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:

$$Score_d = BasicScore_d + ExplorAdd_d, \tag{1}$$

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.

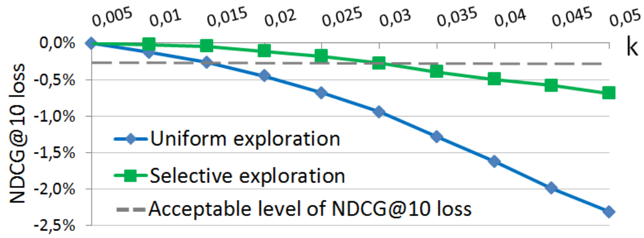

Figure 1: The dependency of NDCG@10 loss in % on value of k on the dataset D_1

where $BasicScore_d$ is some exploitative score optimizing the combination of relevance and user feedback, $ExplorAdd_d$ is a component responsible for exploration of the document d.

Then, to focus our exploration on Good documents, we modify this form in the following way:

$$Score_d = BasicScore_d + ExplorAdd'_d \cdot \hat{P}(d \in Good), \quad (2)$$

where $\hat{P}(d \in Good)$ is the estimated probability of the document d to be labeled as Good. For this prediction, we use a 5-label classifier trained by gradient boosted decision trees (GBDT) [4] maximizing the likelihood of the observed labels and relying on several hundreds of production features. The latter multiplier provides direct dependence of the exploration magnitude on our confidence that the document d is Good. We call the baseline and our exploration algorithms as *uniform* and *selective* respectively.

3. EXPERIMENTAL SETTINGS

Production realization of MAB-based ranking algorithms requires a complex infrastructure, e.g., to provide this algorithm with feedback statistics of each query-document pair. Hence, for simplicity, we use a baseline algorithm of uniform exploration with $ExplorAdd_d = k \cdot r_d$, where k is the exploration magnitude and r_d is sampled uniformly from $[-1, 1]$ independently for each query issue-document pair. For $BasicScore_d$, we use the ranking score of Yandex[1].

Having the model for estimating $\hat{P}(d \in Good)$, we applied an exhaustive search to find the value of k providing the acceptable level of ranking quality on the exploration period, 0.25% NDCG@10 loss in comparison with the production ranking (i.e., $1 - \frac{NDCG@10_{algorithm}}{NDCG@10_{production}} = 0.0025$) on the labeled dataset D_1 of 25K issues of commercial queries (see Fig. 1). The corresponding values of k are 0.009 for the uniform algorithm and 0.021 for the selective one.

Then, we conducted a range of online experiments on commercial queries submitted to Yandex in Moscow region for the same period of 7 days:

- control, i.e., the current production algorithm;
- uniform exploration, $k = 0.009$ (see Equation 1);
- selective exploration, $k = 0.021$ (see Equation 2).

Each experiment ran on a separate portion of 30K users and includes 90K query issues approximately with logged clicks on top-10 results.

To evaluate utility of the user feedback collected by each algorithm, we trained a model of CTR[2] prediction by GBDT maximizing the likelihood of the clicks and skips from the corresponding feedback and relying on several hundreds of production features.

[1]yandex.com

[2]Click-through rate. In this context - a probability of a click on a specific link in case the user has seen that link.

4. EXPERIMENTAL RESULTS

Our experimental results are presented in Table 1. All the metrics are calculated on top-10 results of each query issue with 95% confidence intervals.

The NDCG@10 loss is fixed for both uniform and selective exploration algorithms. Further, relevance losses on dataset D_1 are illustrated in detail by the share of incorrectly ranked (with respect to relevance judgments) pairs of documents from different groups (see rows 2–5). The differences between the two exploration algorithms are insignificant.

Next, expectedly, exploration allows these algorithms to outperform the control by AUC of the model of CTR prediction evaluated on the test sample of 10K query issues collected after the experimental period under the production ranking algorithm. Between the two algorithms, the selective one is only slightly better, what underlines that its aim is not to improve performance in the standard MAB problem setting, but to strengthen performance with respect to the additional criteria keeping the former performance constant. The success is confirmed by rows 7–9 showing significant advantage of the selective exploration against the uniform one in the share of documents changed their positions inside groups of Perfect (no changes for the selective algorithm!) and Excellent. Clearly, we reach this advantage by more active rotation of Good documents.

Table 1: The experimental results for control group, uniform, and selective exploration.

Metric	Control	Uniform	Selective
NDCG@10	1	$0.9975 \pm 0,003$	$0.9975 \pm 0,003$
Good-Fair&Bad	$\mathbf{0.2415} \pm 0,0004$	$0.2434 \pm 0,0004$	$0.2430 \pm 0,0004$
Good-Excellent	$0.2818 \pm 0,0024$	$\mathbf{0.2808} \pm 0,0024$	$0.2809 \pm 0,0024$
Good-Perfect	$\mathbf{0.1102} \pm 0,0065$	$0.1133 \pm 0,0066$	$0.1127 \pm 0,0065$
Excellent-Perfect	$\mathbf{0.184} \pm 0,0302$	$0.185 \pm 0,0303$	$\mathbf{0.184} \pm 0,0302$
AUC	$0.6699 \pm 0,0036$	$0.6798 \pm 0,0021$	$\mathbf{0.6830} \pm 0,0035$
Perfect	0	$0.004 \pm 1.06e\text{-}04$	**0**
Excellent	0	$0.054 \pm 5.73e\text{-}05$	$\mathbf{0.048} \pm 5.41e\text{-}05$
Good	0	$0.233 \pm 9.32e\text{-}06$	$\mathbf{0.239} \pm 9.43e\text{-}06$

5. CONCLUSIONS

In this poster, we proposed the method of focusing exploration in web search ranking for commercial queries on Good documents, and confirmed its effectiveness experimentally. The method allows us to show new relevant sites to users with minimal risk of degrading user experience. Our method can also be used for any application problem formalized as an exploration-exploitation dilemma with additional external penalty for exploration of objects of some type.

6. REFERENCES

[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *SIGIR'2006*, pages 19–26. ACM, 2006.

[2] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

[3] H. K. Dai, L. Zhao, Z. Nie, J.-R. Wen, L. Wang, and Y. Li. Detecting online commercial intention (oci). In *WWW'2006*, pages 829–837. ACM, 2006.

[4] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

[5] A. Vorobev, D. Lefortier, G. Gusev, and P. Serdyukov. Gathering additional feedback on search results by multi-armed bandits with respect to production ranking. In *WWW'2015*, pages 1177–1187, 2015.