Retrieval-Augmented Analysis of Peer Reviews and Acceptance Trends

Halil Yasavul
Graduate School of Informatics
Middle East Technical University
Ankara, Türkiye
halil.yasavul@metu.edu.tr

***Abstract*—** This project explores retrieval-augmented methods to assist peer review processes. We retrieve similar papers using Sentence-BERT embeddings and FAISS, and summarize prior reviewer opinions using RAG-Token models. Additionally, we analyze acceptance and rejection rates among retrieved papers to support reviewer decision-making. Initial analyses on the ASAP-Review dataset confirm data suitability. Our approach aims to enhance peer review quality by providing historical context and acceptance trends.

Keywords— Retrieval-Augmented Generation, Peer Review Analysis, Sentence-BERT, FAISS, Acceptance Rate Analysis, Text Summarization.

I. Literature Review

Retrieval-Augmented Generation (RAG) is a method that combines retrieving information and generating text to answer questions or complete tasks better.

Lewis et al. [1] introduced a RAG model that searches for helpful documents and uses them together with the user's input to generate better answers. Their results showed improvements in many knowledge-heavy tasks compared to models that only generate text without retrieval.

Guu et al. [2] proposed a method called REALM, where the model learns how to retrieve information during training. This helps the model to find important documents by itself while learning language tasks. They showed that it works especially well when there isn't enough data.

Karpukhin et al. [3] presented Dense Passage Retrieval (DPR), which finds documents based on dense vector similarities, rather than simple keyword matching. This made retrieval faster and more accurate, and it can be used to improve the retrieval step inside RAG systems.

Based on these ideas, my project will use RAG to help peer reviewers by finding and summarizing opinions from past reviews on similar research topics.

II. PROPOSAL

The goal of this project is to check if retrieval-augmented generation can help reviewers by giving them a summary of other reviewers' thoughts about similar research topics.

A. Research Questions

- RQ1: Can retrieval-augmented generation improve peer review by summarizing opinions from previous reviews on similar topics?
- RQ2: Can retrieval from past reviews help predict or analyze the acceptance rates of papers on similar research topics?

B. Model Framework

For this project, we propose a two-stage framework to assist peer review processes by utilizing retrieval-augmented methods. The framework addresses two research questions separately as follows:

For RQ1 (Summarizing peer opinions):

- Embedding Model: Sentence-BERT model from Hugging Face Transformers is used to generate dense vector representations of paper titles and abstracts.
- Retriever: FAISS (Facebook AI Similarity Search) is used to retrieve the top-k similar papers and their associated reviews based on embedding similarity.
- Generator: The Hugging Face RAG-Token model (facebook/rag-token-nq) is employed to generate summaries of prior reviewer opinions by combining the query and retrieved documents.
- Frameworks: Hugging Face Transformers, FAISS, PyTorch, and Weights and Biases (for experiment tracking).

For RQ2 (Acceptance and rejection trend analysis):

- Embedding Model: Sentence-BERT model from Hugging Face Transformers is used to embed paper titles and abstracts.
- Retriever: FAISS-based dense retriever is used to identify similar papers from the dataset.
- Metadata Extraction: Acceptance and rejection decisions are extracted from the metadata associated with retrieved papers.
- Analysis Tools: Python standard libraries (pandas, collections, matplotlib) are used for acceptance/rejection counting and visualization.
- Frameworks: Hugging Face Transformers, FAISS, PyTorch, Python.

This two-stage approach allows both summarization of peer opinions and statistical analysis of acceptance trends, providing reviewers with broader historical context to assist in decision-making. As shown in **Figure 1**, the framework starts from preprocessing paper titles, followed by embedding, retrieval, metadata analysis, and final output generation.

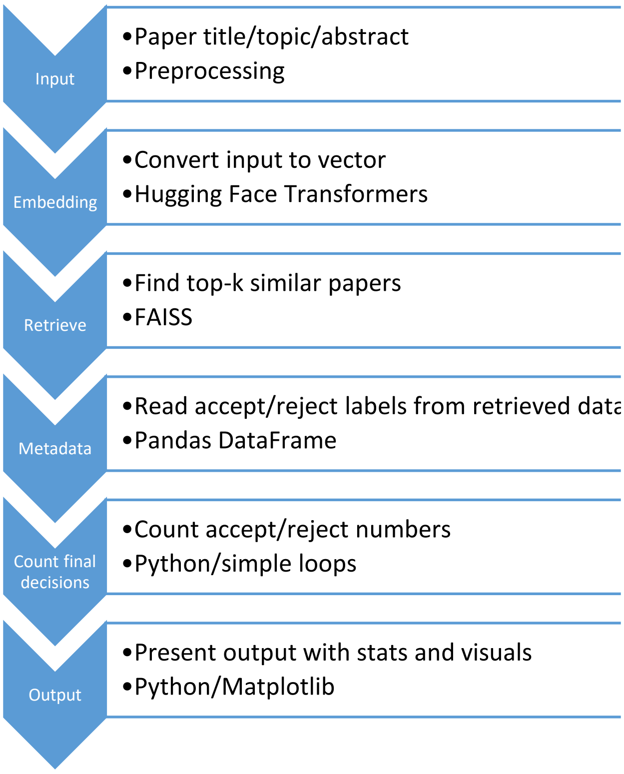

Figure 1. Workflow for retrieval-augmented analysis of peer reviews and acceptance trends.

C. Performance Metrics

- Retrieval: Recall@k (how many times a correct review appears in top-k results)
- Generation: ROUGE-L and BLEU (measuring how good the generated summaries are)

D. Expected Results:

- Create summaries that show different views from earlier peer reviews.
- Help reviewers by giving them a bigger picture of what others thought.
- Show that combining retrieval and generation works better than using generation alone.

E. Why It's Useful:

- Reviewers will be able to write better and more fair reviews.
- It may save time and make reviews more consistent.

III. EDA and Quality Checks

I've conducted descriptive analyses on the ASAP-Review dataset. The dataset includes 8,877 papers and 28,122 reviews from ICLR (2017–2020) and NIPS (2016–2019). Overall, 61% of the papers were accepted, with acceptance rates decreasing over time. Each paper has an average of 3.2 reviews, and the average review length is 374 words. Quality checks show 19 papers missing content metadata and 98 papers missing review files. Additionally, many reviews lacked explicit ratings or confidence scores, especially in older conferences. Overall, the dataset is suitable for retrieval and acceptance trend analysis.

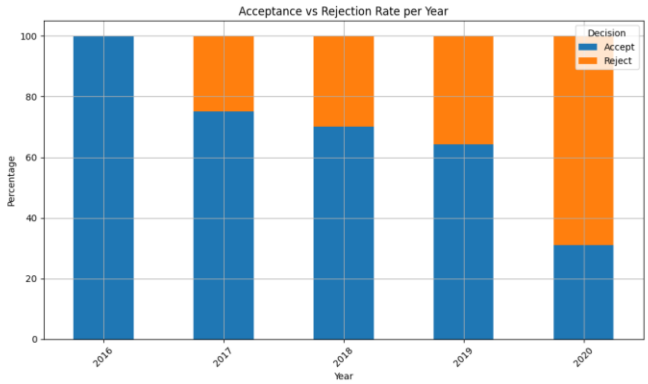

Figure 2. Acceptance vs Rejection per year

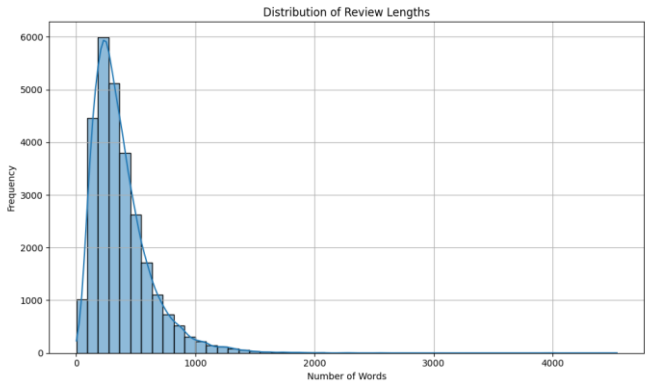

Figure 3. Distribution of review lengths

IV. Git Repository Initializition

A public GitHub repository has been initialized to track the project development. The repository follows a clean and reproducible structure, in line with project guidelines.

https://github.com/halilujah/asap-review-rag.git

References

[1] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Kuttler, M. Lewis, W. Yih, T. Rocktaschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proceedings of Advances in Neural Information Processing Systems (NeurIPS), 2020.

[2] K. Guu, T. Hashimoto, Y. Oren, and P. Liang, "REALM: Retrieval-augmented language model pre-training," in Proceedings of the International Conference on Machine Learning (ICML), 2020.

[3] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W. Yih, "Dense passage retrieval for open-domain question answering," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.