

Paper Reading

A Unified Model for Extractive and
Abstractive Summarization using
Inconsistency Loss

hukuda222

2020/1/15

Introduction

Models

Training Procedures

Exprements

Results

Conclusion

Introduction

Task

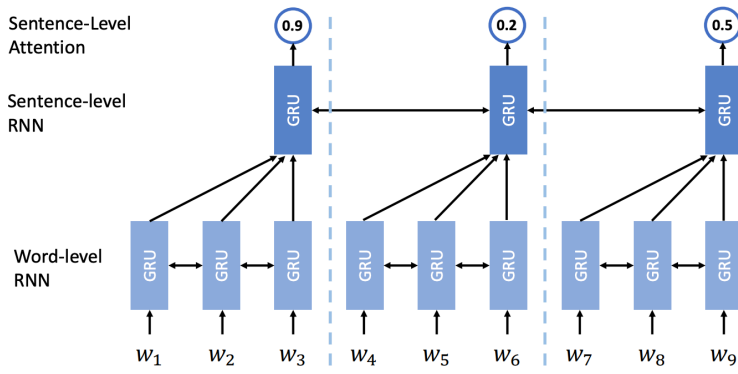
Generate summary from article
(CNN/daily mail).

Contribution

They combine the extractive model
and the abstractive model for
summarization.

They also introduced new loss
functions.

Hierarchical RNN



Hierarchical Attention

α_m^t : the weight of the t -th word in m -th sentence.

β_n : the probability of the n -th sentence been extracted into the summary.

$n(m)$: the position of the sentence with m -th word.

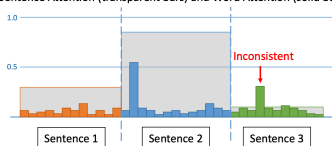
$$\hat{\alpha}_m^t = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}} \quad (1)$$

Inconsistency Loss

$$L_{inc} = -\frac{1}{T} \sum_{t=1}^T \log\left(\frac{1}{|\kappa|} \sum_{m \in \kappa} \alpha_m^t \times \beta_{n(m)}\right) \quad (2)$$

κ is the set of top K attended words and
 T is the number of words in the summary.

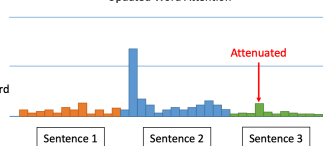
Sentence Attention (transparent bars) and Word Attention (solid bars)



Multiplying and
Renormalizing

→
Sentence and Word
Attentions

Updated Word Attention



Extractive Model

Their extractor is inspired by Nallapati et al. (2017).

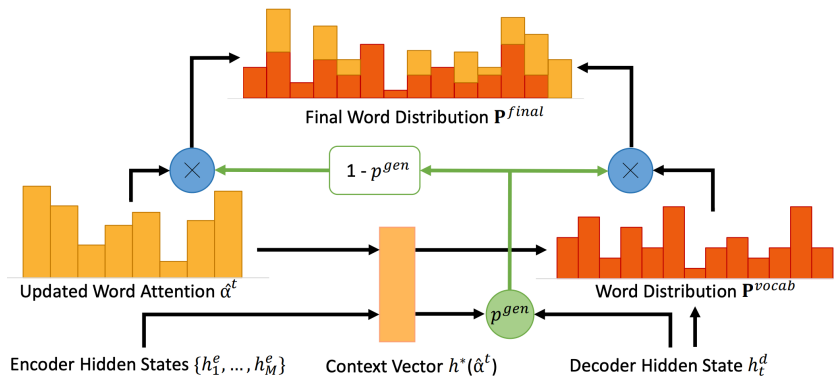
β_n : the probability of the n -th sentence being extracted into the summary.

g_n : the ground-truth label for the n -th sentence.

$g_n \in \{0, 1\}$

$$L_{ext} = -\frac{1}{N} \sum_{n=1}^N \log(g_n \log \beta_n + (1 - g_n) \log(1 - \beta_n)) \quad (3)$$

Abstractive Model



Abstractive Model

$$P_w^{final}(\hat{\alpha}^t) = p_{gen}(h^*(\hat{\alpha}^t))P_w^{vocab}(\hat{\alpha}^t) + (1 - p_{gen}(h^*(\hat{\alpha}^t))) \sum_{m:w_m=w} \hat{\alpha}_m^t \quad (4)$$

$$P^{vocab}(h^*(\hat{\alpha}^t)) = \text{softmax}(W_2(W_1[h_t^d, h^*(\hat{\alpha}^t)] + b_1) + b_2) \quad (5)$$

$$h^*(\hat{\alpha}^t) = \sum_m^M \hat{\alpha}_m^t \times h_m^e \quad (6)$$

The model can output OOV word.

Abstractive Model

\hat{y}^t : the t -th token in the target summary.

$$L_{abs} = \frac{1}{T} \sum_{t=1}^T \log P_{\hat{y}^t}^{final}(\hat{\alpha}^t) \quad (7)$$

$$L_{cov} = \frac{1}{T} \sum_{t=1}^T \sum_{m=1}^M \min(\hat{\alpha}^t, c_m^t) \quad (8)$$

$$c_m^t = \sum_{t'=0}^{t-1} \hat{\alpha}_m^{t'} \quad (9)$$

L_{cov} prevents repetition of the same word.

Two-stages training

1. train the extractive model.
2. train the abstractive model using the output from the extractive model as input.

End-to-end training

It is a method to train two models at the same time.

They use a weighted sum of the four error functions as the error function.

$$L_{e2e} = \lambda_1 L_{ext} + \lambda_2 L_{abs} + \lambda_3 L_{cov} + \lambda_4 L_{inc} \quad (10)$$

Setting

Dataset

Non-anonymized CNN/Daily Mail.

Pre-training for two-stages training

They pre-train two models.

The maximum number of sentence and token is 50

They limit the length of the source text to 400 and the length of the summary to 100.

Setting

Two-stages training

The abstracter takes extracted sentences with $\beta > 0.5$, where β is obtained from the pre-trained extractor.

Setting

End-to-end training

they will minimize four loss functions with $\lambda_1 = 5$ and $\lambda_2 = \lambda_3 = \lambda_4 = 1$.

They set K to 3 for computing L_{inc}
They increase the maximum length of source text to 600.

Setting

Detail

- ▶ Word embeddings size:128
- ▶ Vocabulary size:50k
- ▶ Hidden dimension:
200(extractor),256(abstractor).
- ▶ Summary length:
120 in the testing phase.

Results of Extracted Sentences

Method	ROUGE-1	ROUGE-2	ROUGE-L
pre-trained	73.50	35.55	68.57
end2end w/o inconsistency loss	72.97	35.11	67.99
end2end w/ inconsistency loss	78.40	39.45	73.83
ground-truth labels	89.23	49.36	85.46

Table 1: ROUGE recall scores of the extracted sentences. *pre-trained* indicates the extractor trained on the ground-truth labels. *end2end* indicates the extractor after end-to-end training with the abstracter. Note that *ground-truth labels* show the upper-bound performance since the reference summary to calculate ROUGE-recall is abstractive. All our ROUGE scores have a 95% confidence interval with at most ± 0.33 .

ROUGE recall has improved a lot.

Results of Abstractive Sentences

Method	ROUGE-1	ROUGE-2	ROUGE-L
HierAttn (Nallapati et al., 2016b)*	32.75	12.21	29.01
DeepRL (Paulus et al., 2017)*	39.87	15.82	36.90
pointer-generator (See et al., 2017)	39.53	17.28	36.38
GAN (Liu et al., 2017)	39.92	17.65	36.71
two-stage (ours)	39.97	17.43	36.34
end2end w/o inconsistency loss (ours)	40.19	17.67	36.68
end2end w/ inconsistency loss (ours)	40.68	17.97	37.13
lead-3 (See et al., 2017)	40.34	17.70	36.57

The model of the highest proposed method was SOTA at that time.

Results of Abstractive Sentences

Method	informativity	conciseness	readability
DeepRL (Paulus et al., 2017)	3.23	2.97	2.85
pointer-generator (See et al., 2017)	3.18	3.36	3.47
GAN (Liu et al., 2017)	3.22	3.52	3.51
Ours	3.58	3.40	3.70
reference	3.43	3.61	3.62

Their model scored higher than the reference in informativity and readability.

The effect of Inconsistency Loss

R_{inc} is the probability that the sentence with maximum weighted word have low attention.(i.e., $\beta_{n(\text{argmax}(\alpha^t))} < \bar{\beta}$ in each decoder step t).

$$R_{inc} = \frac{\text{Count}(t_{inc})}{T} \quad (11)$$

Method	avg. R_{inc}
w/o incon. loss	0.198
w/ incon. loss	0.042

Output

Original article (truncated):

A chameleon balances carefully on a branch, waiting calmly for its prey... except that if you look closely, you will see that this picture is not all that it seems. For the 'creature' poised to pounce is not a colourful species of lizard but something altogether more human. Featuring two carefully painted female models, it is a clever piece of sculpture designed to create an amazing illusion. It is the work of Italian artist Johannes Stoetter. Scroll down for video. Can you see us? Italian artist *Johannes Stoetter has painted two naked women to look like a chameleon. The 37-year-old has previously transformed his models into frogs and parrots* but this may be his most intricate and impressive piece to date. *Stoetter daubed water-based body paint on the naked models to create the multicoloured effect, then intertwined them to form the shape of a chameleon.* To complete the deception, the models rested on a bench painted to match their skin and held the green branch in the air beneath them. Stoetter can take weeks to plan one of his pieces and hours to paint it. Speaking about The Chameleon, he said: 'I worked about four days to design the motif bigger and paint it with colours. The body painting took me about six hours with the help of an assistant. I covered the hair with natural clay to make the heads look bald.' Camouflage job: A few finishing touches are applied to the two naked models to complete the transformation. 'There are different difficulties on different levels as in every work, but I think that my passion and love to my work is so big, that I figure out a way to deal with difficulties. My main inspirations are nature, my personal life-philosophy, every-day-life and people themselves.' However, the finished result existed only briefly before the models were able to get up and wash the paint off with just a video and some photographs to record it. (...)

Reference summary:

Johannes Stoetter's artwork features two carefully painted female models. The 37-year-old has previously transformed models into frogs and parrots. Daubed water-based body paint on naked models to create the effect. Completing the deception, models rested on bench painted to match skin.

DeepRL:

Italian artist Johannes Stoetter has painted female models to look like a chameleon. The 37-year-old has previously transformed his models into frogs and parrots but *this may be the work of Italian artist*. He has painted nude models and it is a clever piece of sculpture designed to create an amazing illusion. *It is work of artist Johannes Stoetter.*

GAN:

Italian artist Johannes Stoetter has painted two naked women to look like a chameleon. The 37-year-old has previously transformed his models into frogs and parrots but *this may be his most intricate and impressive piece to date.*

Pointer-generator:

Italian artist Johannes Stoetter has painted two naked women to look like a chameleon. *It is the work of Italian artist Johannes Stoetter.* Stoetter daubed water-based body paint on the naked models to create the multicoloured effect, then intertwined them to form the shape of a chameleon.

Our unified model (with inconsistency loss):

Johannes Stoetter has painted two naked women to look like a chameleon. The 37-year-old has previously transformed his models into frogs and parrots. Stoetter daubed water-based body paint on the naked models to create the multicoloured effect, then intertwined them to form the shape of a chameleon.

Conclusion

- ▶ They propose a unified model combining the strength of extractive and abstractive summarization.
- ▶ The inconsistency loss function is introduced to penalize the inconsistency between two levels of attentions.
- ▶ The loss enables extractive and abstractive summarization to be mutually beneficial.