since 1995

Data Science Natural Language Processing NLP Papers Summary

Day 118: NLP Papers Summary – Extractive Summarization Of Long Documents By Combining Global And Local Context

By Ryan 27th April 2020 No Comments

Objective and Contribution

Proposed a novel extractive summarisation model that combines both global and local context to summarise a single long document. In a given long document, it usually contains various topics. We believe that using these topics to guide summarisation could see improvement. We utilise both the global context (whole document) and the local context (section) to determine if a sentence is informative enough to be in the summary. The contributions are as follo

1. First to use LSTM-minus for text summarisation
2. The model achieved SOTA results on ArXiv and PubMed datasets and has shown to have an increase in performance as document length increase
3. Found that the performance of the model can be mainly attributed to the modelling of local context

Methodology

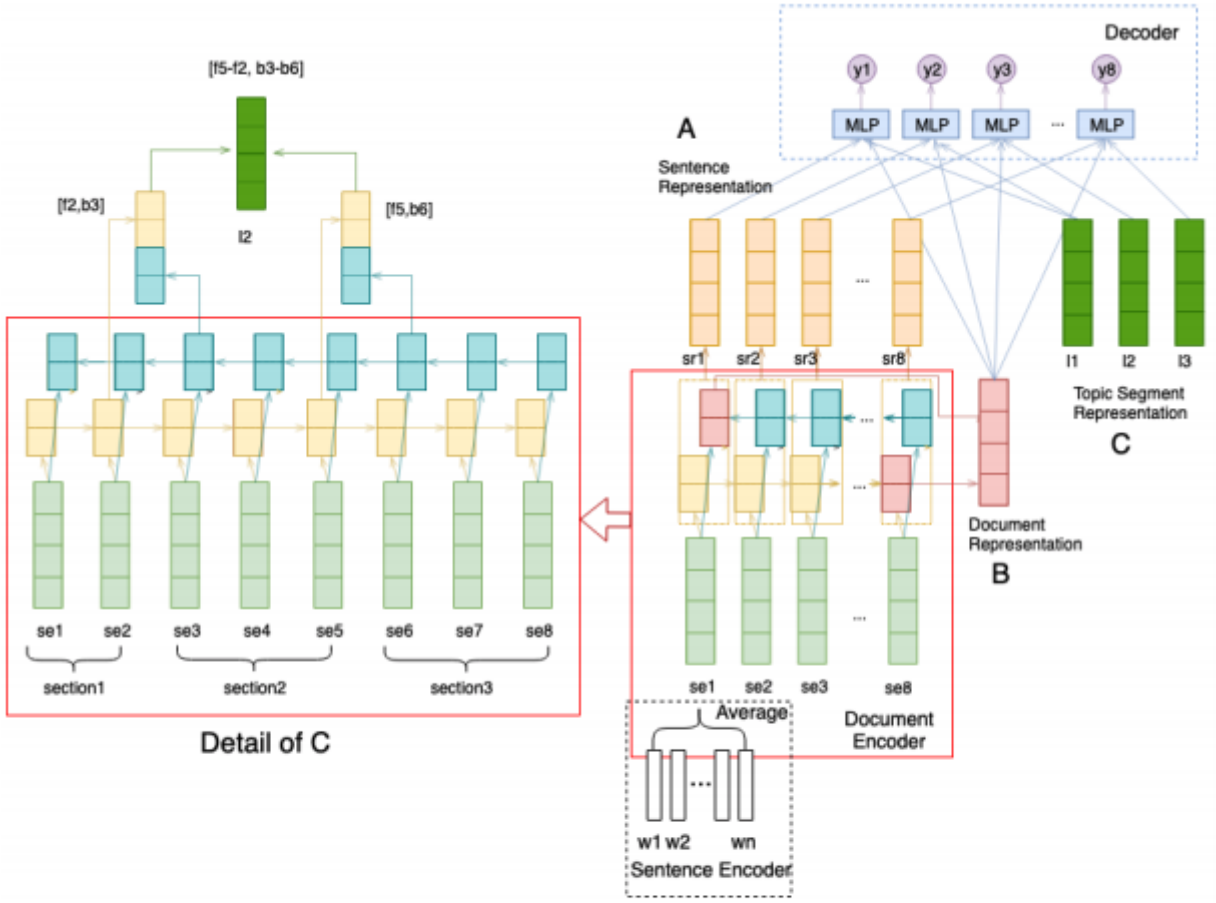

Figure 1: The structure of our model, se_i, sr_i represent the sentence embedding and sentence representation of sentence i, respectively. The binary decision of whether the sentence should be included in the summary is based on the sentence itself (A), the whole document (B) and the current topic (C). The document representation is simply the concatenation of the last hidden states of the forward and backward RNNs, while the topic segment representation is computed by applying LSTM-Minus, as shown in detail in the left panel (Detail of C).

The architecture consists of three components:

1. Sentence encoder
2. Document encoder
3. Sentence classifier

SENTENCE ENCODER

To compute the sentence embedding, we just use the simple method of averaging the word embeddings. This has shown to be just as effective as RNN and CNN when computing sentence embeddings. BERT sentence embedding performed poorly.

DOCUMENT ENCODER

The document encoder is a biGRU, which means that for each sentence there are two hidden states, a forward (yellow) and a backward hidden state (blue). This document encoder creates three outputs:

1. Sentence representation
2. Document representation
3. Topic segment representation

For each sentence, the sentence representation would simply be the concatenation of both the forward and backward hidden states. For document representation, it would be the concatenation of the final end state of both the forward and backward hidden states (in reds). For topic segment representation, we used LSTM-Minus. LSTM-Minus is used for learning text span embeddings. It was first proposed for dependency parsing, where a sentence is divided into three segments (prefix, infix, and suffix) and LSTM-Minus is used to create embeddings for each segment. It works as follows:

1. Apply an LSTM to the whole sentence to obtain hidden states of each word in the sentence
2. To create embeddings for each segment (from word i to word j), compute the difference between hidden state i and hidden state j. The idea behind this is that, given the nature LSTM, each hidden state contains information from previous hidden states plus current word, which allows the model to create segment embeddings using information outside and inside the segments
3. The final topic representation is the concatenation of the forward and backward segment embeddings (Detail of C in the figure above)

SENTENCE CLASSIFIER

With the sentence and document encoder, we now have sentence representations, segment representation, and topic segment representations. These three representations are feed into

the multi-layer perceptron (MLP) to predict whether the sentence should be included in the summary. Two methods of combining these three representations have been explored:

1. *Concatenation*. Simply concatenate all three representations together to form the final representation
2. *Attentive Context*. This is where we compute the weighted context vector for each sentence, allowing the model to learn how much weight to focus on the document and topic segment representations. This weighted context vector is concatenated with the respective sentence representation and feed into the MLP with sigmoid activation function to determine if the sentence should be included in the summary

Experimental Setup and Results

There are two evaluation datasets: arXiv and PubMed. We will be using ROUGE and METEOR as automatic evaluation metrics to compare our models with previous abstractive and extractive models.

MODELS COMPARISON

The models for comparison are split into different categories:

1. *Traditional extractive models*. SumBasic, LSA, and LexRank
2. *Neural abstractive models*. Attention Seq2Seq, Pointer-Generator Network, and Discourse-aware
3. *Neural extractive models*. Cheng & Lapata and SummaRuNNer
4. *Baseline, Lead, and Oracle*. Baseline is only feed sentence representation to MLP (without local or global context), lead is the first k words, and oracle measures the upper bound of the model's performance using ground-truth extractive labels

RESULTS

ROUGE-1 and ROUGE-2 are used to measure the informativeness of a summary whereas ROUGE-L measures the fluency of the summary. Table 2 showcase the results on the ArXiv dataset and table 3 showcase the results on PubMed dataset. As expected, the neural extractive models outperformed both the traditional extractive models and neural abstractive models in ROUGE-1 and ROUGE-2 by a huge margin. Results on ROUGE-L are mixed which could due to the fact that our models are trained on ROUGE-1 ground-truth extractive labels. The Discourse-aware abstractive model outperformed all the extractive models in ROUGE-L and this could due to them being trained on abstract summaries directly.

Low Lead results showcase no sentence position bias exist in scientific papers like in news articles although there might be other position biases that comes with scientific papers, which could be potential future work.

Our models outperformed all the other extractive models in all evaluation metrics. The increase in performance in our models from the baseline models demonstrate the benefit of including local and global context. The goal of our method is to effectively deal with summarising long documents. The figure below showcase that as the document length increase, our models have a good performance gain in comparison to current SOTA models.

ABLATION STUDY

The ablation study is to study the effect of including global and local context into our summarisation approach and assess which contribute more to the overall performance. From the results below, it seems that local topic context drives a strong improvement in the overall performance of the model. In contrast, there seems to be no significant benefit in including the global context. The reasons for this is left for future work.

Conclusion and Future Work

Potential future work would be to investigate methods to deal with redundancy. We could also integrate traditional methods such as feature engineering (e.g. sentence position, saliency) into our neural models. In addition, we could investigate a better structure in representing the document like discourse tree. Evaluation on summarisation has been a big issue in summarisation. ROUGE scores isn't as reliable as an evaluation metric and so it wou' good

to involve human evaluation in the future. Finally, an integration of both extractive and abstractive summarisation approach is worth exploring in the future.

Source: https://www.aclweb.org/anthology/D19-1298.pdf

Ryan

Data Scientist

Previous Post

Day 117: NLP Papers Summary - Abstract Text Summarization: A Low Resource Challenge

Next Post

Day 119: NLP Papers Summary - An Argument-Annotated Corpus of Scientific Publications

Data Science Natural Language Processing NLP Papers Summary

Day 365: NLP Papers Summary – A Survey on Knowledge Graph Embedding

Ryan
30th December 2020

Data Science Implementation Natural Language Processing

Day 364: Ryan's PhD Journey – OpenKE-PyTorch Library Analysis + code snippets for 11 KE models

Ryan
29th December 2020

Data Science Ryan's PhD Journey

Day 363: Ryan's PhD Journey – Literature Review – Knowledge Acquisition – 1st Passes XIV

Ryan
28th December 2020