nlp

by Chaithra Kopparam Cheluvaiah

News Summarization

Instructor: Dr. Lu Xiao

Group 4

Chaithra Kopparam Cheluvaiah, Shuangding Zhu,

Prateek Kumar Kumbar, Vinit Shambhu Horakeri

Text summarization is an example of NLP that will certainly have a huge impact on our lives. With the rise of digital media and the proliferation of publishing, who has time to read documents, news articles, or books to determine if they are beneficial.

Automatic Text Summarization (ATS) is a way to extract a short and coherent summary of text from a variety of sources, blogs, tweets, news stories, including books, emails, and research papers.
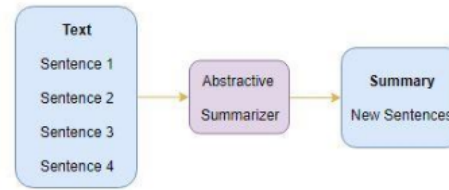
Two approaches to summarizing text:
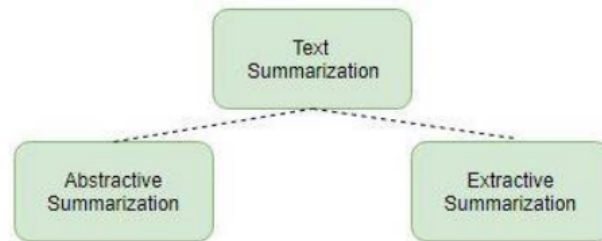
Abstractive Summarization

Abstractive Summarization produces more meaningful human written sentences, rather than being limited to terms from the source text.

Phrases based LSTM-CNN model

- Training data is human generated abstractive summary from news articles in DailyMail and CNN
- ROUGE toolkit is used for determining model performance
- LSTM-CNN model outperformed four previous models by 1-4%

The model development consists of 3 steps:
1. Data pre-processing
2. Phrase Extraction
3. Text summary generation (with the LSTM-CNN model)
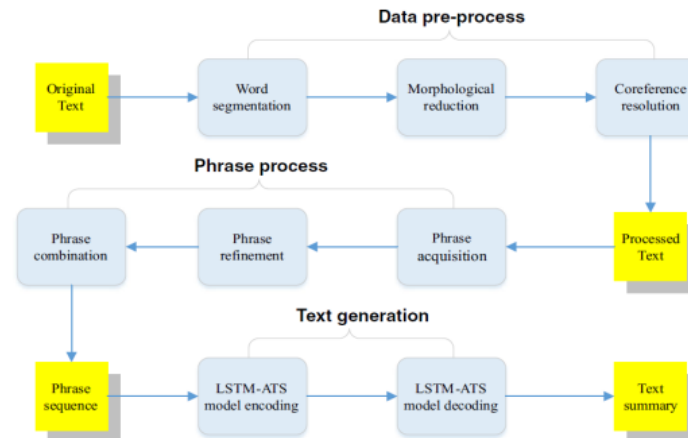
Extractive Summarization

Extractive summarization is a traditional way to generate summaries, such as clipping relevant chunks of the source text and combining to make a rational summary.
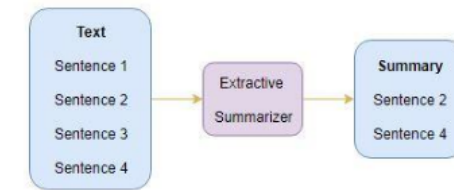
BERT for Extractive Text Summarization

- BERT model is already trained on massive datasets, no additional training is necessary.
- BERT aids in the unlocking of many semantics functions, such as interpreting the document's intent and constructing a similarity model between words.
- To achieve the greatest results in summarizing, it makes use of a powerful flat architecture with inter sentence transform layers.
- The most significant points of a document are thought to be represented by summary sentences.
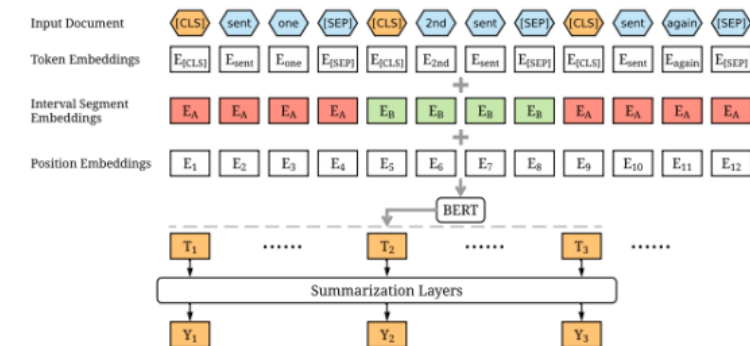
Conclusion

Extractive Text Summarization models focus on syntactic structure, whereas Abstractive Text Summarization models focus on semantics. The strengths of both summarization models should be included in our model. We believe abstractive text summarization is a powerful way of summarizing texts, and we will continue to work on the project with various approaches in the future.

References

Song, Shengli & Huang, Haitao & Ruan, Tongxiao. (2019). Abstractive text summarization using LSTM-CNN based deep learning. Multimedia Tools and Applications. 78. 10.1007/s11042-018-5749-3
https://iq.opengenus.org/bert-for-text-summarization/

nlp