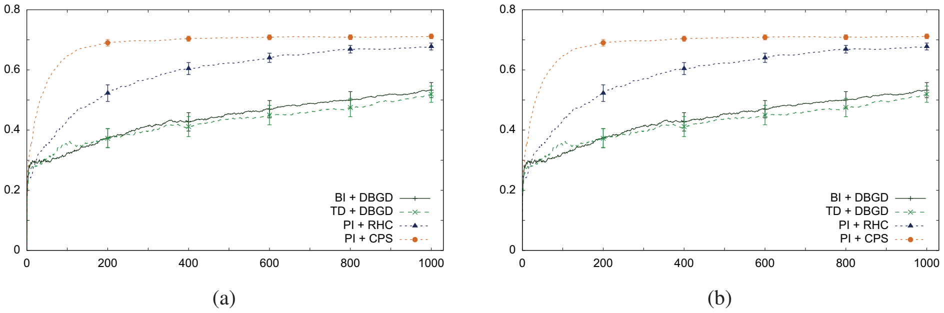
parisons). For evaluation, such data reuse addresses the bottleneck that previously new data had to be collected for each new comparison, which limited the number of possible comparisons. Now, existing data can be used to dramatically reduce the amount of new data required for a comparison.

Figure 2 shows a subset of the results obtained when assessing Probabilistic Interleave. Results were obtained using a simulation setup, where user interactions are obtained from generative click models and annotated learning to rank data sets. Probabilistic Interleave is compared to the existing interleaving methods Balanced Interleave [Joachims 2003; Radlinski et al. 2008], Team Draft [Radlinski et al. 2008], and Document Constraints [He et al. 2009]. We see that Probabilistic Interleave can detect more ranker preferences correctly with smaller amounts of data under live data (Figure 2(a)). Under historical data, it is the only interleaved comparison method that can detect preferences when comparing previously unseen ranker pairs.

In addition to noise in user clicks, we investigated caption bias (i.e., effects of how results were presented) in [Hofmann et al. 2012]. We showed how such bias can be captured using logistic regression models, validated our models in a click prediction task, and then showed how caption bias can affect interleaving outcomes (e.g., when rankers were trained using biased click data).

3. LEARNING: ONLINE LEARNING TO RANK FOR IR WITH NOISY FEEDBACK

Learning directly from user interactions is fundamentally different from the previously dominant supervised learning to rank approaches for IR, where training data was assumed to be sampled i.i.d. from some underlying distribution, and where absolute and reliable labels were provided by professional annotators. 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. Previous research resulted in solutions for learning from relative comparisons, such as the Dueling Bandit algorithm [Yue and Joachims 2009].



(a)

(b)

Fig. 3. Results: offline performance (in NDCG) of online learning to rank approaches under perfect (a) and noisy (informational, b) user interactions.

To enable assessment of online learning to rank for IR approaches, we developed an extension of the test-collection based evaluation framework commonly used in IR evaluation. This extension leverages manually annotated learning to rank data sets and models of click behavior to models various assumptions, e.g., about the amount of noise in user feedback. This framework simulates the system's environment according to these assumptions, and allows us to measure online performance, i.e., the quality of presented rankings while the system is learning [Schuth et al. 2013].

Addressing the requirement for high *online* performance, we investigated approaches for balancing exploration and exploitation in online learning to rank for IR [Hofmann et al. 2011a; 2013a]. Online learning systems need to construct result lists in such a way that users' interactions with these lists allow the systems to infer feedback that is useful for future learning. This is achieved through exploration, e.g., by trying out new rankings. However, such systems also need to exploit what they have already learned, to ensure that the presented result lists satisfy current users' needs and expectations. We developed the first approaches for balancing exploration and exploitation in a learning to rank for IR setting (with relative, noisy feedback), and showed that such a balance can substantially and significantly improve online performance in listwise [Hofmann et al. 2011a] and pairwise [Hofmann et al. 2013a] learning to rank for IR.

Finally, we developed new learning approaches that could make use of historical data, enabled by our Probabilistic Interleave method for interleaved comparisons [Hofmann et al. 2013]. We devised and investigated two such approaches, one that reuses historical data to make ranker comparisons during learning more reliable (RHC), and one that uses historical data for more effective exploration of the solution space (CPS). Both approaches lead to more reliable learning when feedback is noisy. In particular, CPS enables much faster learning than previously possible, learning up to three orders of magnitude faster than methods that do not reuse historical data.

Evaluation results of the proposed learning approaches are shown in Figure 3. We again use the simulation approach discussed above, where user interactions are obtained from generative user models. Here, we show our results for the perfect (low noise and position bias) and informational (high noise and position bias) user models. Compared are the existing learning and comparison methods Dueling Bandit Gradient Descent (DBGD) [Yue and Joachims 2009] with Balanced Interleave and Team Draft, and Probabilistic Interleave with RHC (a learning approach that uses historical data to make comparisons more reliable) and CPS. We see that Probabilistic Interleave with CPS leads to significantly better learning outcomes in both settings. Performance gains are particularly high when user feedback is noisy. These results show that CPS can effectively reuse historical data to quickly eliminate underperforming candidate rankers.

4. CONCLUSION AND FUTURE WORK

In this article, we gave an overview of our recent work on online evaluation and online learning to rank for information retrieval. First, we presented a new interleaved comparison method, probabilistic interleave. This method can detect fine-grained differences between rankers, thereby addressing limitations of previous methods. More importantly, this probabilistic method can be treated using importance sampling, enabling the reuse of previously collected observations to compare new rankers.

In Section 3, we discussed our work on balancing exploration and exploitation in online learning to rank for IR. We showed that such a balance can significantly improve the online performance of such systems, and identified differences in how pairwise and listwise learning approaches react to such a balance. Finally, we showed how data reuse with probabilistic interleave can be used to extend current learning approaches to speed up online learning to rank, especially in settings with noisy feedback.

Future work will focus on smart exploration and long-term planning and learning in search. After showing that balancing exploration and exploitation can make a big difference, developing more sophisticated approaches for doing so is a promising research direction. Finally, while our work so far treated subsequent queries as independent of each other, understanding and modeling such temporal dependencies is expected to further improve online performance.

ACKNOWLEDGMENTS

This article summarizes research presented in Katja Hofmann's PhD dissertation. A copy of the complete dissertation can be obtained from http://katja-hofmann.de and http://dare.uva.nl/record/446342.

This research was supported by the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement nr 288024 (LiMoSINe project), the Netherlands Organisation for Scientific Research (NWO) under project nrs 727.011.005, 612.001.116, HOR-11-10, the Center for Creation, Content and Technology (CCCT), the QuaMerdes project funded by the CLARIN-nl program, the TROVe project funded by the CLARIAH program, the Dutch national program COMMIT, the ESF Research Network Program ELIAS, the Elite Network Shifts project funded by the Royal Dutch Academy of Sciences (KNAW), the Netherlands eScience Center under project number 027.012.105 and the Yahoo! Faculty Research and Engagement Program.

REFERENCES

CHAPELLE, O., JOACHIMS, T., RADLINSKI, F., AND YUE, Y. 2012. Large-scale validation and analysis of interleaved search evaluation. *ACM Transactions on Information Systems 30,* 1, 1–41.

HE, J., ZHAI, C., AND LI, X. 2009. Evaluation of methods for relative comparison of retrieval systems based on clickthroughs. In *CIKM '09*. ACM Press, 2029–2032.

HOFMANN, K., BEHR, F., AND RADLINSKI, F. 2012. On caption bias in interleaving experiments. In *CIKM '12*. ACM Press, 115–124.

HOFMANN, K., SCHUTH, A., WHITESON, S., AND DE RIJKE, M. 2013. Reusing historical interaction data for faster online learning to rank for ir. In *WSDM '13*. ACM Press, 183–192.

HOFMANN, K., WHITESON, S., AND DE RIJKE, M. 2011a. Balancing exploration and exploitation in learning to rank online. In *ECIR '11*. Lecture Notes in Computer Science, vol. 6611. Springer, 251–263.

HOFMANN, K., WHITESON, S., AND DE RIJKE, M. 2011b. A probabilistic method for inferring preferences from clicks. In *CIKM '11*. ACM Press, 249–258.

HOFMANN, K., WHITESON, S., AND DE RIJKE, M. 2012. Estimating interleaved comparison outcomes from historical click data. In *CIKM '12*. ACM Press, 1779–1783.

HOFMANN, K., WHITESON, S., AND DE RIJKE, M. 2013a. Balancing exploration and exploitation in listwise and pairwise online learning to rank for information retrieval. *Information Retrieval Journal 16,* 1, 63–90.

HOFMANN, K., WHITESON, S., AND DE RIJKE, M. 2013b. Fidelity, soundness, and efficiency of interleaved comparison methods. *ACM Transactions on Information Systems 31,* 4.

JOACHIMS, T. 2002. Optimizing search engines using clickthrough data. In *KDD '02*. ACM Press, 133–142.

JOACHIMS, T. 2003. Evaluating retrieval performance using clickthrough data. *Text Mining*.

RADLINSKI, F., KURUP, M., AND JOACHIMS, T. 2008. How does clickthrough data reflect retrieval quality? In *CIKM '08*. ACM Press, 43–52.

SCHUTH, A., HOFMANN, K., WHITESON, S., AND DE RIJKE, M. 2013. Lerot: An online learning to rank framework. In *Living Lab '13: Workshop on Living Labs for Information Retrieval Evaluation*. ACM Press.

YUE, Y. AND JOACHIMS, T. 2009. Interactively optimizing information retrieval systems as a dueling bandits problem. In *ICML '09*. ACM Press, 1201–1208.

Katja Hofmann is a postdoc researcher at Microsoft Research in Cambridge. She completed her PhD at the University of Amsterdam. Her research goal is to understand how computers can effectively learn from natural user interactions. Shimon Whiteson, Anne Schuth, and Maarten de Rijke work in the Intelligent Systems Lab Amsterdam at the University of Amsterdam. Shimon Whiteson is an assistant professor of computer science whose research focuses on decision-theoretic planning and learning, including reinforcement learning, with applications ranging from robotics to information retrieval. Anne Schuth is a PhD student interested in interpreting and learning from users interacting with an information retrieval system. Maarten de Rijke is a full professor of computer science; his core research interest is intelligent information access.