Reviewerly: Modeling the Reviewer Assignment Task as an Information Retrieval Problem

Negar Arabzadeh
Reviewerly
negara@reviewer.ly

Sajad Ebrahimi
Reviewerly
sajad@reviewer.ly

Sara Salamat
Reviewerly
sara@reviewer.ly

Mahdi Bashari
Reviewerly
bashari@reviewer.ly

Ebrahim Bagheri
Reviewerly
bagheri@reviewer.ly

ABSTRACT

The peer review process is a fundamental aspect of academic publishing, ensuring the quality and credibility of scholarly work. In this talk, we will explore the critical challenges associated specifically with the assignment of reviewers to submitted papers. We will introduce Reviewerly, our innovative solution designed to enhance the efficiency and effectiveness of reviewer assignments by leveraging data from diverse sources, including OpenAlex, PubMed, and DBLP. By modeling the reviewer assignment problem as an information retrieval task, we focus on retrieving a pool of relevant and diverse reviewers for each paper. We will highlight the challenges we faced and showcase the benefits of this approach in addressing the reviewer assignment problem.

1 INTRODUCTION

The peer review process is essential in academic publishing, as it guarantees the quality and reliability of scholarly articles. However, assigning suitable reviewers to submitted papers remains a major challenge [9]. Traditional manual methods of reviewer assignment, such as asking reviewers to bid on papers, are labor-intensive and often ineffective. This approach places a significant burden on reviewers and can result in nonconstructive or unfair reviews if the reviewers' expertise does not closely match the paper's content. Moreover, in cases like journal review processing, where submission deadlines might be spread throughout the year, the bidding approach is impractical. Furthermore, the time required to assign reviewers can considerably extend the submission-to-publication timeline, delaying the dissemination of crucial research findings.

Despite the critical nature of the Reviewer Assignment Problem (RAP), only a few efforts [1, 4, 6, 13] have made strides in automating parts of the process, and comprehensive, widely-adopted solutions remain scarce. This gap in the research highlights the need for innovative open-sourced approaches to improve the efficiency and effectiveness of reviewer assignments.

Reviewerly[1]**:** Reviewerly aims to tackle the challenges faced in RAP by utilizing the extensive data from sources like OpenAlex, DBLP, and PubMed to suggest a solution based on well-studied IR approaches and leveraging a vast array of entities including papers, topics, institutions, venues, authors, etc. Reviewerly has a user-friendly interface for exploring papers and research topics as well as ensuring the paper is matched with the suitable set of reviewers. Moreover, it integrates with Open Journal Systems (OJS) to provide smooth reviewer recommendations and offers an API for integrating our solution with other platforms. To tackle RAP, we experimented with various methods, including traditional lexical matching approaches [11], static neural embedding-based approaches [10], and fine-tuning pre-trained language models for academic recommendations [2, 3]. However, we discovered that none of the current open-source approaches provide satisfactory results for this task in real-world scenarios. Our findings revealed that performance is long-tailed over specific papers and biased toward reviewers with a higher number of publications, often neglecting early-career reviewers. Unlike previous work, we conceptualize RAP as an IR task to leverage established IR techniques for efficiently matching papers with the most relevant reviewers. In this context, papers are treated as queries, and reviewers as items to be ranked and evaluation is conducted based on relevance and diversity of the retrieved set of items (reviewers).

Beyond an expert reviewer. Reviewer recommendation is more than just about finding relevant reviewers. We aim to handle conflicts of interest automatically to the best extent possible. In addition, in our evaluations, we also consider diversity in several aspects. We propose a set of novel evaluation metrics to ensure that the set of recommended reviewers includes individuals at different stages of their careers and from various institutional backgrounds, avoiding biases and ensuring a more equitable review process.

In this talk, we offer a comprehensive understanding of RAP by discussing the RAP challenges and presenting a methodology that is not only effective but also addresses the lack of data problem. We also elaborate on different evaluation strategies for RAP, and introduce the tools and APIs we have developed for this task.

2 CHALLENGES AND SOLUTIONS

Data. One of the challenges in RAP is the lack of diverse and comprehensive datasets for both training as well as evaluation of the current approaches due to many reasons such as keeping the anonymity of reviewers. The available datasets are often limited in scope and may not cover the wide range of topics and fields required for accurate reviewer matching. Maintaining and curating these datasets is also challenging, as they need to be regularly updated to reflect the latest research trends and reviewer profiles. To the best of our knowledge there is no dataset that provides a diverse and comprehensive ground truth for this task. There are only limited number of works encompassing the assigned reviewers to the accepted papers as gold standards [14].

Approach. There have been numerous attempts by both academia and industry researchers to tackle RAP. Industrial approaches, while potentially effective, are often proprietary and lack transparency,

[1] https://reviewer.ly/

making it difficult to evaluate and compare their performance. Current approaches can be categorized into topic-based modeling methods such as Latent Dirichlet Allocation (LDA) [7, 8] and lexical matching-based approaches (e.g., TF-IDF [11]). Additionally, by emerging neural-based embeddings, they have been attempts to tackle RAP using static word2vec-based as well as contextualized embeddings [10]. These models often exhibit skewed performance, excelling on popular topics but struggling with long-tailed distributions where papers belong to niche or emerging areas.

RAP has been modeled in different ways, such as classification task[15, 16], topic coverage task [7], recommender system and so on. However, we propose to model RAP as a traditional IR task. In an IR task, the goal is to find relevant documents from a large collection based on a given query. In our context, the queries are the papers submitted for review, represented by their titles, abstracts, and potentially other metadata such as venues and keywords. The collection, on the other hand, consists of the previous works of potential reviewers. By treating the previous works of reviewers as our document collection, we can leverage the similarities between the queries (submitted papers) and documents (reviewers' past works) to make informed reviewer assignments. The task is thus redefined as retrieving authors whose prior work is most similar to the content of the submitted papers.

Advantages. Modeling RAP as IR task offers several advantages: 1) We can use previous sets of papers as queries and their authors as potential ideal reviewers. Therefore, we can train our IR model using paper information and their authors as labels, skipping the challenge of annotation and lacking enough training data. This data could potentially serve well for training and evaluation purposes. 2) The IR task is well-studied, allowing us to leverage established retrieval methods that have previously demonstrated reliable performance on downstream tasks. 3) We can leverage fine-tuned language models for specific topics for the performancer boost. For example, models like SciBERT [2] and SPECTER [3] have shown great performance on representation of scientific contents.

To achieve this, we have collected a dataset specifically for this task, comprising a large corpus of academic papers from DBLP, OpenAlex, and PubMed, focusing on the domains of Computer Science and Medicine for now. We then fine-tuned dense retrievers in a bi-encoder based architecture by the papers information with the goal of retrieving the authors of the papers as their ideal reviewers. Dense retrievers has demonstrated reliable performance on many other benchmarks [5, 12]. Once the relevant authors are retrieved, we further refine the selection by adopting metrics to balance the expertise with diversity considerationssuch as their career stage, institutional background, and other diversity criteria. In the talk, we will go over the evaluation metrics we designed for this task , ensuring a diverse and fair review process. Our results show that our proposed approach in Reviewerly, show a better performance when dealing with real world data for tackling RAP problem.

3 SPEAKER BIOGRAPHIES

Negar Arabzadeh is the data science lead at Reviewerly. She is also a Ph.D. student at the University of Waterloo. Her research is aligned with ad-hoc search in IR. She has published relevant papers in SIGIR, CIKM, ECIR, and IP&M and previously conducted tutorials in SIGIR 2022, ECIR 2023-2024, and WSDM 2023.

Sajad Ebrahimi is serving as a Data Scientist at Reviewerly while pursuing an MASc at the University of Guelph. His expertise lies in Natural Language Processing (NLP) and IR, demonstrated by his publications at prestigious information retrieval conferences.

Sara Salamat is a data scientist at reviewerly. Sara is also a computer engineering MASc student at Toronto Metropolitan University. Her expertise spans NLP and deep learning. She has published her research in top-tier venues such as CIKM and ECIR.

Mahdi Bashari Mahdi Bashari is the co-founder of Reviewerly, a startup aimed at transforming academic collaboration by creating a platform that connects researchers across various disciplines and geographic boundaries. He holds a Ph.D. from the University of New Brunswick, and his research focuses on the application and commercialization of machine learning algorithms.

Ebrahim Bagheri is the co-founder of Reviewerly. He concurrently holds a Canada Research Chair in Social Information Retrieval and an NSERC Industrial Research Chair on Social Media Analytics. In 2019, Ebrahim received the prestigious Government of Canada NSERC Synergy Award for Innovation for his outstanding industry-academia collaborations. He is the Associate Editor for IEEE Transactions on Network Science and Engineering and ACM Transactions on Intelligent Systems and Technology.

4 CONCLUDING REMARKS

In conclusion, Reviewerly demonstrates a novel approach to the Reviewer Assignment Problem by modeling it as an information retrieval task. Future work will focus on refining models, expanding datasets, and integrating additional features to further support journal editors and conference organizers, ultimately enhancing the efficiency and effectiveness of the peer review process.

REFERENCES

[1] Meltem Aksoy, Seda Yanik, and Mehmet Fatih Amasyali. 2023. Reviewer Assignment Problem: A Systematic Review of the Literature. *JAIR* (2023).

[2] Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676* (2019).

[3] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S Weld. 2020. Specter: Document-level representation learning using citation-informed transformers. *arXiv preprint arXiv:2004.07180* (2020).

[4] David Hartvigsen, Jerry C Wei, and Richard Czuchlewski. 1999. The conference paper-reviewer assignment problem. *Decision Sciences* (1999).

[5] Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. *arXiv preprint arXiv:2004.04906* (2020).

[6] Tomasz Kolasa and Dariusz Krol. 2011. A survey of algorithms for paper-reviewer assignment problem. *IETE Technical Review* (2011).

[7] Ngai Meng Kou, Leong Hou U, Nikos Mamoulis, and Zhiguo Gong. Weighted Coverage based Reviewer Assignment. In *ACM SIGMOD 2015*.

[8] David Mimno and Andrew McCallum. Expertise modeling for matching papers with reviewers. In *Knowledge Discovery and Data Mining*.

[9] Kanika Mittal, Amita Jain, and Kunwar Singh Vaisla. 2019. Understanding reviewer assignment problem and its issues and challenges. In *IoT-SIU*. IEEE.

[10] O Ogunleye, T Ifebanjo, T Abiodun, and AA Adebiyi. Proposed framework for a paper-reviewer assignment system using word2vec. In *CUCEN*.

[11] Hongwei Peng, Haoji Hu, Keqiang Wang, and Xiaoling Wang. Time-Aware and Topic-Based Reviewer Assignment. In *DASFAA Workshops*.

[12] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[13] Fan Wang, Ning Shi, and Ben Chen. 2010. A comprehensive survey of the reviewer assignment problem. *IJITDM* 9, 04 (2010), 645–668.

[14] Yichong Xu, Han Zhao, Xiaofei Shi, and Nihar B. Shah. 2018. On Strategyproof Conference Peer Review. *CoRR* abs/1806.06266 (2018). arXiv:1806.06266

[15] Dong Zhang, Shu Zhao, Zhen Duan, Jie Chen, Yanping Zhang, and Jie Tang. 2020. A Multi-Label Classification Method Using a Hierarchical and Transparent Representation for Paper-Reviewer Recommendation. *TOIS* (2020).

[16] Shu Zhao, Dong Zhang, Zhen Duan, Jie Chen, Yan-ping Zhang, and Jie Tang. 2018. A novel classification method for paper-reviewer recommendation. (2018).