

META REVIEW GENERATION

Sri Ram Pavan Kumar Guttikonda

Auburn University

szg0148@auburn.edu

Abstract

Meta-Review generation is relatively new task which involved active research during the past few years. Meta-Review can be viewed as closely related to abstractive summarization but is not simple summarization, it involves other information like decision. It is dependent on many other sub tasks like decision prediction and reviewer sentiment which makes it complex and challenging. The main idea behind this project is based on the intuition that Meta-Review will contain mostly those sentences which are similar to sentences common among the three reviews and additionally it will contain decision. This project will deal with two steps, selecting the most common sentences using the sentence clustering algorithms and then perform abstractive summarization on the sentences selected. The results will be evaluated using BART-Score and reported.

1 Introduction

Peer-Review system is widely used by well known conferences to evaluate the research paper based on the artifacts presented by different reviewers. The Peer-Review system consists of the area chair/program chair who assigns the research paper to independent subject experts. The subject experts evaluate the paper and present their review back to the area chair/program chair. The area chair/program chair makes a decision whether the paper has to be accepted or rejected based on the reviews and leaves comments/meta-review for the author. With the surge in number of conferences, number of submissions to different conferences and lack of enough number of subject experts, the task became complicated due to strict time frame. The quality of the review got degraded because of the time complications, and thus an AI agent is necessary to speedup the meta-review process. In this

project I will design a system capable of generating effective meta-review given peer-reviews.

2 Related Works

Meta-Review generation is still a new task and it has very few works around it. Earlier works include MetaGen(Bhatia et al., 2020) and Decision-Aware Meta-Review generation(Kumar et al., 2021). MetaGen(Bhatia et al., 2020) works on generating an initial draft by selecting the k top sentences using the Random Walk Restart algorithm. The initial draft is used to predict the decision, the decision and the initial draft together forms the final draft which is then used for meta-review generation. On the other hand Decision-Aware MRG(Kumar et al., 2021) enforces a multi-encoder architecture, where the reviews are given as input to multiple encoders and the outputs from the encoder are used to predict the decision and also given as input to a single decoder. The decision predicted is given to the last layer of the decoder and is also a part of the loss function. These are the only research papers present based on Meta-Review generation to the best of my knowledge. These papers have their own limitations which needs to be addressed. Meta-Gen(Bhatia et al., 2020) uses the meta-review data available to update the weights for the sentence graph generated through Random Walk Restart algorithm, this makes it biased to particular conference. Decision-Aware(Kumar et al., 2021) is highly dependent on the decision prediction, If the predicted decision is wrong it will have a huge effect on meta-review making the system sensitive to predicted decision. The upcoming projects are focused mostly on Decision-Aware approach, in my project I am moving forward with the intuition that most of the meta-reviews are based on the common sentences among the reviews and decision.

3 Research Plan

3.1 Dataset

There is currently no pre-defined standard dataset for the task. Dataset is not present due to data confidentiality and copyrights issue. PeerRead is the only dataset which consists of the crawler scripts to get the data from the conference site, but the crawler scripts are outdated because of the transformation of conference websites. Over the past few years conferences have been moving towards the open review system and hence data is available on Openreview website. I have written the crawler script for the same and was able to get around 1000 samples. This could be a potential hindrance to the project, but I am looking for ways to work with small amount of data.

3.2 Method

Traditionally there are two ways through which we can generate a meta-review, 1) Extractive Summarization 2) Abstractive Summarization. Extractive Summarization is the selection of most relevant top k sentences from the input. Text Rank, Lex Rank and LSA are some of the algorithms used to obtain extractive summary. Abstractive Summarization on the other hand is complex compared to extractive summarization. In abstractive summarization we pass the input to a seq2seq model and get the summary generated. We can do abstractive summarization using models like BART. Meta-review cannot be achieved through summarization alone, it is more complex than summarization since it comprises of sub-tasks like finding the decision. We need the model to have capability to generate output based on common sentences and decision, which is what I will be working on for this project.

3.3 Experiments

The earlier works MetaGen(Bhatia et al., 2020) and Decision-Aware(Kumar et al., 2021) have their own limitations which were listed above. Both the research papers used seq2seq transformer for the abstractive summarization part. Transformers generally lag in content selection, they may not be smart enough to select those sentences which should be present in the meta-review. I will work on finding a way to select the most important sentences for meta-review, and according to my intuition that is the common sentences among reviews. I will be using sentence clustering algorithms like (Sarkar, 2009) to cluster all the available sen-

tences to k clusters and pick one sentence from each cluster, this is extractive summary. Next I will be using BART to convert the extractive summary to meta-review. I am looking at ways to alter the attention mechanism of the transformer to make it give importance to few sentences over the other, which is still yet to be discovered. Looking at the feasibility of using Bottom-up abstractive summarization (Gehrmann et al., 2018) and Multi-News (Fabbri et al., 2019) for the purpose of this task.

4 Evaluation

ROUGE has been used as the metric of evaluation for most of the summarization tasks. ROUGE is a good metric for extractive summarization task but is not a good metric to evaluate text generated by abstractive summarization, which is also described in the Decision-Aware(Kumar et al., 2021) paper. Since ROUGE is not a suitable metric for evaluation, I am going to use BartScore as an evaluation metric for my project. Additionally I am trying to ask humans to evaluate few of the generated meta-reviews, whether they are close to the original meta-reviews.

References

- Chaitanya Bhatia, Tribikram Pradhan, and Sukomal Pal. 2020. Metagen: An academic meta-review generation system. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1653–1656.
- Alexander R Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir R Radev. 2019. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. *arXiv preprint arXiv:1906.01749*.
- Sebastian Gehrmann, Yuntian Deng, and Alexander M Rush. 2018. Bottom-up abstractive summarization. *arXiv preprint arXiv:1808.10792*.
- Asheesh Kumar, Tirthankar Ghosal, and Asif Ekbal. 2021. A deep neural architecture for decision-aware meta-review generation. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 222–225. IEEE.
- Kamal Sarkar. 2009. Sentence clustering-based summarization of multiple text documents. *TECHNIA-International Journal of Computing Science and Communication Technologies*, 2(1):325–335.