Paper Review

Title: CONTROLLABLE ABSTRACTIVE SENTENCE SUMMARIZATION WITH GUIDING ENTITIES

by

Anurag Saraswat (M20CS066)

Parsa Revanth (M20CS058)

Date: May 23, 2021

1 SUMMARY

In this paper the author proposes a controllable neural network for abstractive summarization with guideline entities. Author proposed a novel model which not only improved accuracy of content but also improved topic coherence. This is achieved by combining sentence representation with a named entity. As compared to the previous proposed model this model gives us guarantee regarding appearance of entities in the generated output summaries and it also adds more novel entities in the generated summary. To evaluate the informativeness of model , the author presented an informativeness metric which assesses the output quality in semantic vector level. This metric evaluates the model not only for model output and given input summary as label but also for extra and neglected information in summaries. Compared to lexicon based metrics like ROGUE metric, it is more human-like and helps to analyze results more thoroughly.

2 ARCHITECTURE USED IN THE PAPER

In this paper the author proposes a neural abstractive summarization model which leverages named entity information in original articles to generate informative summaries. The architecture of model is illustrated in below figure:

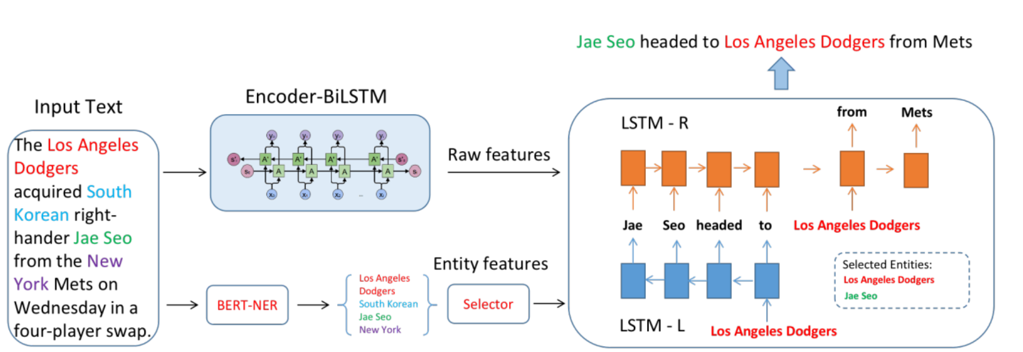

Fig. 1

Model uses an encoder-decoder framework. In the encoding phase , original article representation is extracted using Bi-LSTM layer. For identification of named entities BERT NER model is used. In decoding phases, two LSTM namely LSTM-R and LSTM-L are incorporated. LSTM-L generates the left part of the summary and LSTM-R generates the right part of the summary.
Details of each part is as follow:

2.1 Input Article Encoder

Basic Bi-LSTM is used to extract sequence features. Each token in the sentence is represented with word embedding.

2.2 Entity Encode

For each input article, a pretrained BERT model is incorporated for extracting entities.

2.3 LSTM-L

Till now , Entity is extracted for a given article. LSTM-L is incorporated for generating partial summary to the left part of an entity. Binary cross entropy of this stage is calculated and represented with $Loss_L$.

2.4 LSTM-R

In continuation to LSTM-L, LSTM-R is incorporated for generating partial summary to the right part of an entity. Binary cross entropy of this stage is calculated and represented with $Loss_R$.

2.5 Loss Calculation

Final loss is calculated as :

$Loss = \alpha * Loss_L + (1 - \alpha) * Loss_R$

3 EXPERIMENTS / TRAINING DETAILS

For evaluating performance of the model, English Gigaword corpus and DUC-2004 datasets are used. The pre-processed English Gigaword corpus dataset contains 3.8M training sentence-summary pairs and 1951 testing samples. Both datasets further reduce the input, output and entity vocabularies to 30k and less-frequent words are replaced to "UNK". Maximum article length is set to 50 and the left and right generation step is set to 15. The maximum length of the entity sequence is manually set to 3. For the input sequence without extracted entities, explicit [MID] token is used to replace the guiding entities[1].

For LSTMs, the hidden side is set to 256. Dropout is used on all non-linear connections with a dropout rate of 0.2. Training is done via the Adam optimizer with 1 = 0.9 and 2 = 0.99. Beam search of size 5 is incorporated to generate summaries. Model is implemented by the Tensorflow framework.

4 EVALUATION METRICS

Results are discussed on following metrics:

4.1 Automatic Metrics

It is to evaluate a model on ROGUE metric.

4.2 Human Evaluation

In this scheme randomly choose samples and generate summary using model and ask different evaluators to rate summaries.

4.3 Informativeness Metric

This metric is presented by the author in this paper. Given a candidate summary C and a ground-truth reference G, extract the context features c and g using a pre trained BERT model, respectively. In particular, when an output summary contains multiple sentences, average the sentence representations as the final features. The relevance R of candidate summaries and ground-truth references can be computed by the similarity on the euclidean metric[1]

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + \dots + (p_n - q_n)^2}$$
$$R = d(\mathbf{c}, \mathbf{g})$$

Fig. 2

Then, calculate the extraness of C by performing an orthogonal projection of c to g. The vertical vector c represent the irrelevant information[1].

$$\mathbf{c}_{\perp} = \mathbf{c} - \frac{\mathbf{c} \cdot \mathbf{g}}{||\mathbf{g}||^2}\mathbf{g}$$
$$E = d(\mathbf{c}_{\perp}, \mathbf{c})$$

Fig. 3

The measurement of omission is similar to the extraness, while performing an orthogonal projection of g to c.[1]

$$\mathbf{g}_{\perp} = \mathbf{g} - \frac{\mathbf{g} \cdot \mathbf{c}}{||\mathbf{c}||^2}\mathbf{c}$$
$$O = d(\mathbf{g}_{\perp}, \mathbf{g}),$$

Fig. 4

5 RESULTS

Author reported in paper that the presented model gives best performance against all state of the art models except for RG-2 on Gigaword. Models considered for comparison are Concept-Pointer Model , entity-based model Seq2Seq+E2T, GenParse-Full, ControlCopying.

6 CRITICAL DISCUSSION AND LIMITATIONS

Model is not able to beat other considered models on RG-2. Author does not provide analysis for the same. Maybe it depends on the nature of the corpus or due to other reasons but needed to be interrogated. Comparison of final result is present but author misses to present comparison of time complexity, memory and computational requirement with others model. Author also did not reveal information of hardware used like GPU, processor used for experiments etc.New metric presented by author Informativeness Metric which tests models on semantic vector level makes much more sense than lexicon based metric like ROUGE, BLEU.

7 CONCLUSIONS

Content of the paper is precise and has the following contribution. First, an abstractive summarization model which is more informative than previous works with the guiding entities. Experimental results demonstrate that model can generate entities with better fluency and coherence. Second, a fine-grained informativeness metric to evaluate models compared to previous state-of-the-art methods. There is a possibility to explore more in the comparison section rather than limiting to final results. Overall paper presented a model which makes much more sense for generating summary than previous models and opens up a new dimension of exploration.

REFERENCES

[1] C. Zheng, Y. Cai, G. Zhang, and Q. Li. Controllable abstractive sentence summarization with guiding entities. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5668–5678, Barcelona, Spain (Online), Dec. 2020. International Committee on Computational Linguistics.