Visualizing and Analyzing Neural Machine Translation

Marine Carpuat, Dipanwita Dasgupta, Wen Li, Can Liu, Maria Nadejde, Luis Prieb, Rico Sennrich

Machine Translation Marathon in the Americas 2016

May 21 2016
Return and visualize NMT attention model
Implement copy mechanism
Return and visualize NMT output probabilities
Analyze ensembles and averaged model
Apply NMT for lexical selection task

Code contributed to https://github.com/rsennrich/nematus
Alignment Visualization

Visualizing/Analyzing NMT
Alignment Visualization
Alignment Visualization

Visualizing/Analyzing NMT
Alignment Visualization
a year later, the Fed... Ex officials reversed those cuts.

- alignments off by one???
- same code produces correct alignment for other sentences/languages
- possibly result of information passing through recurrent states
Implementation of [Jean et al., 2015]
Replace unknown words in target sentences
Input: target sentence, source sentence, alignment matrix (probabilities)
Output: updated target sentence (<UNK> replaced with aligned source word)
Steps:
  - find best alignment for each target word
  - If target word is <UNK>, replace it with aligned source word
• confidence seems useful → errors in low-confidence places
• ’eded’ has high probability given ’se...’ → label bias problem?
Analysis: Ensembles

Noteworthy data points
- single system: 26.36
- ensemble: 27.45
- averaged all layers: 27.31
- averaged output layer: 26.63

Experiment
- comparison of three systems:
  - single model
  - ensemble of 4
  - element-wise average of 4
- 42 additional experiments:
  - average one layer at a time hypotheses
    - element-wise average is different from ensembling, but has similar effect
    - benefit of averaging is mostly due to output layer
Thank you!
On Using Very Large Target Vocabulary for Neural Machine Translation.
In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 1–10, Beijing, China. Association for Computational Linguistics.