

1. Citation

See, Liu and Manning (2017). Get To The Point: Summarization with Pointer-Generator Networks. arXiv: 1704.04368

2. Task

See, Liu and Manning use improved sequence-to-sequence method to implement an abstractive text summarization model to solve two problems that typical model encountered. First, the baseline model will produce wrong information. Second, the sentence in the summarizing may duplicate.

In this paper focus on multi-sentence summaries instead of reducing one or two sentences to a single headline, as they say in the paper that it is more challenging and more useful.

3. Data

They use CNN/Daily Mail dataset which contains online news articles and their summaries and preprocessed them with a script created by another paper to generate a dataset that has 287,226 training pairs, 13,368 validation pairs and 11,490 test pairs.

4. Approach

They build 2 text summarization models to compare the outcomes from the original news article to be summarized.

First, the baseline model uses Seq2Seq and Attention to generate the first problem they want to solve: wrong information and nonsensical sentence.

Second, they build an improved model called “pointer-generator” which is a mix of baseline model and pointer network to address the previous errors and found some glitch that it generates some repetitive sentence.

Therefore, they develop a unique mechanism that can be attached to either previous 2 models called “Coverage mechanism” which is an adaptation of coverage model. This add-on mainly solves the repetition problem from the seq2seq models.

5. Evaluation

The evaluation they mention in the paper is ROUGE or Recall-Oriented Understudy for Gisting Evaluation. This is a set of metrics that developed for evaluate the performance of text summarization work in NLP. As the dataset they are using provides a multi-sentence summarization along with the article of each news, they can easily compare the generated and provided summarization in such way.

The comparison among the models they built they also use METEOR score to see how well each model performs and test with different size of dictionary and different truncate of sentences.

They compared not only with the original summarization comes with the dataset, but also other abstract model they are referenced and found that their improved model gets at least 2 points higher in ROUGUE-1, ROUGUE-2, and ROUGUE-L score with a 95% confidence interval.