

# COMP 550 Project Proposal

Fuyuan Lyu, Kian Ahrabian, Aarash Feizi

October 2019

## 1 Introduction

In recent years, there has been a large number of studies trying to measure the ability of natural language processing (NLP) techniques to understand the underlying meaning of texts. At the same time, there has been a rise in the sheer number of research venues and consequently papers submitted to these venues. Fortunately, parallel to this rise, several platforms such as “OpenReivew” containing papers, reviews, and rebuttals have emerged. As a result, with these data available publicly, the question arises whether machines are able to understand and capture the underlying meaning of fairly scientific texts or not? In this work, we plan to study the above-mentioned question from different perspectives.

The rest of the document is organized as follows. In Sec. 2 we will discuss the specific research problem that we want to tackle. Moreover, in Sec. 3 we will introduce related work that has been done previously. Finally, Sec. 4 will contain information about the data that we plan to work on.

## 2 Research Problem

The motivation behind this problem is to measure the extent that NLP techniques can understand scientific articles and assist researchers in the process of publishing articles by providing them with decent auto-generated reviews for their drafts.

Therefore, the ultimate challenge we would like to tackle is building a model that can generate reasonable reviews for a given draft paper as input. Ideally, the intention is to build a model that could understand scientific articles to some degree and generate reviews based on the flaws that it detected.

Although the main goal is to build a model that generates reviews, the solution might be more complex, requiring domain-specific knowledge rather than merely using textual signals with machine learning and NLP techniques [3]. Hence, we propose a pair of alternative tasks, in case achieving the main goal proves to be infeasible due to the limited available time and resource, similar to previously well-studied topics in NLP literature, to measure the ability of the model to understand such articles. Our proposed tasks are as follows

- Predicting expert scores based on the given article and expert reviews.
- Summarizing given articles into a few sentences.

## 3 Related Work

Similar problems have been studied in recent years. In [1], the authors tried to investigate the influence of rebuttals on the final score of a draft paper and moreover, predict the final (after-rebuttal) score based on the author’s responses. In [3], the authors argue that automated scoring, necessitates multidisciplinary collaborations that NLP techniques is merely one of them. The definition of automated scoring is automatically generating feedback and score based on analyzing a student’s response to a question. In [4], the authors aim to find the important components and sentences using both vision-based and language-based methods.

## 4 Datasets

In our literature review, the only relevant dataset that we came across was “PeerRead” [2] which contains 14.7K paper drafts along with 10.7K textual paper reviews. Depending on the time allowance, we might opt to mine more labeled and/or unlabeled data from sources such as arXiv.

## References

- [1] Yang Gao, Steffen Eger, Ilia Kuznetsov, Iryna Gurevych, and Yusuke Miyao. Does my rebuttal matter? insights from a major nlp conference. 03 2019.
- [2] Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Madeleine van Zuylen, Sebastian Kohlmeier, Eduard Hovy, and Roy Schwartz. A dataset of peer reviews (peerread): Collection, insights and nlp applications. *arXiv preprint arXiv:1804.09635*, 2018.
- [3] Nitin Madnani and Aoife Cahill. Automated scoring: Beyond natural language processing. In *COLING*, 2018.
- [4] Shintaro Yamamoto, Yoshihiro Fukuhara, Ryota Suzuki, Shigeo Morishima, and Hirokatsu Kataoka. Automatic paper summary generation from visual and textual information. *CoRR*, abs/1811.06943, 2018.