Balancing Efficiency and Effectiveness Trade-offs in Large Scale Multi-Stage Search Engines

J. Shane Culpepper
RMIT University
Melbourne, Australia

1 ABSTRACT

In this talk, we will discuss recent work on managing tradeoffs between efficiency and effectiveness in modern multi-stage ranking architectures which are comprised of a candidate generation stage followed by one or more reranking stages. In such an architecture, the quality of the final ranked list is often sensitive to the quality of initial candidate pool. We will briefly discuss a few recent related papers from my group, and then discuss future directions. First, we will explore dynamic cutoff prediction in early stage retrieval using query difficulty pre-retrieval features. We will then turn our attention to efficiency and effectiveness trade-offs in the later stage cascaded learning-to-rank algorithms. Specifically, we reexamine the importance of tightly integrating feature costs into multi-stage learning-to-rank (LTR) IR systems, and we present a novel approach to optimizing cascaded ranking models which can directly leverage a variety of different state-of-the-art LTR rankers such as LambdaMART and Gradient Boosted Decision Trees. Finally, we discuss interesting future research directions in multi-stage retrieval systems as modern retrieval tasks continue to evolve towards more complex interactive search systems.

Biography. Associate Professor Shane Culpepper completed his PhD in Computer Science at The University of Melbourne in 2008. He is currently a Vice-Chancellor's Principal Research Fellow and Director for the Centre for Information Discovery and Data Analytics at RMIT University in Melbourne, Australia. His current research focuses on building search systems to effectively and efficiently search web-scale data collections, and understanding how to measure the quality of the answers found. Research interests include efficient and scalable algorithm design, machine learning in information retrieval, and system evaluation. For more information about his research, visit his website at https://www.culpepper.io.

Acknowledgements. This work was supported by the Australian Research Council's *Discovery Projects* Scheme (DP170102231) and a grant from the Mozilla Foundation.

REFERENCES

[1] R.-C. Chen, L. Gallagher, R. Blanco, and J. S. Culpepper. 2017. Efficient cost-aware cascade ranking in multi-stage retrieval. In *Proc. SIGIR*. 445–454.
[2] C. L. A. Clarke, J. S. Culpepper, and A. Moffat. 2016. Assessing efficiency–effectiveness tradeoffs in multi-stage retrieval systems without using relevance judgments. *Inf. Retr.* 19, 4 (2016), 351–377.
[3] J. S. Culpepper, C. L. A. Clarke, and J. Lin. 2016. Dynamic cutoff prediction in multi-stage retrieval systems. In *Proc. ADCS*. 17–24.
[4] D. X. de Sousa, S. D. Canuto, T. C. Rosa, W. S. Martins, and M. A. Gonçalves. 2016. Incorporating risk-sensitiveness into feature selection for learning to rank. In *Proc. CIKM*. 257–266.
[5] B. T. Dinçer, C. Macdonald, and I. Ounis. 2014. Hypothesis testing for the risk-sensitive evaluation of retrieval systems. In *Proc. SIGIR*. 23–32.
[6] C. Macdonald, R. L. T. Santos, and I. Ounis. 2013. The whens and hows of learning to rank for web search. *Inf. Retr.* 16, 5 (2013), 584–628.
[7] J. Mackenzie, F. M. Choudhury, and J. S. Culpepper. 2015. Efficient location-aware web search. In *Proc. ADCS*. 4.1–4.8.
[8] J. Mackenzie, J. S. Culpepper, R. Blanco, M. Crane, C. L. A. Clarke, and J. Lin. 2018. Query Driven Algorithm Selection in Early Stage Retrieval.. In *Proc. WSDM*. 396–404.
[9] H. R. Mohammad, K. Xu, J. Callan, and J. S. Culpepper. 2018. Dynamic shard cutoff prediction for selective search. In *Proc. SIGIR*. 85–94.
[10] J. Pedersen. 2010. Query understanding at Bing. *Invited talk, SIGIR* (2010).
[11] S. Peter, F. Diego, F. A. Hamprecht, and B. Nadler. 2017. Cost efficient gradient boosting. In *Proc. NIPS*. 1551–1561.
[12] L. Wang, P. N. Bennett, and K. Collins-Thompson. 2012. Robust ranking models via risk-sensitive optimization. In *Proc. SIGIR*. 761–770.
[13] L. Wang, J. Lin, and D. Metzler. 2011. A cascade ranking model for efficient ranked retrieval. In *Proc. SIGIR*. 105–114.

DESIRES 2018, August 2018, Bertinoro, Italy