

# Paper Reading

Xinwei Geng

2018-10-23

# Beyond Error Propagation in Neural Machine Translation: Characteristics of Language Also Matter

Lijun Wu, Xu Tan, Di He, Fei Tian, Tao Qin, Jianhuang Lai and Tie-Yan Liu  
EMNLP 2018

# Accuracy drop and error propagation

- Accuracy drop
  - in left-to-right decoding, the right part words in its translation results are usually worse than the left part words in terms of accuracy
- Error propagation
  - if a word is mistakenly predicted during inference, the error will be propagated and the future words conditioned on this one will be impacted
- Two questions
  - Is error propagation the main cause of accuracy drop ?
  - Are there any other causes leading to accuracy drop ?

# Error Propagation is Not the Only Cause

- When translating from right-to-left, the accuracy of the left half is still higher than the right half
- Contradictory to the previous analyses, which regard that accumulated error brought by exposure bias will deteriorate the quality in later part of translation

<b>left-to-right</b>	De-En	En-De	En-Zh
Left	<b>10.17</b>	<b>7.90</b>	<b>7.41</b>
Right	8.39	6.60	5.91
<b>right-to-left</b>	De-En	En-De	En-Zh
Right	7.83	6.45	5.77
Left	<b>9.41</b>	<b>7.11</b>	<b>7.01</b>

# The Influence of Error Propagation

- Exposure bias exists
- Error propagation does exist
- Other causes exist

De-En	left-to-right			right-to-left		
	0	1	Δ	0	1	Δ
Left	<b>10.17</b>	<b>10.71</b>	0.54	<b>9.41</b>	<b>10.41</b>	1.00
Right	8.39	9.25	0.86	7.83	8.45	0.62

# Language Branching Matters

- Language branching is a main cause of accuracy drop
  - Right-branch language: the main subject of the sentence is described first, and is followed by a sequence of modifiers that provide additional information about the subject
  - Left-branch language: put such modifiers to the left of the sentence

		left-to-right		right-to-left	
		0	1	0	1
left	7.90	9.91	7.45	8.95	
	<b>8.70</b>	<b>11.52</b>	<b>9.24</b>	<b>10.59</b>	

Table 4: BLEU scores on En-Jp test set. "0" represents the normal translation results, and "1" represents the teacher-forcing translation results.

# Correlation between Language Branching and Accuracy Drop

- The 2/3-gram frequency as well as the conditional probability of the left half is higher than that of the right half for right-branching languages
- For left-branching language Japanese, the result is opposite

	De-En		En-De	
	2-gram	3-gram	2-gram	3-gram
$F_n^l$	<b>5713.8</b>	<b>3122.7</b>	<b>13811.8</b>	<b>687.1</b>
$F_n^r$	3026.5	1377.6	11692.2	419.9
$RF_n^{l>r}$	<b>59.6%</b>	<b>55.8%</b>	<b>53.8%</b>	<b>53.6%</b>
$RF_n^{l<r}$	38.8%	37.6%	46.0%	45.0%
$\Delta$	20.8%	18.2%	7.8%	8.6%

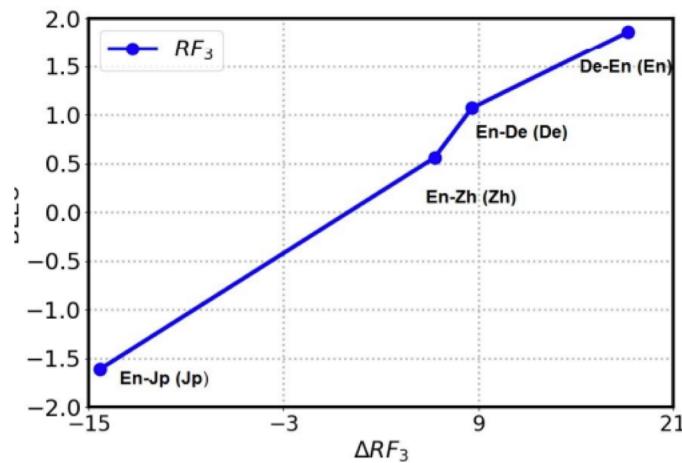
	En-Zh		En-Jp	
	2-gram	3-gram	2-gram	3-gram
$F_n^l$	<b>17707.0</b>	<b>1954.1</b>	18910.0	1350.0
$F_n^r$	16256.4	1250.5	<b>21076.7</b>	<b>1754.0</b>
$RF_n^{l>r}$	<b>51.9%</b>	<b>50.2%</b>	41.2%	38.0%
$RF_n^{l<r}$	46.7%	43.9%	<b>51.7%</b>	<b>52.3%</b>
$\Delta$	5.2%	6.3%	-10.5%	-14.3%

	De-En		En-De	
	2-gram	3-gram	2-gram	3-gram
$P_n^l$	<b>0.137</b>	<b>0.181</b>	<b>0.082</b>	<b>0.155</b>
$P_n^r$	0.092	0.116	0.080	0.148
$RP_n^{l>r}$	<b>59.8%</b>	<b>56.6%</b>	<b>50.6%</b>	<b>51.7%</b>
$RP_n^{l<r}$	38.7%	36.4%	49.2%	47.0%
$\Delta$	21.2%	20.2%	1.4%	4.7%

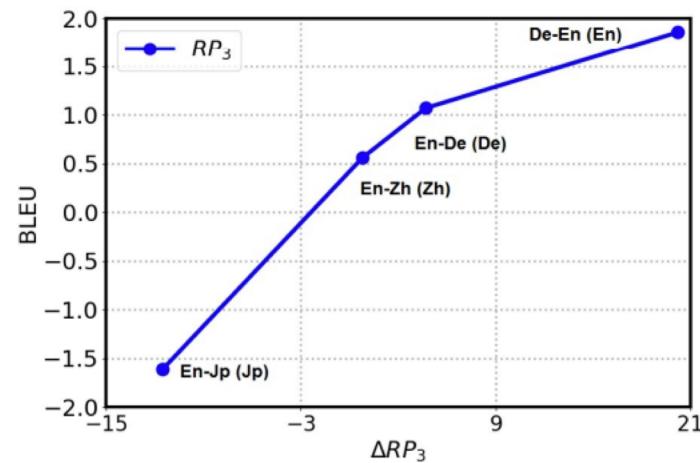
  

	En-Zh		En-Jp	
	2-gram	3-gram	2-gram	3-gram
$P_n^l$	<b>0.064</b>	<b>0.113</b>	0.082	0.171
$P_n^r$	0.055	0.108	<b>0.086</b>	<b>0.191</b>
$RP_n^{l>r}$	<b>52.1%</b>	<b>47.8%</b>	43.9%	39.4%
$RP_n^{l<r}$	46.6%	47.0%	<b>49.2%</b>	<b>50.9%</b>
$\Delta$	5.5%	0.8%	-5.3%	-11.5%

# Visualization the correlation



(a) Accuracy drop v.s 3-gram frequency gap (%).



(b) Accuracy drop v.s 3-gram conditional probability gap (%).

# Dependencies Statistics

- A hypothesize: if the left/right half of sentence contains more dependencies between its intra words, this half should be easier to predict, leading to higher accuracy

	<b>English</b>	<b>Japanese</b>
Left	<b>40242</b>	921735
Right	31509	<b>1570630</b>

# Using RNN-based model

	<b>left-to-right</b>	<b>right-to-left</b>
Full	27.63	25.44
Left	<b>9.17</b>	<b>8.37</b>
Right	7.51	7.25

Table 9: BLEU scores on the left-to-right and right-to-left translation sentences on the De-En test set, with RNN-based model. “Full” means the BLEU score of the whole translation sentence.

# Abstractive Summarization

left-to-right			
	ROUGE-1	ROUGE-2	ROUGE-L
Full	35.55	16.66	33.01
Left	<b>24.44</b>	<b>9.87</b>	<b>23.34</b>
Right	21.31	8.32	20.38

right-to-left			
	ROUGE-1	ROUGE-2	ROUGE-L
Full	35.22	16.55	32.59
Right	21.62	8.41	20.48
Left	<b>23.60</b>	<b>9.54</b>	<b>22.52</b>

Thanks & QA