

Paper Reading: Handling Homographs in Neural Machine Translation (NAACL 2018)

Frederick Liu, Han Lu, Graham Neubig

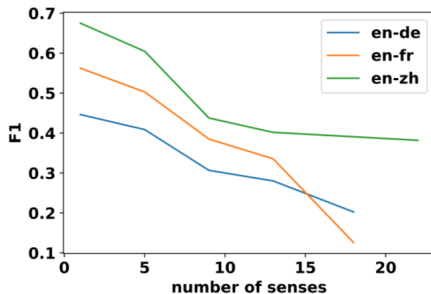
Carnegie Mellon University

Motivation

Standard NMT systems make a significant number of errors on homographs.

Source	Charges against four other men were found not proven .
Reference	对另外四名男子的 指控 最终发现查无实据。(accuse)
Baseline	对四名其他男子的 收费 没有被证明。(fee)
Our Model	对四名男子的 指控 被发现没有被证实。(accuse)

Source	The couch takes up a lot of room .
Reference	Le canapé prend beaucoup de place . (space)
Baseline	Le canapé lit beaucoup de chambre . (bedroom)
Our Model	Le canapé prend beaucoup de place . (space)



Translation performance of words with different numbers of senses.

Neural Word Sense Disambiguation

- ▶ Neural Bag-of-words (NBOW): representing full sentences with a context vector, which is the average of the Lookup embeddings of the input sequence:

$$\mathbf{c}_t = \frac{1}{n} \sum_{k=1}^n \mathbf{M}_c^\top \mathbf{1}(x_k)$$

- ▶ Bi-directional LSTM (BiLSTM): compute the context vector \mathbf{c}_t for t th word:

$$\overrightarrow{\mathbf{c}}_t = \overrightarrow{\text{RNN}}_c(\overrightarrow{\mathbf{c}}_{t-1}, f_c(x_t))$$

$$\overleftarrow{\mathbf{c}}_t = \overleftarrow{\text{RNN}}_c(\overleftarrow{\mathbf{c}}_{t+1}, f_c(x_t))$$

$$\mathbf{c}_t = [\overrightarrow{\mathbf{c}}_t; \overleftarrow{\mathbf{c}}_t]$$

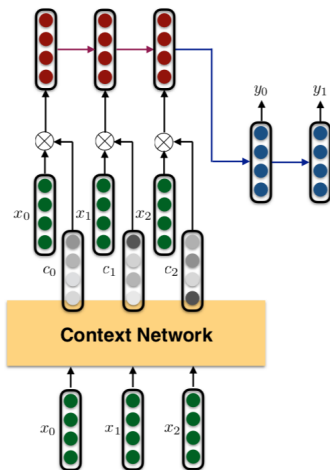
$f_c(x_t)$ is a function that maps a word to continuous embedding space.

- ▶ Held-out LSTM (HoLSTM): compute the context vector c_t for t th word by first replacing t th word with a special symbol. Then feed the replaced sequence to a uni-directional LSTM:

$$\tilde{c}_i = \overrightarrow{\text{RNN}}_c(\tilde{c}_{i-1}, f_c(x_i))$$

$$c_t = \tilde{c}_n$$

Adding Context to NMT



Adding Context to NMT

- Gate:

$$\begin{aligned}f'_e(x_t) &= f_e(x_t) \odot \sigma(\mathbf{c}_t) \\ &= \mathbf{M}_e^\top \mathbf{1}(x_t) \odot \sigma(\mathbf{c}_t)\end{aligned}$$

- Concatenate: concatenating the context vector with the word embeddings.

$$\begin{aligned}f'_e(x_t) &= \mathbf{W}_3[f_e(x_t); \mathbf{c}_t] \\ &= \mathbf{W}_3[\mathbf{M}_e^\top \mathbf{1}(x_t); \mathbf{c}_t]\end{aligned}$$

\mathbf{W}_3 is used to project the concatenated vector back to the original d-dimensional space.

WMT'14,WMT'15 English-German results

Context	Integration	uni/bi	#layers	#params	Ppl	WMT14	WMT15
None	-	→	2	85M	7.12	20.49	22.95
None	-	↔	2	83M	7.20	21.05	23.83
None	-	↔	3	86M	7.50	20.86	23.14
NBOW	Concat	→	2	85M	7.23	20.44	22.83
NBOW	Concat	↔	2	83M	7.28	20.76	23.61
HoLSTM	Concat	→	2	87M	7.19	20.67	23.05
HoLSTM	Concat	↔	2	86M	7.04	21.15	23.53
BiLSTM	Concat	→	2	87M	6.88	21.80	24.52
BiLSTM	Concat	↔	2	85M	6.87	21.33	24.37
NBOW	Gating	→	2	85M	7.14	20.20	22.94
NBOW	Gating	↔	2	83M	6.92	21.16	23.52
BiLSTM	Gating	→	2	87M	7.07	20.94	23.58
BiLSTM	Gating	↔	2	85M	7.11	21.33	24.05

Translation results for homographs and all words in NMT vocabulary

language	System	Homograph			All Words		
		F1	Precision	Recall	F1	Precision	Recall
en → de	baseline	0.401	0.422	0.382	0.547	0.569	0.526
	best	0.426 (+0.025)	0.449 (+0.027)	0.405 (+0.023)	0.553 (+0.006)	0.576 (+0.007)	0.532 (+0.006)
en → fr	baseline	0.467	0.484	0.451	0.605	0.623	0.587
	best	0.480 (+0.013)	0.496 (+0.012)	0.465 (+0.014)	0.613 (+0.008)	0.630 (+0.007)	0.596 (+0.009)
en → zh	baseline	0.578	0.587	0.570	0.573	0.605	0.544
	best	0.590 (+0.012)	0.599 (+0.012)	0.581 (+0.011)	0.581 (+0.008)	0.612 (+0.007)	0.552 (+0.008)

Case Study

English-Chinese Translations	
src	Ugandan president meets Chinese FM , anticipates closer cooperation
ref	乌干达总统 会见 中国外长, 期待增进合作(come together intentionally)
best	乌干达总统 会见 中国调频, 预期更密切合作(come together intentionally)
base	乌干达总统 符合 中国调频, 预期更加合作(satisfy)
src	Investigators are trying to establish whether Kermiche and Petitjean had accomplices in France and whether they had links with Islamic State , which has claimed responsibility for the attack .
ref	调查人员正试图 确定 克尔米奇和帕迪让在法国是否有同谋, 以及是否与伊斯兰国武装分子有联系, 伊斯兰国武装分子声称对此次袭击负责。(get proof of something)
best	调查人员正试图 确定 Kermiche 和Petitjean 在法国是否有同谋, 他们是否与伊斯兰国有联系, 声称对这次袭击负责。(get proof of something)
base	调查人员正在努力 建立 法国的同谋和他们是否与伊斯兰国有联系, 该国声称对这次袭击负有责任。(to start)
src	The decrease of transaction settlement fund balance in the securities market in July was smaller than that in June , while the net bank @-@ securities transfers stood at negative RMB 66.6 billion .
ref	7 月证券市场交易结算资金 余额 减少额较6 月大幅降低, 银证转账变动净额为- 666 亿元。(money left)
best	7 月份证券市场交易结算资金 余额 的减少小于6 月份, 而银行证券转让净额为negative 亿元。(money left)
base	七月证券市场交易结算基金 平衡 的减少比六月份小, 而净银行证券转让则为负元。(equal weight or force)
src	Initial reports suggest that the gunman may have shot a woman , believed to be his ex @-@ partner .
ref	据初步报告显示, 开枪者可能击中一名妇女, 据信 是他的前搭档。(been accepted as truth)
best	初步的报道表明, 枪手可能已经射杀了一个女人, 被认为 是他的前伙伴。(been known as)
base	最初的报道显示, 枪手可能已经射杀了一名妇女, 相信 他是他的前伙伴。(accept as truth)
src	When the game came to the last 3 ' 49 ' ' , Nigeria closed to 79 @-@ 81 after Aminu added a layup .
ref	比赛还有3 分49 秒时, 阿米努上篮得手后, 尼日利亚将比分 追成 了79-81 . (narrow)
best	当这场比赛到了最后三个“ 49 ”时, 尼日利亚在Aminu 增加了一个layup 之后 MISSING TRANSLATION 。
base	当游戏到达最后3 “ 49 ” 时, 尼日利亚已经 关闭 了Aminu . (end)

Conclusion

- ▶ Context-aware word embedding successfully improving the NMT model.
- ▶ Inspiration: Context information can help improve NMT either bleu score or other specific task.