Selective Exploration of Commercial Documents in Web Search

Alexander Shishkin, Ekaterina Gladkikh, Aleksandr Vorobev
Yandex
16 Leo Tolstoy St., Moscow, 119021 Russia
{sisoid, kglad, alvor88}@yandex-team.ru

ABSTRACT

Implicit user feedback is known to be a strong signal of user preferences in web search. Hence, solving the exploration-exploitation dilemma [5] became an important direction of improvement of ranking algorithms in the last years. In this poster, in the case of commercial queries, we consider a new negative effect of exploration on the user utility – distracting and confusing users by shifting well-known documents from their common positions – and propose an approach to take it into account within Multi-Armed Bandit algorithms, usually applied to solve the dilemma.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Learning to rank; I [**Computing methodologies**]: Online learning settings

Keywords

exploration-exploitation dilemma, web search ranking, implicit user feedback, Multi-Armed Bandit

1. INTRODUCTION

A standard scheme of collecting training data for web search ranking model is manual labeling. However, for *commercial queries* [3] (ones with commercial user intention), which form a large share of search traffic, there are many documents which present offers of the same or similar products or services and differ from each other only by specific parameters not specified in the query, such as product brand, model and price, whose influence on the user preferences between these products is difficult, if possible at all, to be analyzed by an assessor. In the standard 5-grade relevance scale (Perfect, Excellent, Good, Fair, Bad), it is more typical for Good ones.

At the same time, implicit user feedback contains a strong signal about these user preferences. Hence, it is rational to use both the explicit relevance judgments and the implicit user feedback as ground truth for training a ranking algorithm, e.g., as it is done in [1, Section 3.1]. Then, as an effective method for collecting additional implicit user feedback, exploration of unobserved documents (shown only on low positions before and, thereby, lacking user feedback) by showing them on high positions becomes especially important [5]. However, permutations of Perfect and Excellent documents are undesirable because they are well-known for users and used to have fixed positions over long periods. Users have got used to see these documents at these positions and may be confused if we change them. This effect not only strengthens the user utility decrease resulted from such permutations, but also makes the user feedback on the permutated documents biased and noisy. It does not mean that we should not explore Perfect and Excellent documents at all, but this exploration should be organized in a different way (e.g., with some period of showing a new result list just to allow users to get accustomed to it and with collecting feedback only after it).

Hence, and taking into account that exploration of Fair and Bad documents would hardly provide a significant profit, we focus on the exploration of Good documents only. It allows to show new relevant documents to users with minimal risk of degrading user experience. Formally, we are looking for the exploration algorithm which provides the minimum level of transpositions of Perfect and Excellent documents and the maximum utility of the collected feedback, given some fixed level of the ranking quality during the exploration period – period of algorithm running.

Note that Multi-Armed Bandit (MAB)-based ranking algorithms, applied to similar exploration-exploitation dilemma in previous papers [5], permute Perfect and Excellent documents if their exploitative scores (optimizing the combination of relevance and user feedback in our case) are close to each other. In this poster, we suggest a simple approach to adopt such algorithms to our restrictions on permutations of Perfect and Excellent documents, not formalized in terms of MAB problem. By large-scale experiments on real user feedback, we show that this approach outperformes the standard MAB-based ranking algorithms in our problem.

2. APPROACH

The core idea of our approach is to explore more actively documents which are more similar to Good ones. Let us consider the following general form of a MAB-based ranking score (covering, e.g., the case of UCB-1 [2]) of a document d as a baseline:

$$Score_d = BasicScore_d + ExplorAdd_d; \qquad (1)$$

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889376.

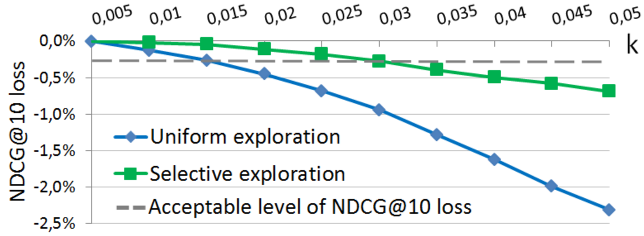

Figure 1: The dependency of NDCG@10 loss in % on value of k on the dataset D_1

where $BasicScore_d$ is some exploitative score optimizing the combination of relevance and user feedback, $ExplorAdd_d$ is