Text Summarization Report

Dayou(Wally) Wu

4.21.2023

Part 1: Literature Review

As a popular research area within the field of natural language processing, automatic text summarization (ATS) encompasses numerous approaches across multiple dimensions. Broadly categorized, these approaches include Extractive Summarization, Abstractive Summarization, and Hybrid methods that combine elements of both. These methods are mainly evaluated based on the Rouge (Recall-Oriented Understudy for Gisting Evaluation) Score.

1. Extractive Approach

Method	**State-of-the-art Approach**	**Dataset**	**Evaluation Method**
Vector-Based	Darling (2010): Enhanced associations by employing a linear combination of unigram and bigram into vectors.	DUC 2004 &TAC 2010	ROUGE-2, ROUGESU-4
SVD-Based	Khurana and Bhatnagar (2022): Employed a Non-Negative Matrix Factorization (NMF) technique to reveal probability distributions.	DUC 2001, DUC 2002, CNN/Daily Mail	ROUGE-1, ROUGE-2 and ROUGE-L
Supervised machine learning-based	Liu and Lapata (2019): Modified BERT transformer (Devlin et al., 2018) for extracting the features in the internal layers.	CNN/Daily Mail	ROUGE-1, ROUGE-2, and ROUGE-L
Reinforcement learning-based methods*	Narayan et al. (2018a): Approached ATS as a sentence ranking problem.	CNN/Daily Mail	ROUGE-1, ROUGE-2, and ROUGE-L

*It is worth noting that RL-based method has the highest Rouge Score among extractive approaches.

2. Abstractive Approach

Linguistic-based abstractive approaches are not included in this article, as most of these methods are not trained based on dataset, but author's understanding of certain contexts. Other than linguistic approaches, the main methods used now are Sequence-to-sequence deep learning methods. It's an RNN model which first encodes the input text and the match an output text with the trained model and decodes it.

Method	State-of-the-art Approach	Dataset	Evaluation Method
BART	Lewis et al. (2019): It is a bidirectional encoder model with an autoregressive decoder architecture.	CNN/Daily Mail and XSUM	ROUGE-1, ROUGE-2, and ROUGE-L
Transformers (GPT-3 / PEGASUS)	Brown et al. (2020) / Zhang et al. (2020): The Transformer architecture utilizes self-attention mechanisms to capture long-range dependencies between words in a sequence, making it particularly effective for modeling sequential data such as text.	CNN/Daily Mail and XSUM	ROUGE-1, ROUGE-2, and ROUGE-L

Part 2/3: Implementation & Evaluation:

In this study, I tried to implement the BART model introduced by Lewis et al. and a Chrome extension based on this model. As most state-of-the-art Abstractive Summarization approaches use the Transformers model (explained above), I used the Hugging Face Transformers Library. Since the Transformers purpose is not for summarization directly. I fine-tuned it into a summarization model by training based on Large-CNN Daily Mail Dataset and generate the tokenizers of it. The work of it would be shown in the demonstration of Implementation and Chrome Extension. The table below listed the Rouge-1 and Rouge-L evaluation of the model based on a list of testing document:
https://en.wikipedia.org/wiki/Artificial_intelligence
https://en.wikipedia.org/wiki/Automatic_summarization

Wiki_Page	ROUGE-2 Precision	ROUGE-2 F-measure(F1)	ROUGE-L Precision	ROUGE-L F-measure
AI	0.5227	0.5542	0.5227	0.5542
ATS	0.5970	0.7143	0.6851	0.7872

The reference summary used for comparison here is the summary generated by the Wikipedia python database (wikipedia.summary("page", sentences =2)). From the result shown above, we could see that the BART model performs a high ROUGE Score which is even much higher than its performance shown in the ATS survey result (roughly 0.2541 for ROUGE-L of CNN corpus). Several reasons may make this happen:

1. Wikipedia content structure: Wikipedia articles typically have a well-organized structure, with the most important information presented at the beginning of the article. BART might be able to capture the key points effectively, leading to summaries that align well with the reference summaries generated by the Wikipedia Python library.
2. Reference summary: wikipedia.summary("page") function is not necessarily a human-generated summary but rather an extraction of the lead section of the Wikipedia article. The lead section is meant to provide an overview of the article and is often a good summary of the topic, but it is not specifically crafted as a summary by human editors.
3. Parameters: In my model, I used a max_length of 500 and num_beams (the number of beams) of 4 to fine-tune the BART model. It could be possible that the output summary with larger word limit captures more information than typical summaries, allowing it to better align with the reference summary.

Part 4: Conclusion

Through this practice, I learned about the state-of-the-art approaches of Automatic Text Summarization and how these methods perform in real life cases. In fact, it was beyond my expectations that the Transformers model did a great job in general text summarization. From my Chrome extension, I can clearly grasp the overall gist of the entire article from the obtained information, even though my implementation contains a significant amount of noise, which does not affect the overall generation of results. However, perhaps because my test mainly involves Wikipedia pages, my extension includes a large amount of extractive information, which is different from my initial expectation. In order to better leverage and adjust my extension's functionality, I believe it is possible to optimize the model by fine-tuning the BART model into different types according to specific requirements. On the other hand, preprocessing the content extracted from web pages can also contribute to generating better summaries.

Reference:

Cajueiro, Daniel O., et al. "A comprehensive review of automatic text summarization techniques: method, data, evaluation and coding." *arXiv preprint arXiv:2301.03403* (2023).

Yadav, Hemant, Nehal Patel, and Dishank Jani. "Fine-Tuning BART for Abstractive Reviews Summarization." *Computational Intelligence: Select Proceedings of InCITe 2022*. Singapore: Springer Nature Singapore, 2023. 375-385.

Lenka, Ranjita Kumari Biswal, et al. "Evaluation of Extractive and Abstract Methods in Text Summarization." *Data Science and Emerging Technologies: Proceedings of DaSET 2022*. Singapore: Springer Nature Singapore, 2023. 535-546.