Co-optimization of Multiple Relevance Metrics in Web Search

Dong Wang[1,2,*], Chenguang Zhu[1,2,*], Weizhu Chen[2], Gang Wang[2], Zheng Chen[2]

[1]Institute for Theoretical Computer Science
Tsinghua University
Beijing, China, 100084
{wd890415, zcg.cs60}@gmail.com

[2]Microsoft Research Asia
No. 49 Zhichun Road
Haidian District
Beijing, China, 100080
{v-dongmw, v-chezhu, wzchen, gawa, zhengc}@microsoft.com

ABSTRACT
Several relevance metrics, such as NDCG, precision and pSkip, are proposed to measure search relevance, where different metrics try to characterize search relevance from different perspectives. Yet we empirically find that the direct optimization of one metric cannot always achieve the optimal ranking of another metric. In this paper, we propose two novel relevance optimization approaches, which take different metrics into a global consideration where the objective is to achieve an ideal tradeoff between different metrics. To achieve this objective, we propose to co-optimize multiple relevance metrics and show their effectiveness.

Categories and Subject Descriptors
H.3.3 [**Information Systems**]: Information Search and Retrieval;

General Terms
Algorithms, Design, Experimentation, Theory.

Keywords
Learning to Rank, User Feedback, LambdaRank.

1. INTRODUCTION
Recent advances in search relevance have positioned it as a very important aspect of information retrieval (IR), and traditional works to improve search relevance can be grouped into two different categories based on the kinds of metrics used for optimization. The first one aims to improve relevance from explicitly judged labeled data by learning a ranking model to optimize a metric, like NDCG [4]. We call this kind of metric an explicit relevance metric since it's based on the explicit data. The other category looks for ways to improve search relevance by leveraging large-scale implicit user behavior log data from commercial search engines, and optimize another kind of metric, like CTR [2], pSkip [5]. We call this kind of metric an implicit relevance metric since it's based on implicit data.

However, to the best of our knowledge, previous works mostly focus on optimizing one metric to improve search relevance, though both the explicit relevance metric and implicit metric have their own merits [3]. Yet, we empirically observe that the exclusive optimization of one metric cannot always achieve the optimal ranking of another metric. For example, directly optimizing NDCG on the explicit data often results in a non-optimal relevance for pSkip on the implicit data, and vice versa. We may see this conflict from a lot of real examples. As an instance, for a query q, we will only consider its three URLs: u_1, u_2 and u_3. For a case that u_1 and u_2 are both rated as Excellent while u_2 has a higher click frequency than u_1, if we only optimize NDCG, the NDCG is maximized if we put $u_1 > u_2$, where > means the right part is put below the left part in the search result; however, the pSkip doesn't achieve the optimal result since we put u_2 with higher click frequency below u_1. In this extreme case, if we can optimize NDCG and pSkip simultaneously, we may put $u_2 > u_1$, so NDCG and pSkip can both achieve the optimal result. For another case: u_2 is a duplicate of u_1, so most users won't click u_2 and will likely jump to u_3 if they are unsatisfied with u_1. So if u_1 and u_2 are more relevant than u_3, maximizing NDCG will rank them as $u_1 > u_2 > u_3$, while optimizing pSkip will rank them as $u_1 > u_3 > u_2$ based on the click frequency. All of these real cases illustrate that we cannot solve this kind of conflict if we only consider one metric in optimization. Conversely, if we can take both metrics into consideration, it's possible for us to find an ideal tradeoff to optimize both metrics simultaneously.

In this paper, we propose to co-optimize the explicit relevance metric and implicit relevance metric simultaneously with our objective being to find an ideal co-optimization approach. Especially, we aim to answer the question: how can we maximize one metric without even slightly sacrificing another metric? For example, we aim to find a ranking function that optimizes pSkip with the constraint that the decrease of the NDCG score is less than 0.1 percent. To achieve this objective, we propose two novel methods from different machine learning approaches to co-optimize multiple relevancies.

2. LEARNING MODELS
Exclusive optimization for explicit metric cannot always achieve the optimal value for implicit metric, and verse vice. Here we propose two combination models.

2.1 Indirect Optimization Model
Firstly, we propose *indirect optimization model*. In this model, we try to integrate CTR into the calculation of NDCG. In order

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

*This work was done when the first and second authors were visiting Microsoft Research Asia.