Abstract

Extractive Summarization with Neural Networks

Virág Kulcsár

supervised by

Gábor Recski, Senior lecturer, Department of Automation and Applied Informatics

Gábor Borbély, Assistant lecturer, Department of Algebra

Summarization is the shortening of text without losing much of its information. In the news domain, there is a great demand for this task, as generating summaries by humans is expensive and slow.

Summarization has quite a history, due to being a long standing problem with numerous subquestions. One of these subquestions is how to evaluate summaries, since there is not one ultimate summary. We will talk about the problems surrounding the topic and the Rouge score, which is today's accepted method. Another unsolved question is how to successfully generate text, therefore our main focus was extractive summarization or highlighting, which means not paraphrasing the original text but rather selecting important sentences to create a summary.

The technique of extractive summarization has come a long way. At first, statistical methods were the most successful, then the graph based methods, and nowadays neural networks represent the state-of-the-art solutions. The `nnsum` is a framework of such neural networks that we used for our experiments. We used the averaging encoder and the Seq2Seq extractor on the CNN/Daily Mail data, and created our own extractive labels from the given abstractive summaries.

We trained the network with different labels to determine if we can think of summarization as a ranking problem instead of a classification problem. We saw that our models' results are close to the original's, and the performances of the new models are indistinguishable from each other, thus we concluded that further testing with different hyperparameters is needed to have a definitive answer.

2019