Does String-Based Neural MT Learn Source Syntax?
Xing Shi, Inkit Padhi and Kevin Knight

Motivation

Whether does encoder learn syntactic information of source sentence?
What kind of syntactic information does the encoder learn?
How much syntactic information does the encoder learn?

Where is Neural Machine Translation?

Question

1. Whether does encoder learn syntactic information of source sentence?
2. What kind of syntactic information does the encoder learn?
3. How much syntactic information does the encoder learn?

Conclusion

1. NMT encoder learns quite a lot syntactic information, but still lost subtle details.
2. Not too much difference between E2F and E2G
3. NMT encoder is better than Autoencoder if you want to convert sentence into vector.

Method 1: Syntactic labels = logistic-regression(hidden states)

Method 2: Extract whole parse tree from NMT encoder

Analysis done by Berkeley Parser Analyzer (Kummerfeld et al., 2012)