Studies In Aspects of Peer Review: Novelty, Scope, Research Lineage, Review Significance, and Peer Review Outcome

Tirthankar Ghosal
Department of Computer Science and Engineering
Indian Institute of Technology Patna, India
tirthankar.pcs16@iitp.ac.in

Abstract

The process of peer-review is considered as the *sentinel of science* in scholarly communications. However, with the deluge of research articles and the overload of scholarly information, it is increasingly becoming difficult for humans to keep up with the pace of the latest research. This thesis uses Machine Learning (ML) and Natural Language Processing (NLP) to investigate critical issues relevant to scholarly communications and peer review. Detecting whether a document contains sufficient new information to be deemed as *novel* is of immense significance in this age of data duplication. We investigate if we can automatically identify the *newness* in a document based on its information content against a set of relevant documents. We create two benchmark datasets: TAP-DLND 1.0 and TAP-DLND 2.0, and devise several methods for novelty detection at the document level. One particular challenge in peer review is that the journal editors and conference program chairs are overwhelmed with the ever-increasing rise in article submissions. Studies show that many submissions are not well-informed and do not fit within the scope of the intended journal or conference. We investigate how an AI could assist the editors and program chairs in identifying potential *out-of-scope* submissions based on the past accepted papers of the particular journal conference. We experiment with several methods, including feature-based machine learning, deep multimodal architectures, to multiview clustering. Again, the recent unprecedented growth in paper submissions in major Machine Learning and AI conferences is a grand challenge to the community to maintain high-quality reviewing practices while ensuring fair evaluation of the manuscripts. We investigate if we can identify which reviews were significant enough to be considered in the decision-making. We design an end-to-end deep network that leverages on *exhaustiveness* and *strength* of the peer review in a multi-task fashion to comprehend it's *informativeness*. Our *DeepSentiPeer* model is one of the first attempts to predict decisions in peer review automatically. Finally, finding the relevant literature or identifying the works that have inspired the current work under scrutiny is central to an evaluation in peer review. We analyze the context of citations within paper full-text to identify a given research lineage. The idea is to leverage citation significance classification to build a research lineage via a *significant citation graph*. We argue that finding the lineage of given research would help identify the true academic impact of a paper beyond quantitative citation counts.

Awarded by: Indian Institute of Technology Patna, India **on** 28 August, 2021.
Supervised by: Asif Ekbal and Pushpak Bhattacharyya.
Available at: https://www.iitp.ac.in/~ai-nlp-ml/resources/dissertations/PhD/Tirthankar_Ghosal_PhD_Thesis_IIT_Patna.pdf.

Acknowledgement

Tirthankar Ghosal is a Visvesvesvaraya Ph.D. fellowship awardee (# VISPHD-MEITY-2518) and acknowledges Digital India Corporation, Ministry of Electronics and Information Technology, Government of India, to support his doctoral studies.

Selected Publications

Tirthankar Ghosal, Vignesh Edithal, Asif Ekbal, Pushpak Bhattacharyya, George Tsatsaronis, and Srinivasa Satya Sameer Kumar Chivukula. Novelty goes deep. a deep neural solution to document level novelty detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2802–2813, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics. URL https://aclanthology.org/C18-1237.

Tirthankar Ghosal, Ashish Raj, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya. A deep multimodal investigation to determine the appropriateness of scholarly submissions. In Maria Bonn, Dan Wu, J. Stephen Downie, and Alain Martaus, editors, *19th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019*, pages 227–236. IEEE, 2019a. doi: 10.1109/JCDL.2019.00039. URL https://doi.org/10.1109/JCDL.2019.00039.

Tirthankar Ghosal, Rajeev Verma, Asif Ekbal, and Pushpak Bhattacharyya. DeepSentiPeer: Harnessing sentiment in review texts to recommend peer review decisions. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1120–1130, Florence, Italy, July 2019b. Association for Computational Linguistics. doi: 10.18653/v1/P19-1106. URL https://aclanthology.org/P19-1106.

Tirthankar Ghosal, Tanik Saikh, Tameesh Biswas, Asif Ekbal, and Pushpak Bhattacharyya. Novelty Detection: A Perspective from Natural Language Processing. *Computational Linguistics*, pages 1–42, 12 2021a. ISSN 0891-2017. doi: 10.1162/coli_a_00429. URL https://doi.org/10.1162/coli_a_00429.

Tirthankar Ghosal, Piyush Tiwary, Robert Patton, and Christopher Stahl. Towards establishing a research lineage via identification of significant citations. *Quantitative Science Studies*, pages 1–19, 2021b. URL https://doi.org/10.1162/qss_a_00170.