

Stopping Methods Based on Point Processes: Recent Developments

Mark Stevenson¹, Reem Bin-Hezam^{1,2}

¹*School of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield, United Kingdom*

²*Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Saudi Arabia*

Abstract

Technology Assisted Review (TAR) develops methods to assist high-recall retrieval, the process of identifying a large proportion (or all) of the relevant documents within a collection. Stopping methods aim to determine when enough relevant documents have been identified to achieve a desired level of recall (the *target recall*), potentially leading to a substantial reduction in the effort required to screen a collection. We summarise two recent publications describing a TAR stopping method based on point processes [1, 2].

Point processes are statistical models that can represent the occurrence of events, such as encountering a relevant document when examining a ranked list. We estimate the total number of relevant documents in a ranked collection by obtaining relevance judgements for the highest ranked documents and then applying a point process to estimate the number of relevant documents in the remainder of the collection. This approach was evaluated using standard collections within TAR research and compared against eight baseline methods employing a range of stopping strategies. Results indicated that the proposed approach performs well in comparison with the baselines and is Pareto optimal in almost all configurations.

The approach was extended by training a text classifier using existing relevance judgements which is then applied to the remaining documents and the information produced used to refine the point process's estimate of the number of relevant documents. Experiments showed that the approach substantially reduced the cost (often by as much as 50%) of the original approach with only minimal (statistically insignificant) reduction in reliability.

Keywords

Technology Assisted Review, TAR, Stopping methods, Point processes

References

- [1] M. Stevenson, R. Bin-Hezam, Stopping Methods for Technology-assisted Reviews Based on Point Processes, *ACM Trans. Inf. Syst.* 42 (2023). doi:10.1145/3631990.
- [2] R. Bin-Hezam, M. Stevenson, Combining Counting Processes and Classification Improves a Stopping Rule for Technology Assisted Review, in: *Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore, 2023, pp. 2603–2609. doi:10.18653/v1/2023.findings-emnlp.171.

3rd Workshop on Augmented Intelligence for Technology-Assisted Review Systems (ALTARS 2024): Evaluation Metrics and Protocols for eDiscovery and Systematic Review Systems. March 28, 2024, Glasgow, UK.

✉ mark.stevenson@sheffield.ac.uk (M. Stevenson); rybinhezam@pnu.edu.sa (R. Bin-Hezam)

🌐 <https://www.sheffield.ac.uk/dcs/people/academic/mark-stevenson> (M. Stevenson)

🆔 0000-0002-9483-6006 (M. Stevenson); 0000-0002-9156-6186 (R. Bin-Hezam)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).