Challenges and Opportunities in Extreme Systematic Reviews for Environmental Policymaking

Jean- Jacques Dubois[1], Yue Wang[2]

[1]United States Environmental Protection Agency, Research Triangle Park, North Carolina, USA
[2]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Extended Abstract

Since 1977, the U.S. Clean Air Act (CAA) has directed the United States Environmental Protection Agency (USEPA) to synthesize the most recent scientific research at regular intervals for each of six criteria ambient air pollutants. These documents, called Integrated Science Assessments (ISAs), aim to provide an updated, comprehensive review of the state of the science on the health and welfare effects of these air pollutants, and lay the scientific foundation for the policymakers setting the National Ambient Air Quality Standards (NAAQS) [1]. Because an ISA document has far-reaching implications on environmental policies and subsequent impacts on public health and welfare, the USEPA invests substantial resources to ensure that the document covers up-to-date policy-relevant literature as comprehensively as possible. As a result, ISAs can be viewed as extreme systematic reviews distinguished by unprecedented scale (often citing thousands of references screened from search results containing hundreds of thousands of entries), the pursuit of maximum recall (95% or higher coverage of relevant literature), periodic updates (every five years as per CAA's requirement), and integration of studies that constitute the evidence basis for environmental policymaking across a very wide range of topics and disciplines [2].

The USEPA has had some success using commercial implementations of technology-assisted review (TAR) technologies that employ active learning. It has been leveraging the systematic and recurring nature of ISAs to accelerate an active learning process used in screening recent literature. For example, the number of references a newly published article shares with a prior ISA can be used to assign a pseudo-relevance label and warm-start the active learning process. However, these techniques have not reduced the time and labor spent on reviewing literature that will not be cited in an ISA as drastically as needed. The USEPA has been actively exploring new paths towards improving current approaches to technology-assisted literature screening. We would like to share some of these paths with the participants of the ALTARS workshop.

Data Release: The USEPA has recently released a unique data resource associated with the problem of ISA literature screening [2]. The resource demonstrates the special nature of

ALTARS 2023: 2nd Workshop on Augmented Intelligence for Technology-Assisted Reviews Systems, April 2, 2023, Dublin, Ireland

dubois.jeanjacques@epa.gov (J. J. Dubois); wangyue@unc.edu (Y. Wang)
0000-0002-0278-2347 (Y. Wang)

the problem and calls for new solutions on many fronts, including new document ranking models for machine-assisted review, new methods that simultaneously guarantee an ultra high recall while minimize excessive reviewing efforts, and new human-in-the-loop workflows and interactive designs that effectively leverage the knowledge of a diverse team of human scientists in triaging a large amount of information.

Recall Estimation: The USEPA has been investigating methods for reliably estimating the recall of relevant articles cited in systematic literature reviews to allow for termination of the review process as early as possible without compromising recall. This is especially important given the large scale of ISAs and their extremely high recall requirement. Although several early-stopping rules and recall estimation methods have been proposed in recent literature [3, 4, 5], they need to be rigorously tested against a target recall at 95% or higher on ISA corpora.

Workflow Innovation: The USEPA has been looking into the possibility of recasting the screening procedure from batch mode to stream mode. Given the recurring nature of ISAs, the literature screening work for a future ISA does not have to wait until the set of all potentially relevant articles have been published and are ready to be screened in one batch. Instead, the stream of potentially relevant articles are continuously generated over multiple years and can be screened as soon as they are published. This provides two advantages. First, amortizing the reviewing efforts across a longer timeline lowers the human stress in reviewing an exceedingly large batch of articles. Second, an increasing number of carefully reviewed relevant (and non-relevant) documents found in the stream-screening process might continuously improve a relevance ranking model over time, eliminating the cold-start problem in a batch-mode active learning process.

The special features of USEPA's ISAs provide an ideal environment to explore specific TAR topics of interest, such as ultra-high recall, warm start techniques, and ranker selection approaches in TAR. We hope that the workshop participants find that our work can contribute to solutions in less extreme contexts.

References

[1] U. S. E. P. Agency, Integrated Science Assessments (ISAs), https://www.epa.gov/isa, 2023.

[2] J. Hou, X. Wang, J.-J. Dubois, R. B. Rice, A. Haddock, Y. Wang, Extreme systematic reviews: A large literature screening dataset to support environmental policymaking, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 4029–4033.

[3] B. E. Howard, J. Phillips, A. Tandon, A. Maharana, R. Elmore, D. Mav, A. Sedykh, K. Thayer, B. A. Merrick, V. Walker, et al., Swift-active screener: Accelerated document screening through active learning and integrated recall estimation, Environment International 138 (2020) 105623.

[4] D. D. Lewis, E. Yang, O. Frieder, Certifying one-phase technology-assisted reviews, in: Proceedings of the 30th ACM international conference on information & knowledge management, 2021, pp. 893–902.

[5] E. Yang, D. D. Lewis, O. Frieder, Heuristic stopping rules for technology-assisted review, in: Proceedings of the 21st ACM Symposium on Document Engineering, 2021, pp. 1–10.