Length Control in Abstractive Summarization by Pretraining Information

Yizhu Liu[1] **Qi Jia**[2] **Kenny Q. Zhu**[3*]
Shanghai Jiao Tong University
Shanghai, China
{[1]liuyizhu, [2]Jia_qi}@sjtu.edu.cn, [3]kzhu@cs.sjtu.edu.cn

—— ACL 2022

Code :https://github.com/yizhuliu/lengthcontrol

2022.10.2 • ChongQing

Reported by Wenhao Chen

1.Introduction

2.Method

3.Experiments

Introduction

Source Document
... iranians erupted in celebration as young people waved flags from their sunroofs , blasted music from stereos and chatted online with the hashtag #irantalks . the excitement came after a breakthrough nuclear deal with the united states and other world powers ...

Length	Reference Summary
10	iranians celebrate the deal online and in the streets .
30	after a breakthrough nuclear agreement deal with the united states and other world powers , celebration broke out in iranians . young people waved flages and chatted online .

Table 1: The reference summaries of one source document with lengths as 10 and 30.

The content of the reference summary with 10 tokens is the celebration of iranians. The reference summary with 30 to kens contains the reason for the celebration.

Introduction

Generated Summaries (Desired Length=10)
BART (Lewis et al., 2020) + Ad-hoc (Rush et al., 2015) (10 tokens)
iranians erupted in celebration as young people waved flags from
LenAtten (Yu et al., 2021) (12 tokens)
the agreement on the final day of persian new year festivities ,
LPAS (Saito et al., 2020) (22 tokens)
iranians erupted in celebration . the excitement came after a breakthrough nuclear deal with the united states and other world powers .

Table 2: The summaries generated by different models.

The summaries of ad-hoc and LenAtten in Table 2 are not complete and lose the information about “deal”.

Table 2 shows that LPAS con- tains redundant information about “deal” and its length is much longer than the reference summary.

Method

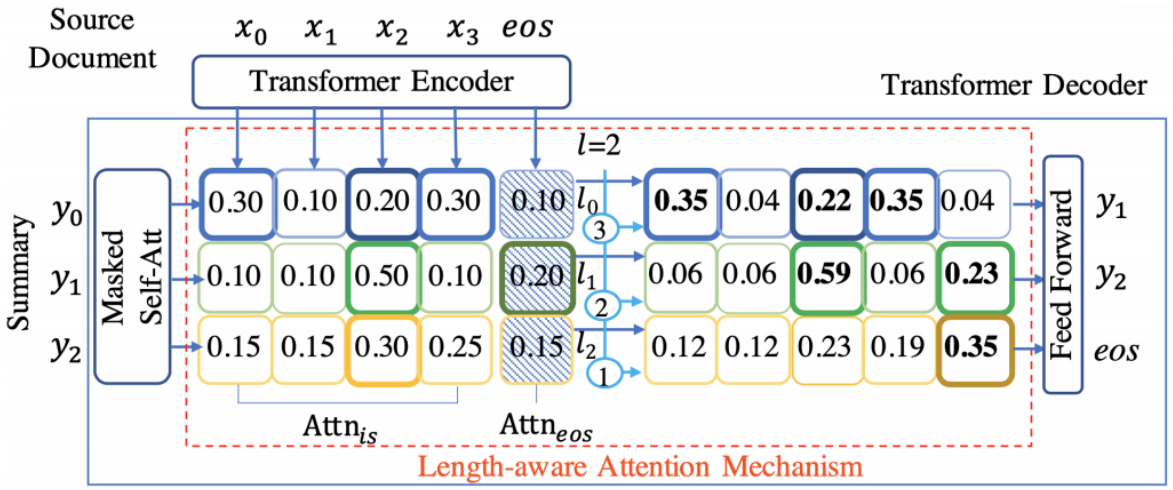

Figure 1: Overview of LAAM on Transformer Seq2seq. The bold values are boosted attention scores. The shadow boxes denote the attention scores of *eos*.

The model takes the source document $\mathbf{x} = (x_0, x_1, ..., x_m)$ and the desired length l as inputand the summary $\mathbf{y} = (y_0, y_1, ..., y_n)$ as output. x_m and y_n are *eos* tokens.

$$p(\mathbf{y}|\mathbf{x}, l) = \prod_{t}^{n} p(y_t|y_1, y_2, ..., y_{t-1}, \mathbf{x}, l) \quad (1)$$

$$\mathbf{h} = \{h_0, h_1, ..., h_m\},\ \mathbf{h} \in \mathbb{R}^{m \times d},$$
$$\mathbf{z} = \{z_0, z_1, ..., z_n\},\ \mathbf{z} \in \mathbb{R}^{n \times d}.$$

$$\mathbf{A} = \text{softmax}(\mathbf{z} \cdot \mathbf{h}^T) \quad (2)$$

where $\mathbf{A} \in R^{n \times m}$ is an attention matrix. $A_t = \{a_{t,0}, a_{t,1}, ..., a_{t,m}\}$ shows the attention scores of y_t. $a_{t,i}$ is the attention score between y_t and x_i.

Method

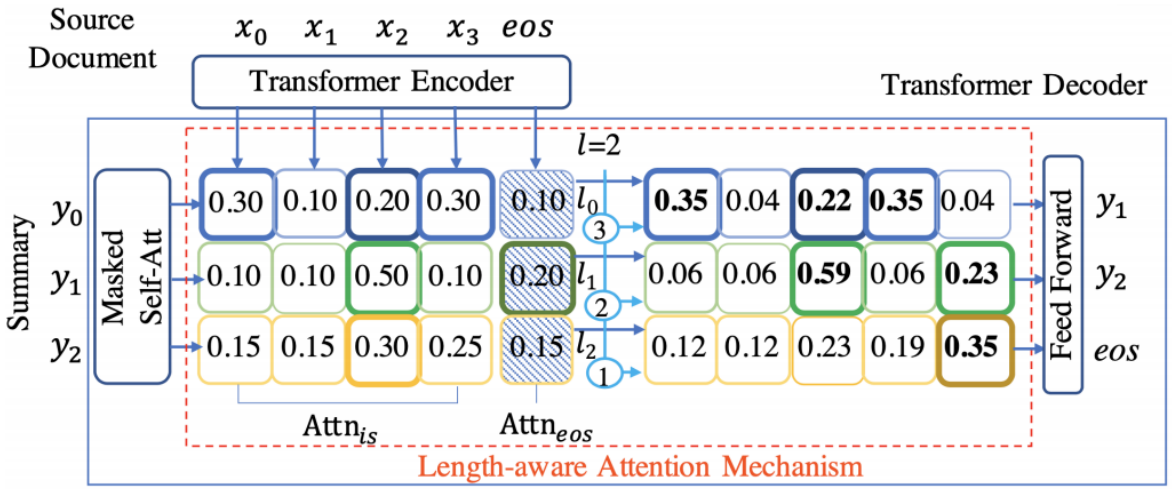

Figure 1: Overview of LAAM on Transformer Seq2seq. The bold values are boosted attention scores. The shadow boxes denote the attention scores of eos.

$\boldsymbol{Attn_{is}}$.

$$l_t = \begin{cases} l + 1 - t, & 0 \leq t \leq l \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

$$a'_{t,i} = w_{t,i} \times a_{t,i} \tag{4}$$

$$w_{t,i} = \begin{cases} 1, & p_i = 0 \\ l_t, & p_i = 1 \end{cases} \tag{5}$$

vector $\mathbf{p} = \{p_0, p_1, ..., p_m\}$ to label the indics of the top l_t tokens with the highest attention scores in A_t as 1 and others as 0.

Method

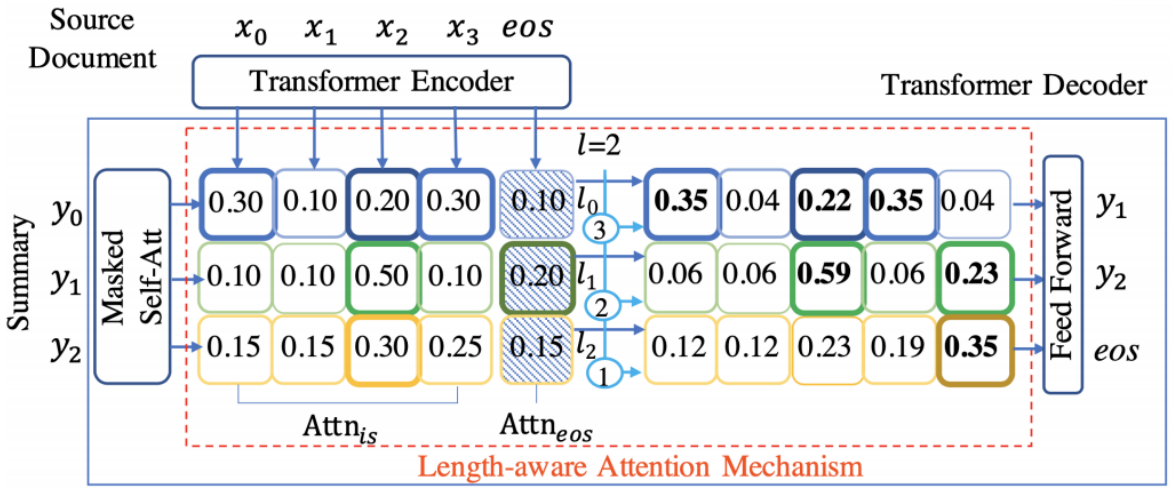

Figure 1: Overview of LAAM on Transformer Seq2seq. The bold values are boosted attention scores. The shadow boxes denote the attention scores of eos.

$\boldsymbol{Attn_{eos}}$.

$$a'_{t,m} = (l + 1 - l_t) \times a_{t,m} \quad (6)$$

$$p(y_t|y_{i<t}, \mathbf{x}, l) = \text{softmax}(W\mathbf{c}_{t-1} + b) \quad (7)$$

$$\mathbf{c_t} = \sum_0^m \tilde{a}_{t,i} h_i \quad (8)$$

$$\tilde{a}_{t,i} = \frac{a'_{t,i}}{\sum_{i=0}^m a'_{t,i}} \quad (9)$$

where W and b are trainable parameters.

Method

Algorithm 1 Creating Training Set of LBD

Input: the training set T
Output: the training set T'

$rec()$ computes the R-1 recall score between two texts.
$len()$ computes the length of token sequence.
for each training pair $(src, ref) \in T$ **do**
    $src = \{s_0, s_1, ...\}$, where s_t is the t^{th} sentence in src.
    for $i = 0 \rightarrow k$ **do**
        min and max denote minimum and maximum length of length range b_i, respectively.
        $e_i \leftarrow \emptyset$
        while $S = \{s | s \in src \cap len(e_i \cup s) \leq max\}$ **do**
            Select the s_{sel} with best $rec(e_i \cup s_{sel}, ref)$ from S.
            if $rec(e_i \cup s_{sel}, ref) > rec(e_i, ref)$ **then**
                $e_i \leftarrow s_{sel}$; $src \leftarrow src - s_{sel}$
            else
                break
        if $len(e_i) > min$ **then**
            Add $(src, e_i, rec(e_i, ref))$ to $S(b_i)$.
$S(b_i) \leftarrow$ top $\lceil |T|/k \rceil$ samples from $S(b_i)$ sorted by $rec(e_i, ref)$
$T' \leftarrow S(b_1) \cup S(b_2) \cup \cdots \cup S(b_k)$
return T'

we propose a heuristics to create a length-balanced dataset(LBD) by extracting summaries with various lengths from each document in original dataset and makeing lengths of these extractive summaries evenly distributed in different ranges.

Experiments

	CNNDM			XSUM		
	R-1	R-2	R-L	R-1	R-2	R-L
BART [4]	43.13	20.05	39.32	44.61	21.19	36.00
LenEmb	32.74	13.78	24.50	28.45	8.92	23.13
LC	35.45	14.50	26.02	31.87	11.23	25.94
GOLC	38.27	16.22	34.99	32.94	14.38	26.11
LenAtten	39.82	17.31	36.20	37.20	16.05	31.24
LPAS	42.55	20.09	39.36	43.64	19.81	35.22
BLPAS	42.95	20.29	39.76	44.94	20.31	35.98
LAAM	43.55	20.44	40.63	45.30	21.77	36.64
PtLAAM	**44.17**	**20.63**	**40.97**	**45.48**	**21.80**	**36.84**

Table 5: Gold length test with soft length control. The LAAM and PtLAAM are statistically significantly better than BLPAS with $p<0.05$ according to t-test.

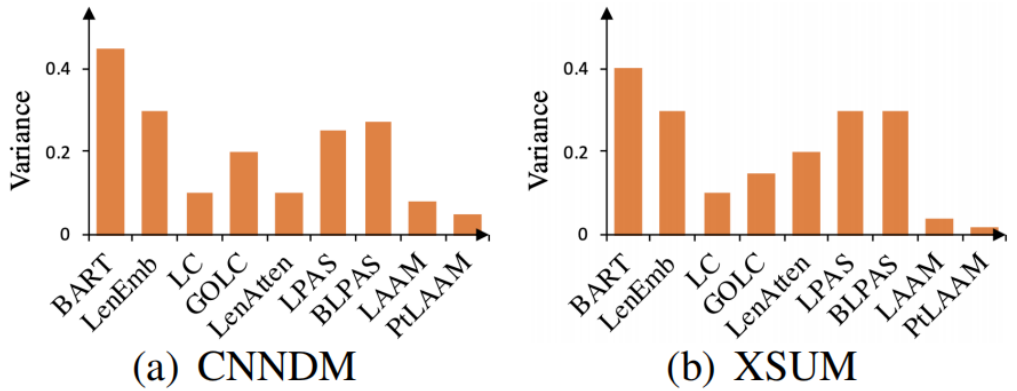

Figure 2: Variance of generated summary lengths in gold length test with soft length control.

Experiments

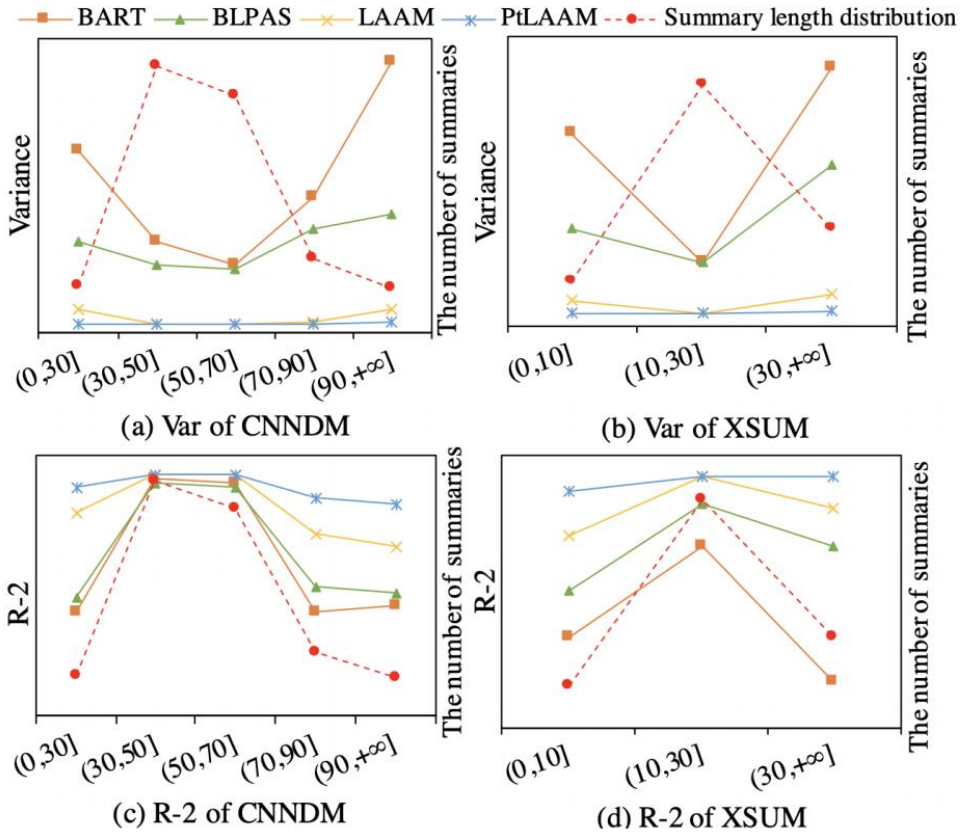

Figure 3: Var and R-2(F1) scores of gold length test with soft length control on divided test sets.

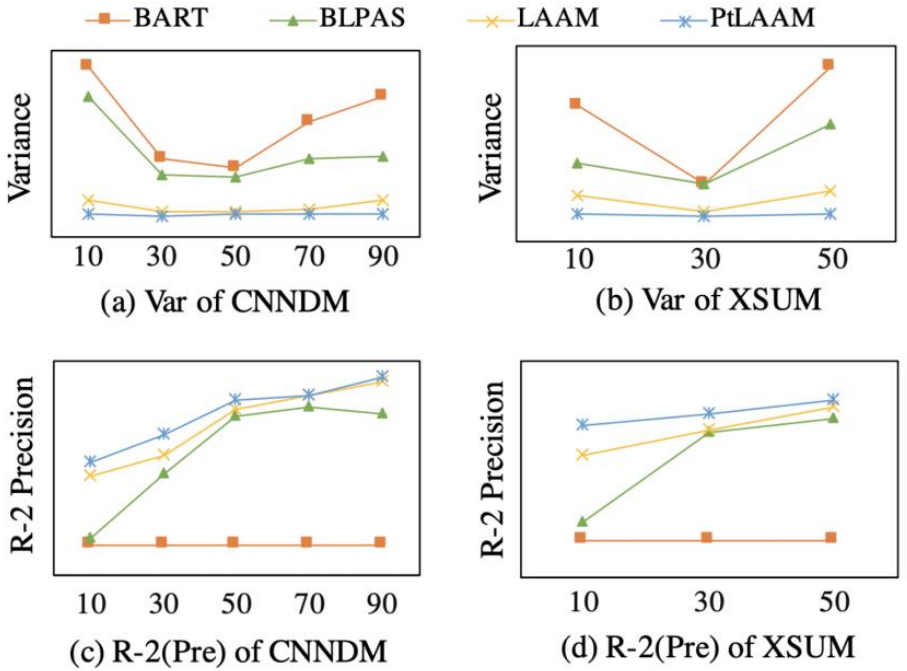

Figure 4: Var and R-2 (Pre) of arbitrary length test with soft length control on complete test sets.

Experiments

Approach	CNNDM			XSUM		
	R-1	R-2	R-L	R-1	R-2	R-L
BART	43.43	20.11	39.52	44.82	21.34	36.23
BLPAS	43.15	20.52	40.01	45.03	20.57	36.02
LAAM	43.63	20.76	40.63	45.38	21.77	36.64
PtLAAM	**44.21**	**20.77**	**40.97**	**45.53**	**21.82**	**36.85**

Table 7: The ROUGE scores of models in gold length test with hard length control.

Len	BLPAS Summaries	PtLAAM Summaries
10	iranians erupted in celebration , as young people waved flages	iranians celebrate online and in the streets after deal .
30	iranians erupted in celebration as young people waved flages , blasted music from stereos and chatted online . the agreement on the final day of persian new year festivities .	the excitement came after a breakthrough nuclear deal with the united states and other world powers . iranians erupted in celebration as young people waved flags and chatted online .

Table 8: Generated summaries of two different lengths from the source document in Table 1.

Experiments

Data	Model	R-1	R-2	R-L	Var(%)
CNNDM	LAAM	**43.63**	**20.76**	**40.63**	**0.05**
	w/o $Attn_{is}$	42.77	19.32	39.13	0.06
	w/o $Attn_{eos}$	43.10	20.17	37.45	0.13
XSUM	LAAM	**45.38**	**21.77**	**36.64**	**0.03**
	w/o $Attn_{is}$	43.45	20.64	34.79	0.03
	w/o $Attn_{eos}$	44.62	21.32	35.03	0.08

Table 9: Usefulness of two kinds of attentions.

Dataset	Length	Approach	R-1	R-2	R-L	Var(%)
Soft length control						
CNNDM	(0, 30]	BLPAS	33.04	14.83	29.42	0.14
		LAAM	33.52	15.20	30.54	0.05
		PtLAAM	**33.65**	**15.77**	**31.26**	**0.03**
XSUM	(0, 10]	BLPAS	34.37	19.54	31.66	0.10
		LAAM	34.49	20.07	32.10	0.03
		PtLAAM	**35.16**	**20.55**	**32.47**	**0.02**
Hard length control						
CNNDM	(0, 30]	BLPAS	30.25	12.51	26.98	-
		LAAM	33.64	15.23	30.76	-
		PtLAAM	**33.78**	**15.89**	**31.30**	**-**
XSUM	(0, 10]	BLPAS	32.55	17.16	29.52	-
		LAAM	34.83	20.15	32.10	-
		PtLAAM	**35.16**	**20.58**	**32.49**	-

Table 10: Results of zero-shot length control.

Thank you!