When and Why is Document-level Context Useful in Neural Machine Translation?

Yunsu Kim, Duc Thanh Tran, Hermann Ney
RWTH Aachen University, Aachen, Germany
`kim@cs.rwth-aachen.de`

Document-level NMT Hype

Dozens of methods for integrating document-level context in NMT:

[Jean & Lauly[+] 17] [Tiedemann & Scherrer 17] [Tu & Liu[+] 18] [Cao & Xiong 18]

[Zhang & Luan[+] 18] [Voita & Serdyukov[+] 18] [Maruf & Haffari 18] [Kuang & Xiong[+] 18b]

[Kuang & Xiong 18a] [Bawden & Sennrich[+] 18] [Miculicich & Ram[+] 18] [Stojanovski & Fraser 18]

[Voita & Sennrich[+] 19] [Maruf & Martins[+] 19] [Junczys-Dowmunt 19]

$\vdots$

- 14 out of 15 papers: trained **only with** (restricted) **document-level** data
  - ▷ TED, News commentary, Europarl, OpenSubtitles, ...
  - ▷ For both sentence-level and document-level training
- In average: $\sim +1.0$ Bleu $[\%]$ over a sentence-level baseline

Question: Where do all improvements come from? Coreference? Coherence?

Causes of Improvements

How often is document-level context utilized for coreference/coherence?

	#sentences	
	en-it	**en-de**
Total	1,147	2,998
Total TER improved	379	1,246
↪ Coreference	21	2
↪ Topic-aware lexical choice	66	33
↪ Others	**292**	**1,211**

- **92.5%** of improvements ("Others"): **irrelevant** to document-level context
- Our hypothesis: context causes the regularization effect

Stronger Sentence-level NMT

Document-level training on top of **well-regularized sentence-level models**

Training Data	Dropout	System	Bleu [%]	
			en-it	**en-de**
Small	0.1	Sentence-level	31.4	28.9
		Document-level	**32.5**	**30.3**
	0.3	Sentence-level	**33.7**	**32.3**
		Document-level	33.5	32.0
Large	0.1	Sentence-level	-	**40.2**
		Document-level	-	39.9

- No improvements in Bleu over strong sentence-level systems
- Targeted test sets: improvements might not carry over to real scenarios

When and Why is Document-level Context Useful in Neural Machine Translation?
— DiscoMT 2019, November 3, 2019

RWTH AACHEN UNIVERSITY

Remarks

1. Do **not** sell the improvements in Bleu by document-level context
2. Check the **real** document-level improvements **manually**
3. Build the sentence-level system **as strong as possible** first

Come to our poster to also check:

- How much modeling power is needed for context encoding?
- Is the entire context sentence needed?
- Is a long-range context useful?

When and Why is Document-level Context Useful in Neural Machine Translation?

Yunsu Kim, Duc Thanh Tran, Hermann Ney

References

[Bawden & Sennrich+ 18] R. Bawden, R. Sennrich, A. Birch, B. Haddow.

Evaluating discourse phenomena in neural machine translation.

In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1304–1313, 2018.

[Cao & Xiong 18] Q. Cao, D. Xiong.

Encoding gated translation memory into neural machine translation.

In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3042–3047, 2018.

[Jean & Lauly+ 17] S. Jean, S. Lauly, O. Firat, K. Cho.

Does neural machine translation benefit from larger context?

arXiv preprint arXiv:1704.05135, Vol., 2017.

[Junczys-Dowmunt 19] M. Junczys-Dowmunt.

Microsoft translator at wmt 2019: Towards large-scale document-level neural machine translation.

In *Proceedings of the Fourth Conference on Machine Translation (WMT), Volume 2: Shared Task Papers*, pp. 424–432, Florence, Italy, August 2019.

[Kuang & Xiong 18a] S. Kuang, D. Xiong.

Fusing recency into neural machine translation with an inter-sentence gate model.

In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 607–617, 2018.

[Kuang & Xiong+ 18b] S. Kuang, D. Xiong, W. Luo, G. Zhou.

Modeling coherence for neural machine translation with dynamic and topic caches.

In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 596–606, 2018.

[Maruf & Haffari 18] S. Maruf, G. Haffari.

Document context neural machine translation with memory networks.

In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1275–1284, 2018.

When and Why is Document-level Context Useful in Neural Machine Translation?
— DiscoMT 2019, November 3, 2019

References

[Maruf & Martins+ 19] S. Maruf, A. F. Martins, G. Haffari.

Selective attention for context-aware neural machine translation.

In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 3092–3102, 2019.

[Miculicich & Ram+ 18] L. Miculicich, D. Ram, N. Pappas, J. Henderson.

Document-level neural machine translation with hierarchical attention networks.

In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2947–2954, 2018.

[Stojanovski & Fraser 18] D. Stojanovski, A. Fraser.

Coreference and coherence in neural machine translation: A study using oracle experiments.

In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pp. 49–60, 2018.

[Tiedemann & Scherrer 17] J. Tiedemann, Y. Scherrer.

Neural machine translation with extended context.

In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pp. 82–92, 2017.

[Tu & Liu+ 18] Z. Tu, Y. Liu, S. Shi, T. Zhang.

Learning to remember translation history with a continuous cache.

Transactions of the Association for Computational Linguistics, Vol. 6, pp. 407–420, 2018.

[Voita & Sennrich+ 19] E. Voita, R. Sennrich, I. Titov.

When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion.

In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, pp. 1198–1212, Florence, Italy, July 2019.

[Voita & Serdyukov+ 18] E. Voita, P. Serdyukov, R. Sennrich, I. Titov.

Context-aware neural machine translation learns anaphora resolution.

In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1264–1274, 2018.

References

[Zhang & Luan⁺ 18] J. Zhang, H. Luan, M. Sun, F. Zhai, J. Xu, M. Zhang, Y. Liu.
Improving the transformer translation model with document-level context.
In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 533–542, 2018.

When and Why is Document-level Context Useful in Neural Machine Translation?
— DiscoMT 2019, November 3, 2019