Behavioral Dynamics from the SERP's Perspective: What are Failed SERPs and How to Fix Them?

Julia Kiseleva[1] Jaap Kamps[2] Vadim Nikulin[3] Nikita Makarov[3]

[1]Eindhoven University of Technology, Eindhoven, The Netherlands, j.kiseleva@tue.nl
[2]University of Amsterdam, Amsterdam, The Netherlands, kamps@uva.nl
[3]Yandex, Moscow, Russian Federation, {vnik,nkmakarov}@yandex-team.ru

ABSTRACT

Web search is always in a state of flux: queries, their intent, and the most relevant content are changing over time, in predictable and unpredictable ways. Modern search technology has made great strides in keeping up to pace with these changes, but there remain cases of failure where the organic search results on the search engine result page (SERP) are outdated, and no relevant result is displayed. Failing SERPs due to temporal drift are one of the greatest frustrations of web searchers, leading to search abandonment or even search engine switch. Detecting failed SERPs timely and providing access to the desired out-of-SERP results has huge potential to improve user satisfaction.

1. RESULTS

Our main research question was: By analyzing behavioral dynamics at the SERP level, can we detect an important class of detrimental cases (such as search failure) based on changes in observable behavior caused by low user satisfaction?We presented an overview of prior work on topic and concept drift, behavioral dynamics, and user satisfaction on the web, with a special focus on the $SERP$ level. We conducted a conceptual analysis of success and failure at the SERP level in order to answer our first research question: How to include the SERP into the conceptual model of behavioral dynamics on the web? How to identify (un)-successful SERPs in terms of drastic changes in observable user behavior? Specifically, we introduced the concept of a successful and failed SERP and analyzed their behavioral consequences identifying indicators of success and failure. By analyzing success and failure in light of changing query intents over time, we identified an important case of SERP failure due to query intent drift. This suggested an approach to detect a failed SERP due to query intent drift by significant changes in behavioral indicators of failure.

We continued our analysis of different types of drifts in query intent over time, answering our second research question: Can we distinguish different types of SERP failure due to query intent drift (e.g., sudden, incremental), and when and how should we update the SERP to reflect these changes? Inspired by the literature on concept drift [1], we studied different changes in query intent: sudden, incremental, gradual and reoccurring, and identified relevant parameters, such as the window of change, volume or popularity of queries, and relevant behavioral indicators, such as the probability of reformulation, abandonment rates, and click through rates. For the two main categories of intent drift, we define an unsupervised approach to detect failed SERPs caused by drift, requiring only a single pass through a transaction log.

Finally, we ran experiments on massive raw search logs, answering our third research question: How effective is our approach on a realistic sample of traffic of a major internet search engine? We ran a simplified version of our algorithm and detected over 200,000 pairs of $\langle Q, SERP\rangle$ suspected of failing due to drifting query intents, observing a reasonable accuracy of drift detection (72%) and a high accuracy of candidate URLs to be included on the SERP of the original query. For incremental change over the longer detection period of 14 days, we detected failed SERPs due to query intent drift with an 80% accuracy, but under the specific conditions of the recency optimized search engine the performance for detecting sudden change over shorter periods was less effective.

Our analysis of behavioral dynamics at the SERP level gives new insight in one of the primary causes of search failure due to temporal query intent drifts. Our overall conclusion is that the most detrimental cases in terms of (lack of) user satisfaction lead to the largest changes in information seeking behavior, and hence to observable changes in behavior we can exploit to detect failure, and moreover not only detect them but also resolve them.

Acknowledgments This is an extended abstract of [2].

2. REFERENCES

[1] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4):44:1–44:37, 2014.

[2] J. Kiseleva, J. Kamps, V. Nikulin, and N. Makarov. Behavioral dynamics from the serp's perspective: What are failed serps and how to fix them? In J. Bailey, A. Moffat, C. C. Aggarwal, M. de Rijke, R. Kumar, V. Murdock, T. Sellis, and J. X. Yu, editors, *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, pages 1561–1570. ACM, 2015.

DBDBD'15 December 16, 2015, Amsterdam, The Netherlands.