Attention Focusing for Neural Machine Translation by Bridging Source and Target Embeddings

Motivation
- In NMT, source and target word embeddings locate at the two ends of a long information processing procedure.
- The associations and interactions between source and target words are loose in NMT.

Examples of NMT output with incorrect alignments of source and target words.
- Statistics on our development data show that as high as 50% of target side eos do not properly align to source side eos.
- Source words are translated into target items that are not their possible translations.

Solution
- Focusing on modeling connections between source and target words.
- Three methods: Source-side bridging model, Target-side bridging model, Direct bridging model.

Experiments
- Chinese-English bilingual corpora (1.25M sentence pairs).

BLEU scores on the test data

Analysis of Over and Under Translation

Analysis of Long Sentence Translation
- During training, the maximum source sentence length limit is set to 50, thus making the learned models not ready to cope well with sentences over this maximum length limit.
- For long input sentences, NMT systems tend to stop early in the generation of the translation.

Analysis of Word Alignment
- The closest target words obtained via the transformation matrix of our direct bridging method are very consistent with those obtained from the SMT lexical table.
- Our improved model has a good capability of capturing the translation equivalence between source and target word embeddings.

Analysis of Learned Word Embeddings

Data

System | NIST03 | NIST04 | NIST05 | NIST06 | NIST07 | NIST08 | Avg.
--- | --- | --- | --- | --- | --- | --- | ---
BLEU | 33.02 | 37.28 | 36.21 | 38.43 | 28.30 | 30.91 | 34.00
RNNSearch* | 32.70 | 36.71 | 37.24 | 40.25 | 27.40 | 29.05 | 31.70
Source bridging | 36.69 | 39.04 | 37.63 | 41.41 | 23.98 | 36.27 | 34.71
Target bridging | 36.47 | 38.77 | 38.42 | 40.85 | 27.61 | 36.42 | 34.71
Direct bridging | 38.29 | 40.65 | 39.12 | 41.12 | 31.54 | 38.04 | 39.04

ter | 59.70 | 57.19 | 59.25 | 57.88 | 64.22 | 59.78 | 59.70
RNNSearch* | 59.13 | 58.28 | 57.33 | 55.02 | 62.15 | 58.01 | 59.13
Source bridging | 58.11 | 56.27 | 57.36 | 55.33 | 62.12 | 58.12 | 58.11
Target bridging | 57.28 | 55.08 | 57.29 | 55.62 | 62.49 | 58.02 | 57.28

Better word alignment

A lower score indicates better alignment.

Confusion matrix for translation by part-of-speech (POS).