Global Encoding for Abstractive Summarization

Junyang Lin[1,2] Xu Sun[2] Shuming Ma[2] Qi Su[1]

[1]School of Foreign Languages, Peking University

[2]MOE Key Laboratory of Computational Linguistics, Peking University

{linjunyang, xusun, shumingma, sukia}@pku.edu.cn

Abstract

- Problems in the seq2seq (repetition and semantic irrelevance);
- Our global encoding mechanism: **CNN and self attention**;
- Improved performances on the benchmark datasets;
- Generate summaries with **less repetition and higher semantic consistency** to the source text.

Sequence-to-Sequence as Baseline

- Encoder: RNN is more popular, usually LSTM and GRU
- Decoder: RNN for sequential decoding. Usually training is with teacher forcing.
- Attention mechanism: additive attention or global attention for the relevant source-side information

Example

- Problems in the seq2seq (repetition and semantic irrelevance);

Text: the mainstream fatah movement on monday officially chose Mahmoud abbas, chairman of the Palestine liberation organization (plo), as its candidate to run for the presidential election due on jan. #, ####, the official wafa news agency reported.

seq2seq: fatah officially officially elects abbas as candidate for candidate.

Gold: fatah officially elects abbas as candidate for presidential election.

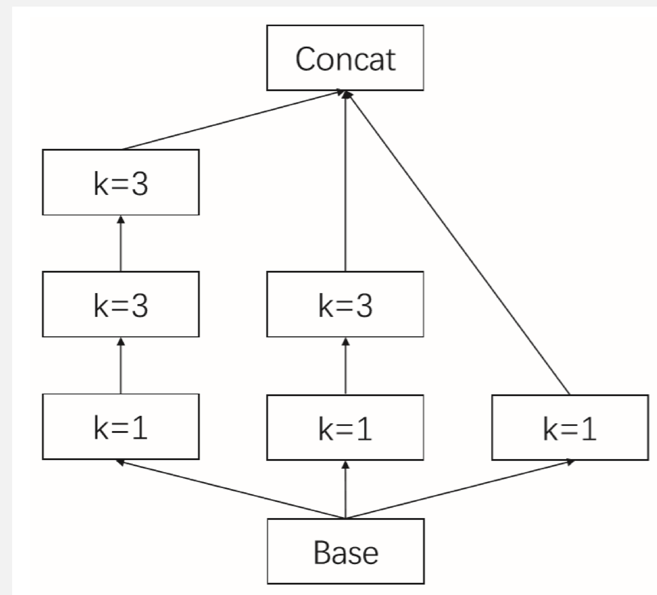
Problems

- **Noise** in the source context.
- **Relationship** between the source and the target is different from the alignment in machine translation, and correct alignment does not always indicate good summary.
- Source annotation at each time step **lacks global information of the context**, which may provide unnecessary information for summary.

Global Encoding

- **Convolutional Neural Networks** over the source annotations.
- **Self attention** for the connections to the global context.
- Collaboratively build a **gate** for the original source annotations.

CNN (Inception-like structure), Self Attention and Gate



Self Attention

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

Gate

$$\tilde{h} = h \odot \sigma(g)$$

Experiments

- Datasets:

LCSTS and **Gigaword**

Model	R-1	R-2	R-L
RNN	21.5	8.9	18.6
RNN-context	29.9	17.4	27.2
CopyNet	34.4	21.6	31.3
SRB	33.3	20.0	30.1
DRGD	37.0	24.2	34.2
seq2seq (Our impl.)	33.8	23.1	32.5
+CGU	**39.4**	**26.9**	**36.5**

Table 2: **F-Score of ROUGE on LCSTS.**

Model	R-1	R-2	R-L
ABS	29.6	11.3	26.4
ABS+	29.8	11.9	27.0
Feats	32.7	15.6	30.6
RAS-LSTM	32.6	14.7	30.0
RAS-Elman	33.8	16.0	31.2
SEASS	36.2	17.5	33.6
DRGD	**36.3**	17.6	33.6
seq2seq (Our impl.)	33.6	16.3	31.3
+CGU	**36.3**	**18.0**	**33.8**

Table 3: **F-Score of ROUGE on Gigaword.**

Qualitative Analyses

Source: 较早进入中国市场的星巴克，是不少小资钟情的品牌。相比在美国的平民形象，星巴克在中国就显得“高端”得多。用料并无差别的一杯中杯美式咖啡，在美国仅约合人民币12元，国内要卖21元，相当于贵了75%。第一财经日报 Starbucks, which entered Chinese market early, is a brand appealing to young people of petit bourgeoisie. Compared with its ordinary image in the United States, Starbucks seems to be of higher class in China. A Tall Americano sells about 12RMB in the United States, but 21RMB in China, which means it is 75% more expensive.	
Reference: 媒体称星巴克美式咖啡售价中国比美国贵75%。 Media report that the price of Starbucks Americano in China is 75% more expensive than that in the United States.	
seq2seq: 星巴克中国美式咖啡在中国。 Starbucks China Americano in China.	
+CGU: 星巴克美式咖啡中国贵75%。 Starbucks Americano is 75% more expensive in China.	

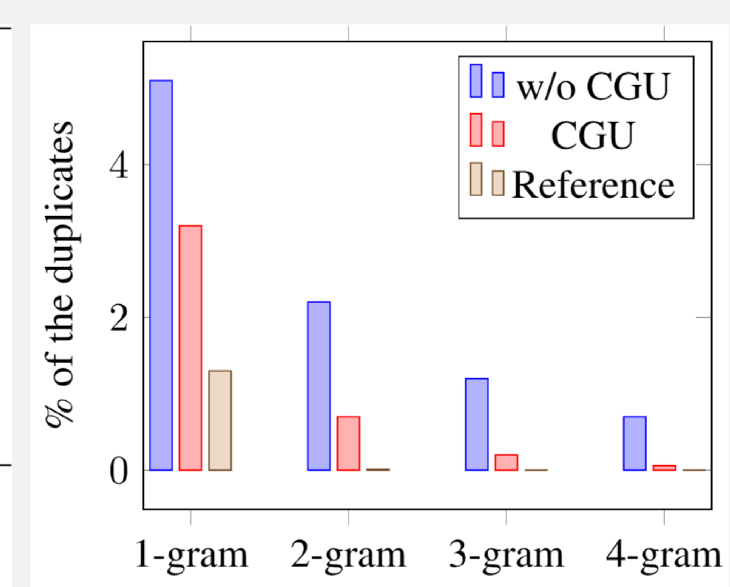


Figure above demonstrates the percentage of duplicates of n-gram

Conclusion

- Conventional Seq2Seq requires a mechanism to improve the source annotations so that they can provide summary-oriented information for the attention.
- Global encoding can improve the quality of generated summaries, which is reflected in both the ROUGE evaluation and the case study.
- It still requires future work to figure out what it filters and how it improves the performance of the model.