Read Top News First: A Document Reordering Approach for Multi-Document News Summarization

Chao Zhao, Tenghao Huang, Somnath Basu Roy Chowdhury, Muthu Kumar Chandrasekaran, Kathleen McKeown, Snigdha Chaturvedi

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

UNC Chapel Hill, Columbia University

Overview

Current concatenation-based multi-document news summarization (MDS) approaches neglect the relative importance of documents. In this work,

- We propose two simple yet effective approaches to reorder the input documents according to their relative importance before applying a summarization model;
- Our reordering approach significantly outperforms the state-of-the-art methods with more complex model architectures.

Motivation

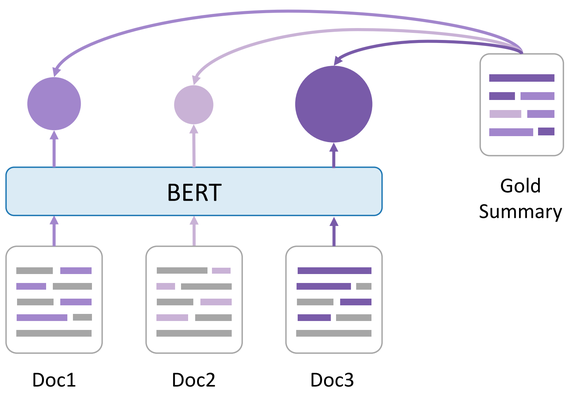
Recent approaches reformulate the problem of MDS as a single-document summarization (SDS) problem by concatenating all documents into a single meta-document and then using an SDS model to summarize it.
However, many SDS systems pay more attention to the beginning of the document. Therefore, in MDS, it is important to consider the order in which the documents are concatenated to form the meta-document before applying the SDS model.

Supervised Reordering

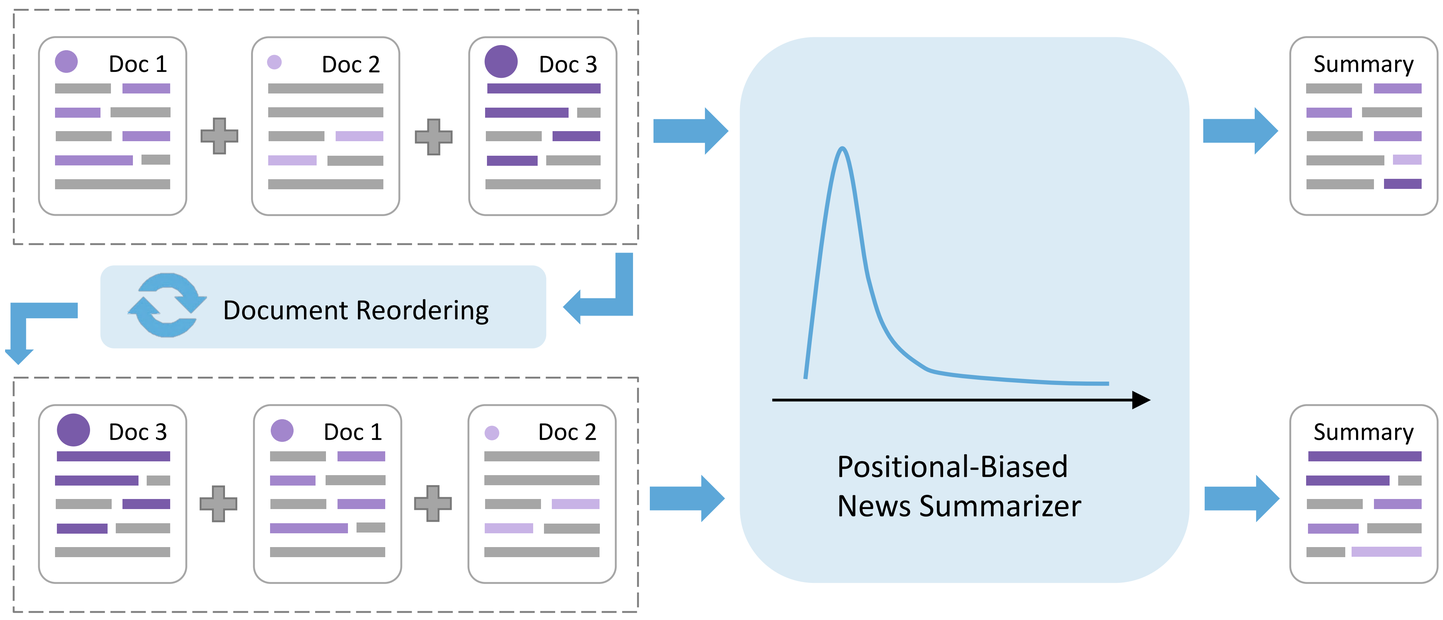
During training, we determine the oracle importance score of each document d_i as the normalized ROUGE-1 F score between d_i and the gold abstractive summary S:

$$y_i = \frac{\text{ROUGE}(d_i, S)}{\sum_i \text{ROUGE}(d_i, S)}. \quad (1)$$

Our learning objective is to minimize the Kullback–Leibler divergence divergence between the predicted distribution and the oracle distribution of importance scores.

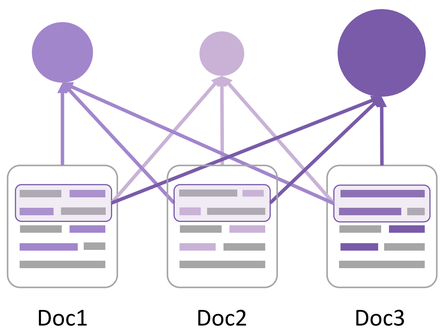


Document Reordering Framework



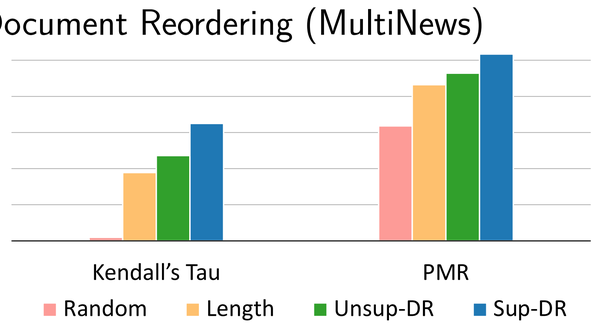
Unsupervised Reordering

To compute the centrality of a document d_i, we first represent the topic of the input cluster, T_i, by concatenating the top-3 sentences of each document except d_i, and then calculate the centrality as ROUGE(d_i, T_i). We choose top-3 sentences to represent the topic as it is a strong unsupervised summarization baseline.

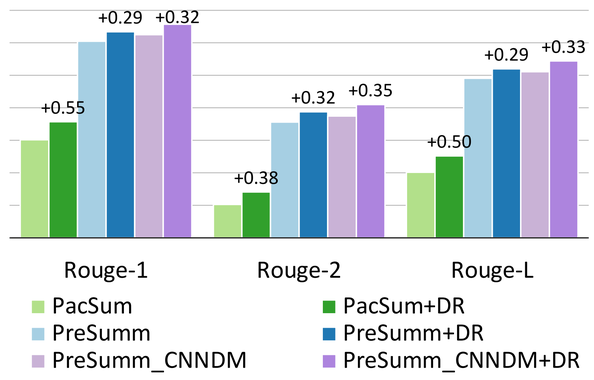


Results

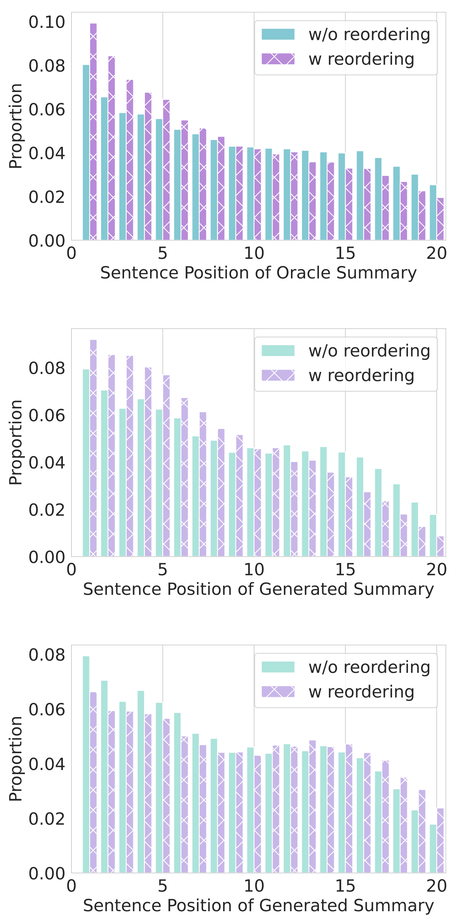
Document Reordering (MultiNews)



Document Summarization (MultiNews)



Analysis



Since the base summarization model pays more attention to the beginning of the input (Fig 2), by moving important content towards the beginning (Fig 1), the reordering method helps the summarization model also focus on information that was scattered across the original unordered input (Fig 3).

Contact Info

zhaochaocs.github.io/
zhaochaocs@gmail.com
zhaochaocs/MDS-DR

Paper Here

ACL 2022 @ Dublin, Ireland