On Biases in Information Retrieval Models and Evaluation

Aldo Lipani
TU Wien
Vienna, Austria
aldo.lipani@tuwien.ac.at

Abstract

The advent of the modern information technology has benefited society as the digitization of content increased over the last half-century. While the processing capability of our species has remained unchanged, the information available to us has been notably increasing. In this overload of information, Information Retrieval (IR) has been playing a prominent role by developing systems capable of separating relevant information from the rest. This separation, however, is a difficult task rooted in the complexity of understanding of what is and what is not relevant. To manage this complexity, IR has developed a strong empirical nature, which has led to the development of grounded retrieval models, resulting in the development of retrieval systems empirically designed to be biased towards relevant information. However, other biases have been observed, which counteract retrieval performance. In this thesis, the reduction of retrieval systems to filters of information, or sampling processes, has allowed us to systematically investigate these biases.

We study biases manifesting in two aspects of IR research: retrieval models and retrieval evaluation. We start by identifying retrieval biases in probabilistic IR models and then develop new document priors to improve retrieval performance [1, 2]. Next, we discuss the accessibility bias of retrieval models, and for Boolean retrieval models we develop a mathematical framework of retrievability [3]. For retrieval evaluation biases, we study how test collections are built using the pooling method and how this method introduces bias [4]. Then, to improve the reliability of the evaluation, we first develop new pooling strategies to mitigate this bias at test collection build time [5, 6, 7, 8] and then, for two IR evaluation measures, Precision and Recall at cut-off (P@n and R@n), we develop new pool bias estimators to mitigate it at evaluation time [9, 10, 11].

Through a large scale experimentation involving up to 15 test collections, four IR evaluation measures and three bias measures, we demonstrate that including document priors based on verboseness improves the performance of probabilistic retrieval models; that the accessibility bias of Boolean retrieval models quickly worsens for conjunctive queries with the increase of the query length (while slightly improving for disjunctive queries); that the test collection bias can be lowered at test collection build time by pooling strategies inspired by a well-known problem in reinforcement learning, the multi-armed bandit problem; and that this bias can also be improved at evaluation time by analyzing the runs participating in the pool. For this last point in particular, we show that for P@n, bias reduction is done

by quantifying the potential of the new system against the pooled runs, and for R@n, this is done instead by simulating the absence of a pooled run from the set of pooled runs.

This thesis contributes to the IR field by giving a better understanding of relevance through the lens of biases in retrieval models and retrieval evaluation. The identification of these biases, and their exploitation or mitigation, leads to the development of better performing IR models and the improvement of the current IR evaluation practice.

Advisor: Prof. Allan Hanbury;

Second advisor: Dr. Mihai Lupu;

The thesis is available on the website of the author.

References

[1] Aldo Lipani, Mihai Lupu, Allan Hanbury, and Akiko Aizawa. Verboseness fission for BM25 document length normalization. In *Proc. of ICTIR '15*. ACM.

[2] Aldo Lipani, Thomas Roelleke, Mihai Lupu, and Allan Hanbury. A systematic approach to normalization in probabilistic models. *Information Retrieval Journal*, Jun 2018.

[3] Aldo Lipani, Mihai Lupu, Akiko Aizawa, and Allan Hanbury. An initial analytical exploration of retrievability. In *Proc. of ICTIR '15*. ACM.

[4] Aldo Lipani. Fairness in information retrieval. In *Proc. of SIGIR '16*. ACM.

[5] Aldo Lipani, Guido Zuccon, Mihai Lupu, Bevan Koopman, and Allan Hanbury. The impact of fixed-cost pooling strategies on test collection bias. In *Proc. of ICTIR '16*. ACM.

[6] Aldo Lipani, Joao Palotti, Mihai Lupu, Florina Piroi, Guido Zuccon, and Allan Hanbury. Fixed-cost pooling strategies based on IR evaluation measures. In *Proc. of ECIR '17*. Springer.

[7] Aldo Lipani, Mihai Lupu, Joao Palotti, Guido Zuccon, and Allan Hanbury. Fixed budget pooling strategies based on fusion methods. In *Proc. of SAC '17*. ACM.

[8] Aldo Lipani, Mihai Lupu, and Allan Hanbury. Visual pool: A tool to visualize and interact with the pooling method. In *Proc. of SIGIR '17*. ACM.

[9] Aldo Lipani, Mihai Lupu, and Allan Hanbury. Splitting water: Precision and anti-precision to reduce pool bias. In *Proc. of SIGIR '15*. ACM.

[10] Aldo Lipani, Mihai Lupu, and Allan Hanbury. The curious incidence of bias corrections in the pool. In *Proc. of ECIR '16*. Springer.

[11] Aldo Lipani, Mihai Lupu, Evangelos Kanoulas, and Allan Hanbury. The solitude of relevant documents in the pool. In *Proc. of CIKM '16*. ACM.