Enhanced Neural Machine Translation by Learning from Draft

Aodong Li

NLP Group, CSLT, Tsinghua University

Route Map

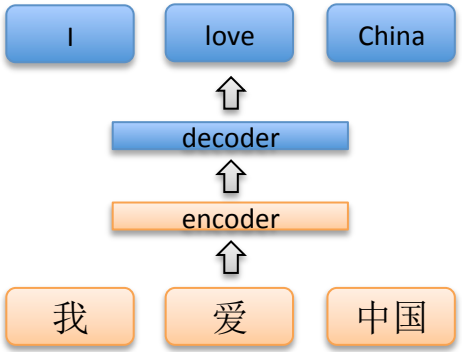
- Introduction
- Learning from draft
- Experiments
- Conclusions
- Future work

Introduction (1)

- Statistical Machine Translation (SMT) is a translation paradigm that focuses on latent feature design.
  - word alignment
  - source sentence length
  - target sentence length
  - …
- Benefit: Its latent structure is explainable.
- Shortcomings:
  - expert designed feature
  - expert designed translation process
  - long-distance dependency
  - …
- In recent years, Neural Machine Translation (NMT) has achieved the state-of-the-art results on many language pairs, e.g., English-to-French, English-to-German, Chinese-to-English, etc.

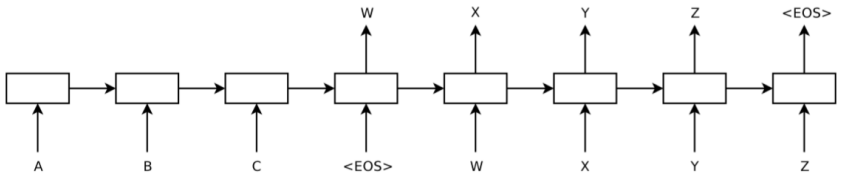
Introduction (2)

- Neural Machine Translation usually adopts an encoder-decoder structure to accommodate paired languages.
- The decoder acts as a language model which incorporates the left context but **ignores the right context**. $p(y_t|y_{i<t}, X)$
- Is this right context useful?
- If it's useful, how can we use the right context?



Introduction (3)

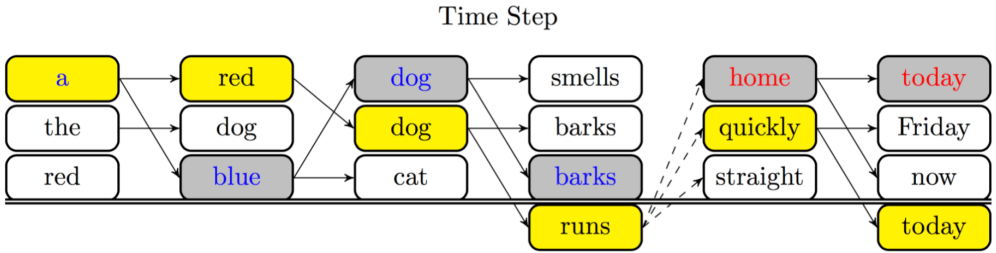
- Is the right context useful?
  - Yes!
  - Sutskever et al. found that the sequence-to-sequence model achieved a promising improvement when reversing the source sentence “a, b, c” to “c, b, a” [1].
  - A significant improvement could be obtained when using a bi-directional RNN rather than a uni-directional RNN [2].



[1] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *Proc. of NIPS*, NIPS’14, pages 3104– 3112, 2014.
[2] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

Introduction (4)

- How can we use this right context?
  - Beam search is a way to utilize the right context, but to a limited extent [1][2].
  - We propose a two-stage translation approach with the idea of drafting-and-refinement to tackle this problem. The draft contains the right context.
  - Novak et al. proposed a similar iterative translation approach in which they correct the words again and again [3].



[1] Wiseman S, Rush A M. Sequence-to-sequence learning as beam-search optimization[J]. arXiv preprint arXiv:1606.02960, 2016.
[2] Liang Huang. *Forest-based algorithms in natural language processing*. PhD thesis, University of Pennsylvania, 2008.
[3] Roman Novak, Michael Auli, and David Grangier. Iterative refinement for machine translation. *arXiv preprint arXiv:1610.06602*, 2016.

Learning from draft

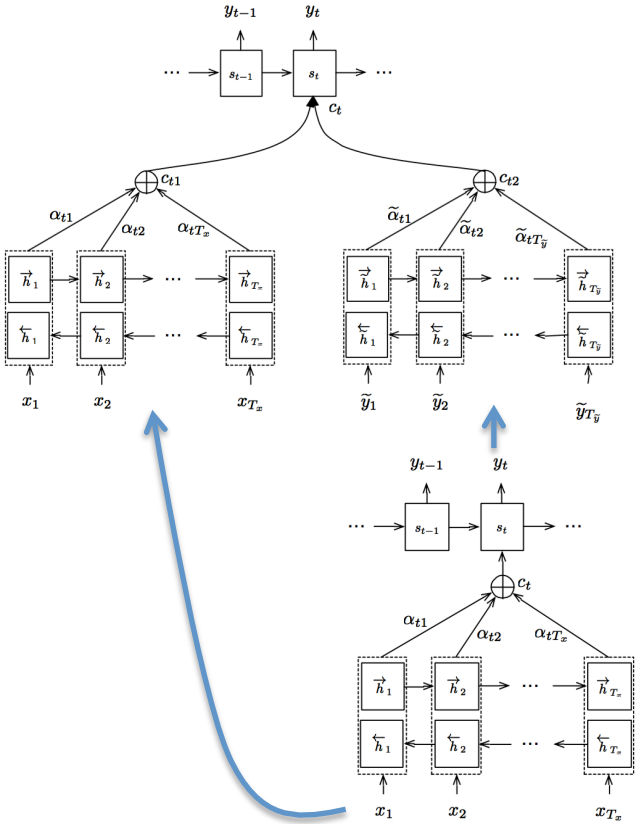
- Two-stage translation approach: drafting-and-refinement.
- Drafting: the source sentence *X* is translated into a draft . $\widetilde{Y} = (\widetilde{y}_1, \widetilde{y}_2, ..., \widetilde{y}_{T_{\widetilde{y}}})$

$$p(y_t|y_{i<t}, C) = g(y_{t-1}, s_t, c_t)$$

- The right context can be obtained from the draft.
- Refinement: both the original source sentence and the draft are translated together.

$$p(y_t|y_{i<t}, C_1, C_2) = g(y_{t-1}, s_t, c_t)$$

$$c_t = [c_{t1}^\top; c_{t2}^\top]^\top$$



Experiments

- Experiments were conducted on two Chinese-English tasks.
- Dataset:
  - large NIST corpora with 1M parallel training data.
  - small IWSLT corpora with 44K parallel training data.
- Comparison systems:
  - Moses: a widely-used SMT system.
  - Attention-based NMT: a popular NMT system.
- Evaluation metric:
  - we used the case-insensitive 4-gram NIST BLEU score.

TABLE I
BLEU SCORES ON CHINESE-ENGLISH TRANSLATION

SYSTEM	NIST	IWSLT
Moses	30.6	**52.5**
Attention-based NMT	30.83	43.83
Double-attention NMT	**31.71**	46.32

Conclusions

- The target sentence's right context is informative and provides an important regulation for current generated word.
- Our two-stage translation approach can utilize the right context to enhance the neural translation model.

Future work

- Is the right context more informative than the left context?
- Can the model be constructed in a uniform framework so that we don’t have to conduct two separate training processes?
- More experiments on different language pairs need to be done to confirm our approach’s effect.

Thanks! Q&A