Recent Advances In Document-level Neural Machine Translation

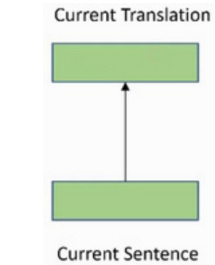
Lorenzo Lupo

Supervisors: Laurent Besacier, Marco Dinarelli

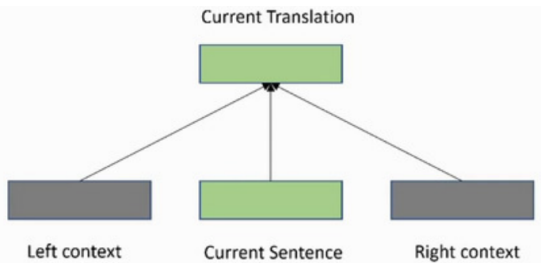
July 10, 2020

What is Document-level Machine Translation

Sentence-level MT

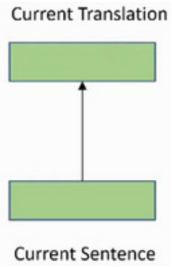


Document-level MT

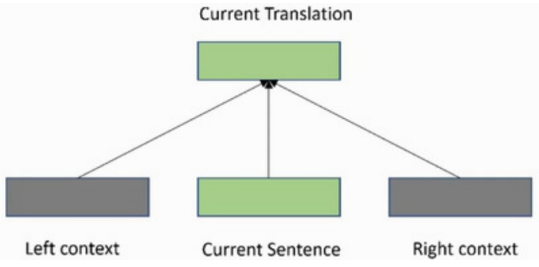


Document-level MT ↔ Context-aware MT

Context-agnostic MT



Context-aware MT



Why Document-level NMT ?

Why Document-level NMT ?

- Some recent results suggest that neural machine translation (NMT) "approaches the accuracy achieved by average bilingual human translators [on some test sets] [Wu et al., 2016]

Why Document-level NMT ?

- Some recent results suggest that neural machine translation (NMT) "approaches the accuracy achieved by average bilingual human translators [on some test sets] [Wu et al., 2016]
- "In a pairwise ranking experiment, human raters assessing **adequacy** and **fluency** show a stronger preference for human over machine translation when evaluating documents as compared to isolated sentences." [Lubli et al., 2018]

Sentence-level NMT is inconsistent

B: How are you today?

Sentence-level NMT is inconsistent

B: How are you today?

SENTENCE-LEVEL TRANSLATION

B: Comment vas-tu aujourd'hui ?

Sentence-level NMT is inconsistent

A: Good Morning, Mr. President.

B: How are you today?

SENTENCE-LEVEL TRANSLATION

B: Comment vas-tu aujourd'hui ?

Sentence-level NMT is inconsistent

A: Good Morning, Mr. President.

B: How are you today?

SENTENCE-LEVEL TRANSLATION

B: Comment vas-tu aujourd'hui ?

CONTEXT-AWARE TRANSLATION

B: Comment allez-vous aujourd'hui ?

How frequent are inconsistencies ?

[Voita et al., 2019] undertake a human study on context agnostic translation :

- 2000 pairs of consecutive English sentences (S1 + S2) from OpenSubtitles2018
- translate to Russian with Transformer model [Vaswani et al., 2017]

How frequent are inconsistencies ?

[Voita et al., 2019] undertake a human study on context agnostic translation :

- 2000 pairs of consecutive English sentences (S1 + S2) from OpenSubtitles2018
- translate to Russian with Transformer model [Vaswani et al., 2017]

all	one/both bad	both good	
		bad pair	good pair
2000	211	140	1649
100%	11%	7%	82%

Which kind of inconsistencies?

type of phenomena	frequency
deixis	37%
ellipsis	29%
lexical cohesion	14%
ambiguity	9%
anaphora	6%
other	5%

Figure: Types of phenomena causing inconsistencies between English-Russian context-agnostic translations of consecutive sentences when placed in the context of each other.

Objectives

Objectives

- **Design translation models and learning techniques** that solve inconsistencies by taking context into account;

Objectives

- **Design translation models and learning techniques** that solve inconsistencies by taking context into account;
- **Evaluate such models** in a proper way;

Plan

Thank you for your attention!

References I

Lubli, S., Sennrich, R., and Volk, M. (2018).
Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation.
arXiv:1808.07048 [cs].
00038 arXiv: 1808.07048.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017).
Attention Is All You Need.
arXiv:1706.03762 [cs].
05728 arXiv: 1706.03762.

References II


Voita, E., Sennrich, R., and Titov, I. (2019).
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*, pages 1198–1212, Florence, Italy. Association for Computational Linguistics.
00007.


Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, ., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. (2016).
Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.

References III

arXiv:1609.08144 [cs].
00000 arXiv: 1609.08144.