Paper Report

NGO HO Anh Khoa

October 13, 2017

1 IBM models

2 Unsupervised Neural HMMs [3]

3 What does Attention in NMT pay Attention to ? [1]

3.1 Main idea

This research compares between Attention Models (Non-recurrent attention model/ Global attention and Recurrent attention/ Input-feeding model) and known Word Alignment. The result is that their differences depends on the word being generated.

3.2 How to compare ?

Higher consistency between Attention and Alignment leading to better translation.

- **Spearman's rank correlation between attention quality** (Attention loss compared known human alignment) **and translation quality** (Word prediction loss). Higher correlation means a closer relationship between translation quality and consistency of attention versus alignment.

  - Attention loss
  - Word prediction loss

- Attention concentration: Entropy of attention distribution (Soft-hard attention problem)

3.3 Analysis

The analysis is based on POS tags experiments.

- Impact of Attention between Non-recurrent (NR) and Recurrent attention (IF): IF has lower AER (Hard attention) and Attention loss (Soft attention).

- Translation quality: Consistency between attention and word alignment depends on POS tags. For example,

  - There is a higher consistency in the case of nouns. However, attention captures other information in the case of verbs.
  - Translation quality of Verbs is better than Nouns. It means attention does not follow alignment for translating Verbs.

- Attention concentration: Review the case of Verbs and Nouns.

  - Nouns have a lower attention entropy (Higher concentration), lower attention loss (Closer to Alignment), which is that attention entropy can be used as a measure of closeness of attention to alignment in the case of nouns
  - Verbs have a lower correlation between attention entropy and word prediction loss, which means that attention concentration is not necessary for translating verbs.

- Attention distribution: It shows how a POS tag of target sentence depends on other POS tags of source sentence.

4 Confidence through Attention [2]

4.1 Main idea

References

[1] H. Ghader and C. Monz. What does Attention in Neural Machine Translation Pay Attention to? *ArXiv e-prints*, October 2017.

[2] M. Rikters and M. Fishel. Confidence through Attention. *ArXiv e-prints*, October 2017.

[3] M. Ke Tran, Yonatan Bisk, Ashish Vaswani, Daniel Marcu, and Kevin Knight. Proceedings of the workshop on structured prediction for nlp. pages 63–71. Association for Computational Linguistics, 2016.