Scientific Paper Recommender

Abstract

In the recent years, recommender systems are emerging in many domains such as e commerce, music streaming, social media platforms etc.

The main motivation behind these is to enhance the user experience, save time, revenue and competitive advantage etc.

People working in research face difficulties(In terms of time consumption) to find papers/publications related to the field they are working on.

So, We propose a system that can recommend papers to achieve the same.
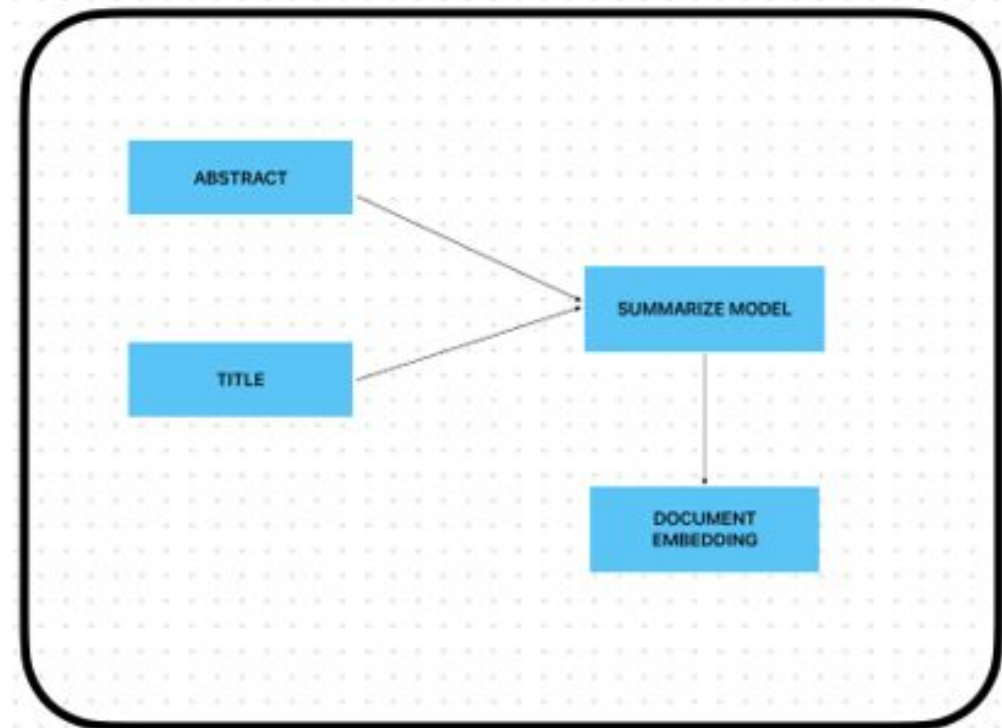
Model Training

- The abstract and title are summarized using pre-trained models.

Models that can be used here - **TextRank**, **BERTSum**, **T5**, **GPT**, etc.

- Feeding the summarized text into embedding model.

Models that can be used here – **Doc2Vec**, **BERT**, etc.

- Using these embeddings to find relevant papers.

Data

- We will be using Cornell University **arXiv** and AllenAI **SciREPEval** dataset.

- The Cornell University arXiv dataset is a collection of research papers in various fields, including computer science, physics, mathematics, and statistics and AllenAI on STEM.

- The papers in the dataset are in JSON format and contain information such as the paper title, author names, abstract, publication date, and references.

- The datasets combined contains over 2.2 million papers.

Conclusion

- To provide relevant research papers based on titles and abstract using different embeddings and models.

- Comparing performances for different approaches.

References

Datasets -
- [arXiv dataset](#)
- [ScirepEval dataset](#)

Embeddings -
- [SPECTER Embedding](#)
- [Doc2Vec](#)

Models -
- [BERTSum](#)
- [BERT](#)