NMT WITH ATTENTION

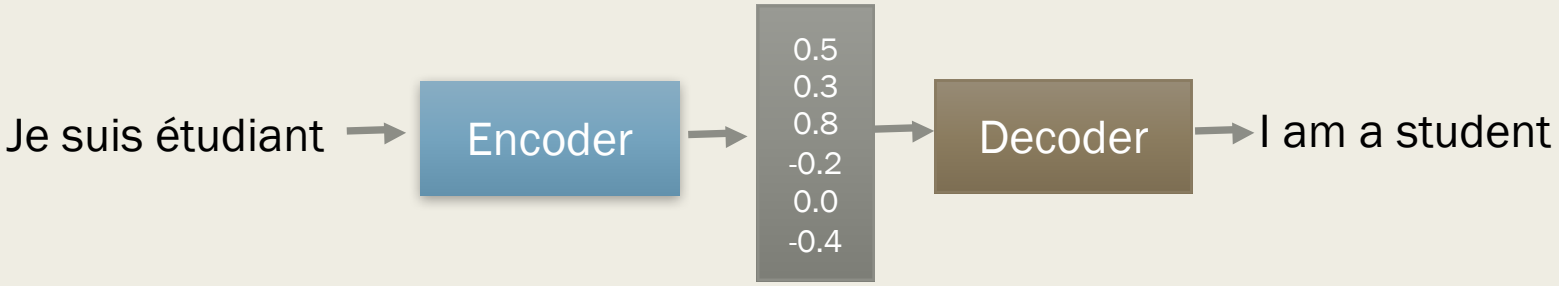
Seq2Seq neural machine translator with attention mechanisms

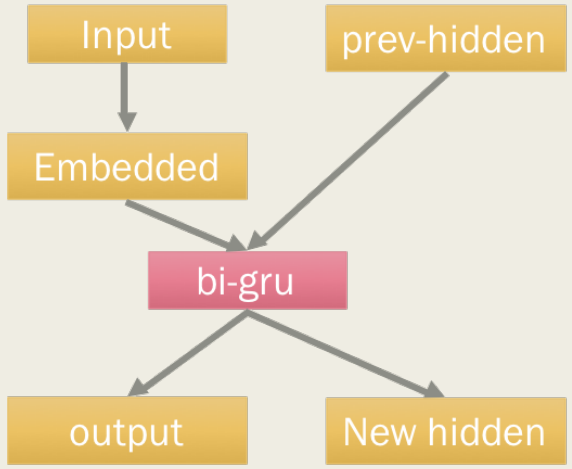
Wenyue LIU

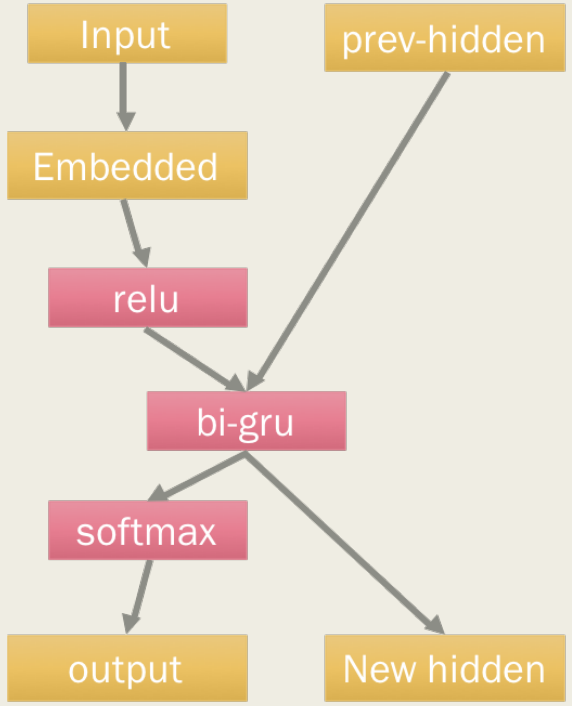
Seq2Seq

- Encode French sentence to a vector using gated recurrent unit.
- Decode the output from encoder to English.

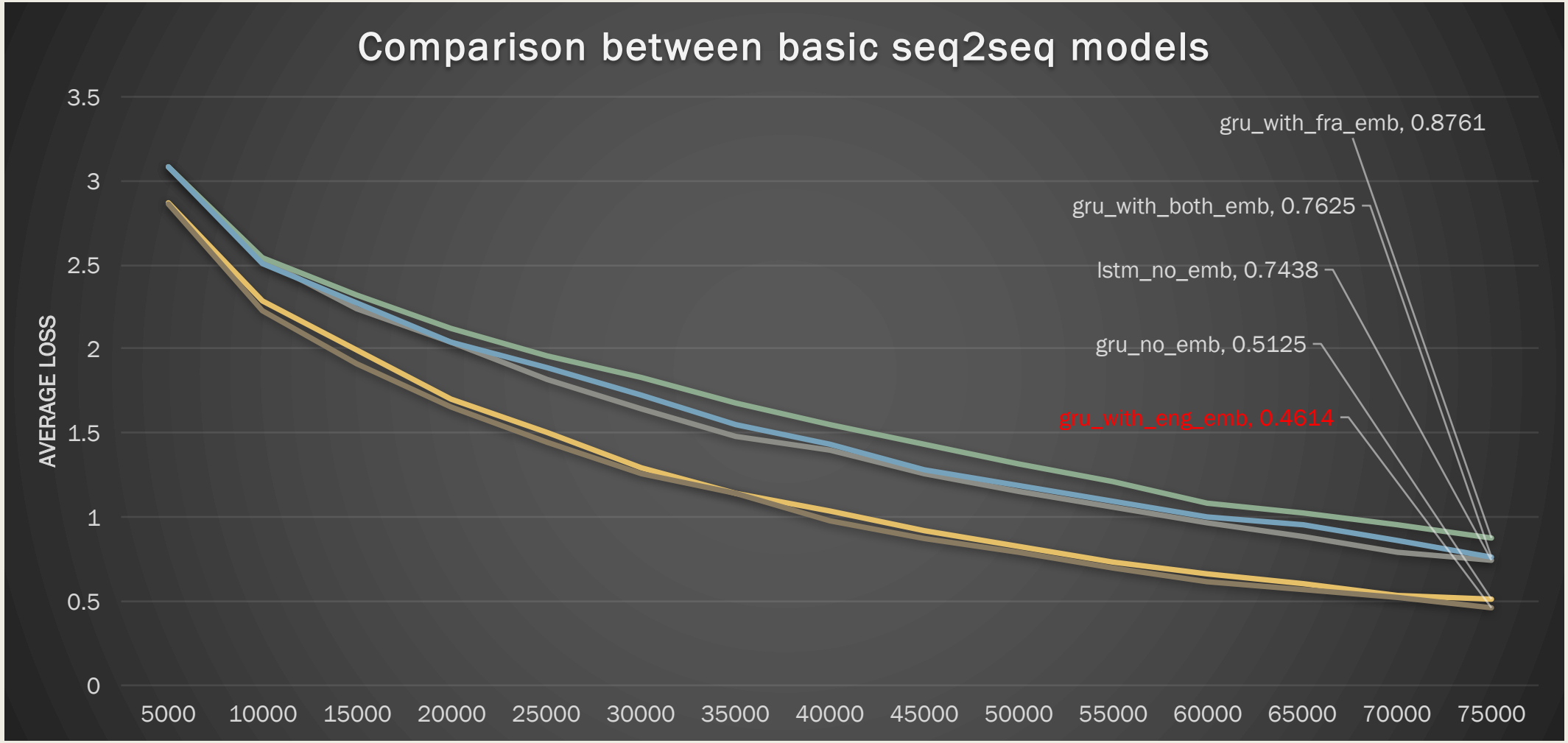
- GRU performs better than LSTM
- Glove embedding on decoder applied, however, word2vec embedding on encoder hurt the performance.





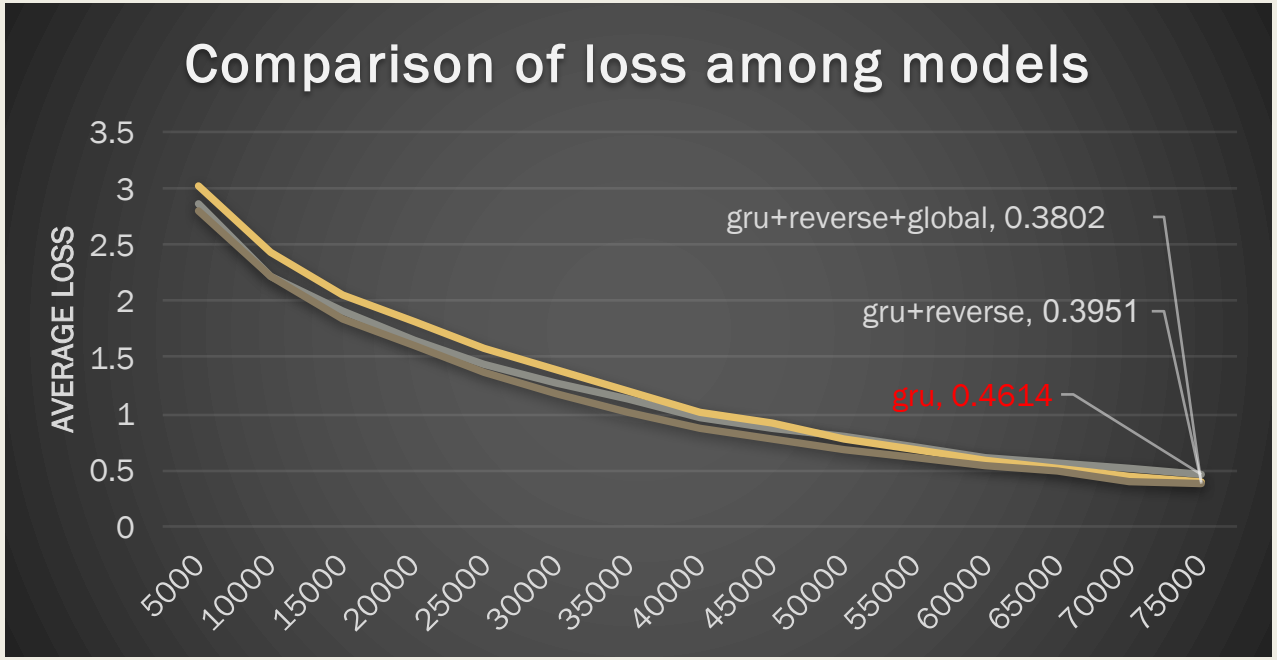


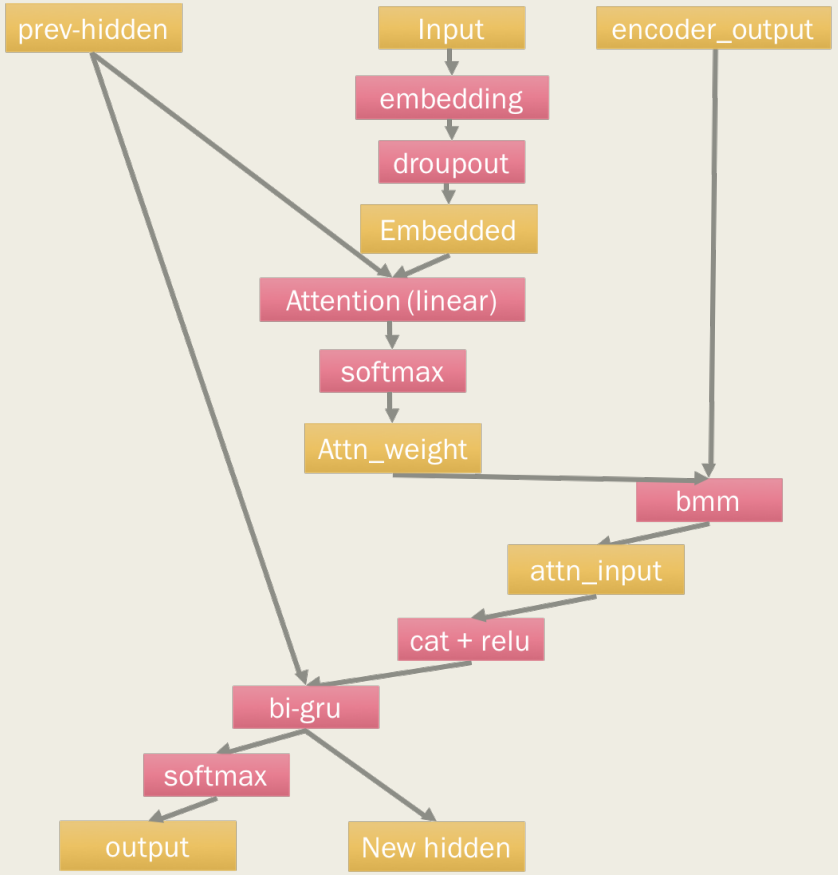
Some performance results



bi-gru and attention in decoder

- Bi directional gru has much better performance.
- Added global attention into decoder[1]
- Add local attention into decoder (not yet implemented)
- Allow decoder focus on parts of encoder output.



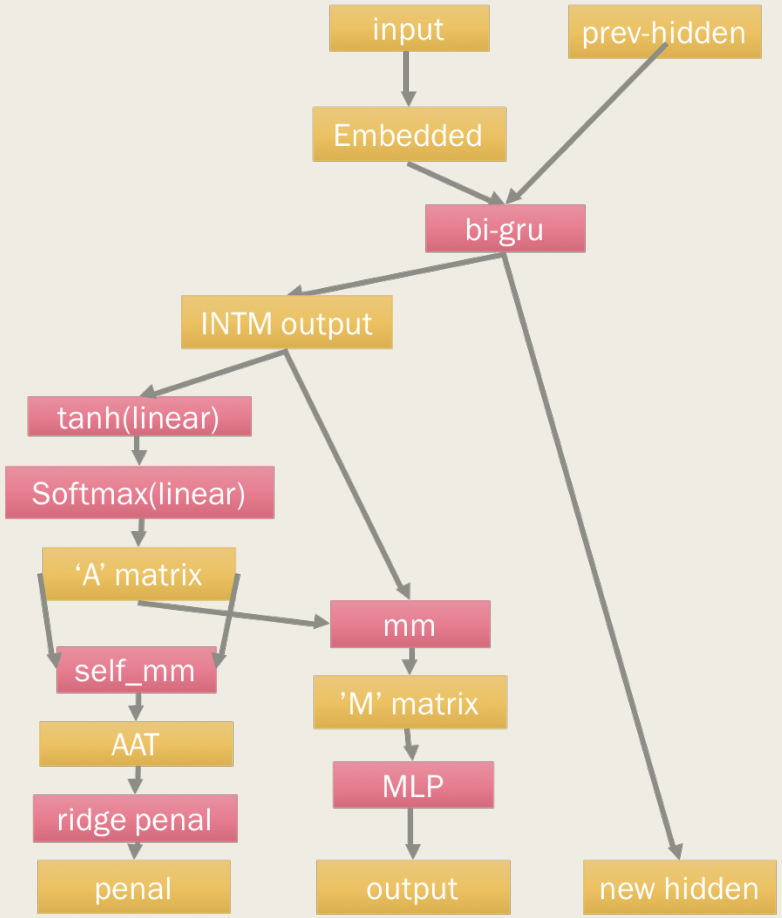


1. Effective Approaches to Attention-based Neural Machine Translation arXiv:1508.04025v5

Self attention mechanism on encoder

- Apply attention mechanism on encoder layer?
- We do not have the '**encoder output**' as additional info.
- Self attention[2]
- Without the encoder input, we used intermediate output to build attention unit.
- Hyper-parameter r: number of hops (attention units) matters.
- Penalization item controls volatility introduced by attention mechanism.

- By using the optimal r, performance of self-attentive bi-gru model works slightly better than basic bi-gru model.



2. A Structured Self-attentive Sentence Embedding arXiv:1703.03130v1

Next steps:

- Try to implement the local attention mechanism
- Understand why French embedding does not work.

- Theoretically possible to use attention mechanism in both encoding and decoding? Implementation

Thanks
Q&A