

# Context-Aware NMT Learns Anaphora Resolution

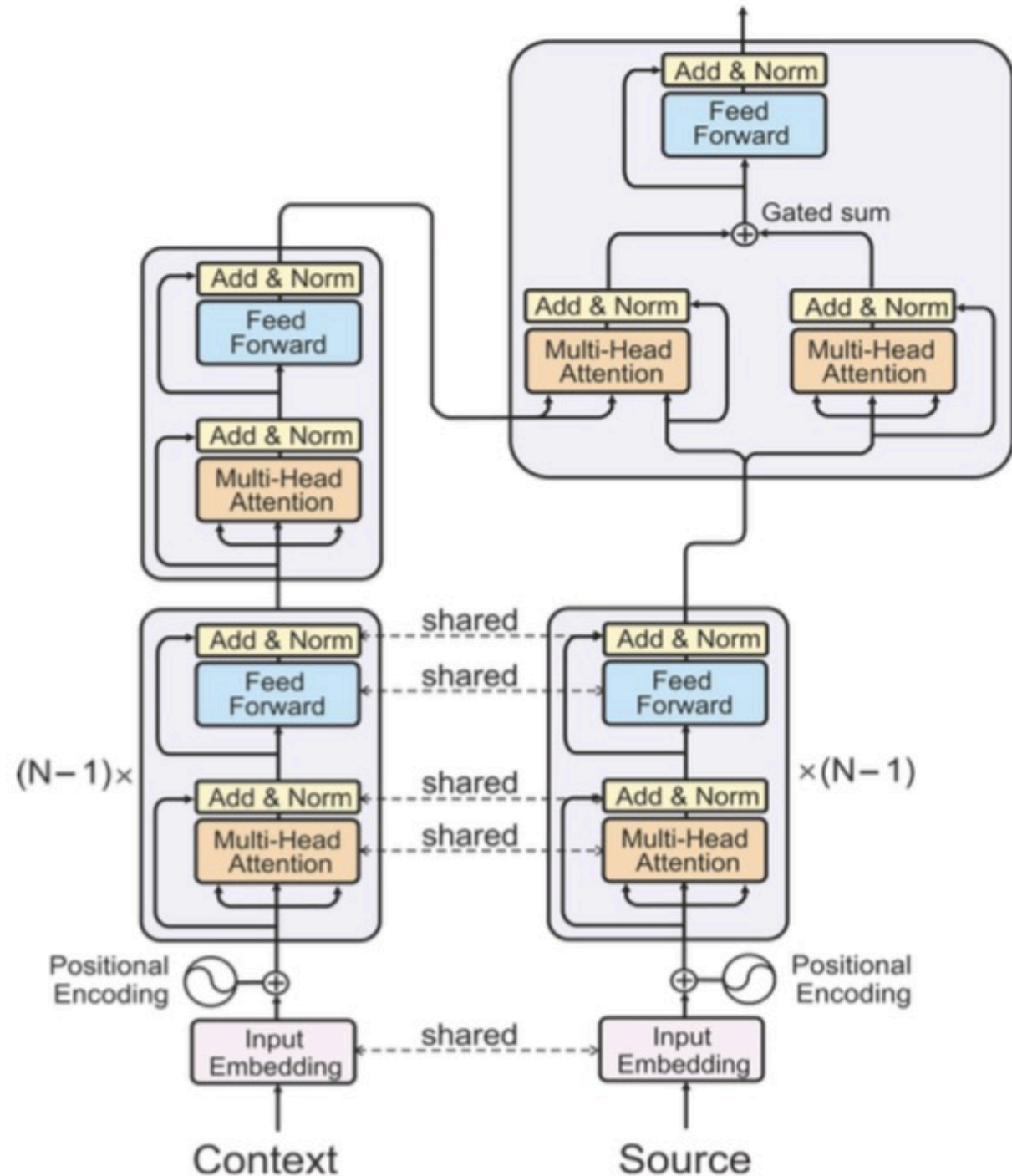
Pressentor: Baosong Yang

# Motivation

- ▶ Discourse phenomena is important in translation
- ▶ Using extended context, beyond the single source sentence, should be beneficial in ambiguous cases.
- ▶ construct a simple discourse-aware model

# Architecture

- Source Encoder: The last layer incorporates contextual information using a gate.
- Context Encoder:
  - using separate encoders does not yield an accurate model.
  - share the parameters of the first  $N - 1$  layers with the source encoder.
  - add <bos> to context sentences, but not source sentence so as to distinguish.



# Experiment

- ▶ Data: OpenSubtitles2018 corpus (2M/1w/1w)
- ▶ Setting: Base model

model	BLEU
baseline	29.46
concatenation (previous sentence)	29.53
context encoder (previous sentence)	<b>30.14</b>
context encoder (next sentence)	29.31
context encoder (random context)	29.69

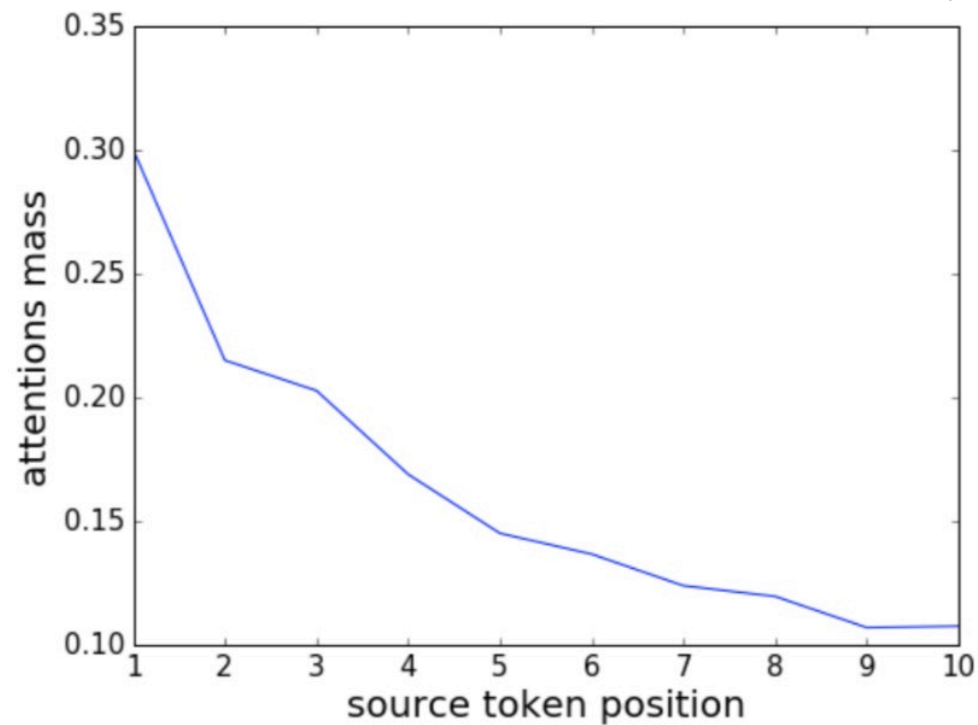
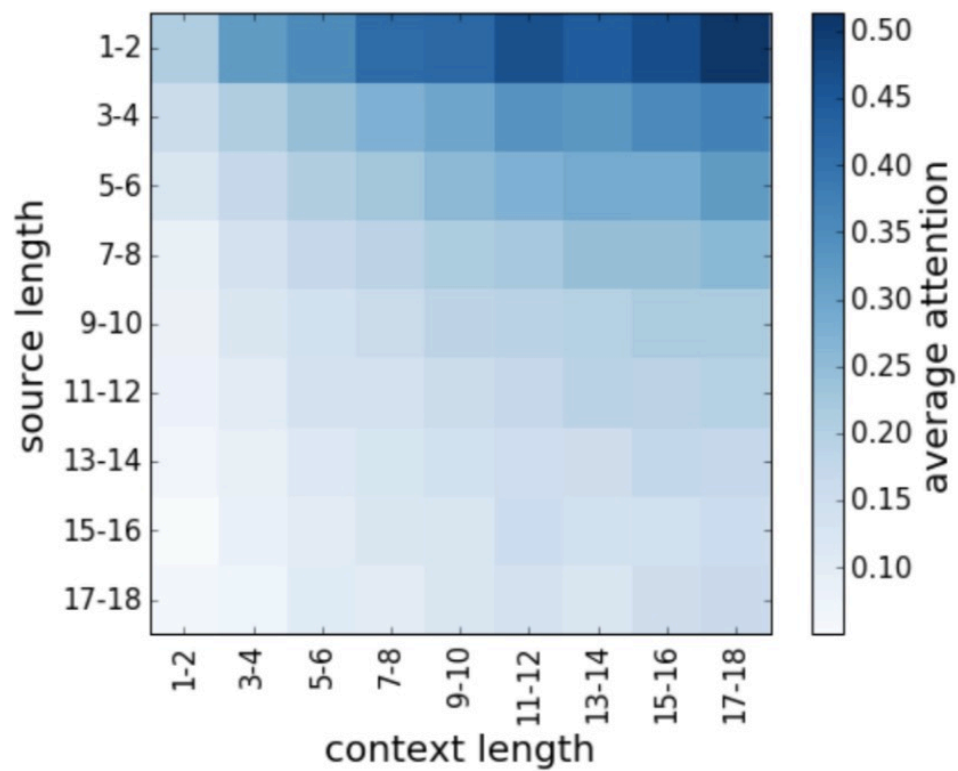
# Analysis

- For the translation of which words does the model rely on contextual history most

word	attn	pos	word	attn	pos
it	0.376	5.5	it	0.342	6.8
yours	0.338	8.4	yours	0.341	8.3
yes	0.332	2.5	ones	0.318	7.5
i	0.328	3.3	'm	0.301	4.8
yeah	0.314	1.4	you	0.287	5.6
you	0.311	4.8	am	0.274	4.4
ones	0.309	8.3	i	0.262	5.2
'm	0.298	5.1	's	0.260	5.6
wait	0.281	3.8	one	0.259	6.5
well	0.273	2.1	won	0.258	4.6

# Analysis

- Dependence on sentence length and position



# Analysis

## ► Analysis of pronoun translation

<b>pronoun</b>	<b>N</b>	<b>#pronominal antecedent</b>	<b>baseline</b>	<b>our model</b>	<b>difference</b>
it	11128	6604	25.4	26.6	<b>+1.2</b>
you	6398	5795	29.7	30.8	<b>+1.1</b>
yours	2181	2092	24.1	25.2	<b>+1.1</b>
I	8205	7496	30.1	30.0	-0.1

<b>word</b>	<b>N</b>	<b>baseline</b>	<b>our model</b>	<b>diff.</b>
it	4524	23.9	26.1	<b>+2.2</b>
you	693	29.9	31.7	<b>+1.8</b>
I	709	29.1	29.7	<b>+0.6</b>

Table 4: BLEU for test sets of pronouns having a nominal antecedent in context sentence. *N*: number of examples in the test set.

<b>type</b>	<b>N</b>	<b>baseline</b>	<b>our model</b>	<b>diff.</b>
masc.	2509	26.9	27.2	<b>+0.3</b>
fem.	2403	21.8	26.6	<b>+4.8</b>
neuter	862	22.1	24.0	<b>+1.9</b>
plural	1141	18.2	22.5	<b>+4.3</b>

Table 5: BLEU for test sets of pronoun “it” having a nominal antecedent in context sentence. *N*: number of examples in the test set.



# Conclusion

- ▶ Components share parameters with encoder.
- ▶ Discourse-Context is beneficial in ambiguous cases.
- ▶ More linguistic analysis may be interesting.