ACL 2018 Trip Report

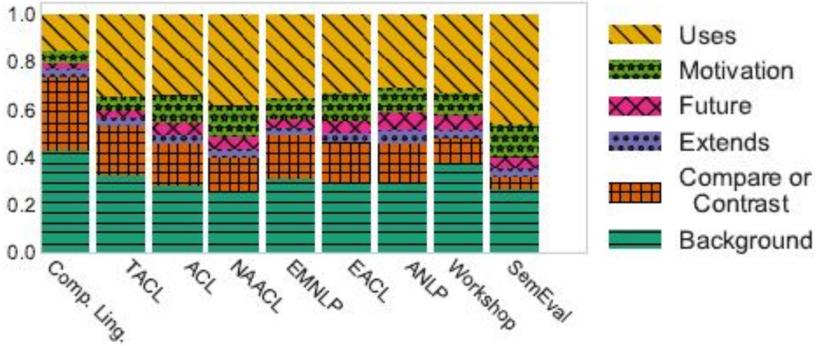
Qingyun Wang

Content

1. Measuring the Evolution of a Scientific Field through Citation Frames
2. Unsupervised Neural Machine Translation with Weight Sharing
3. The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation

Measuring the Evolution of a Scientific Field through Citation Frames

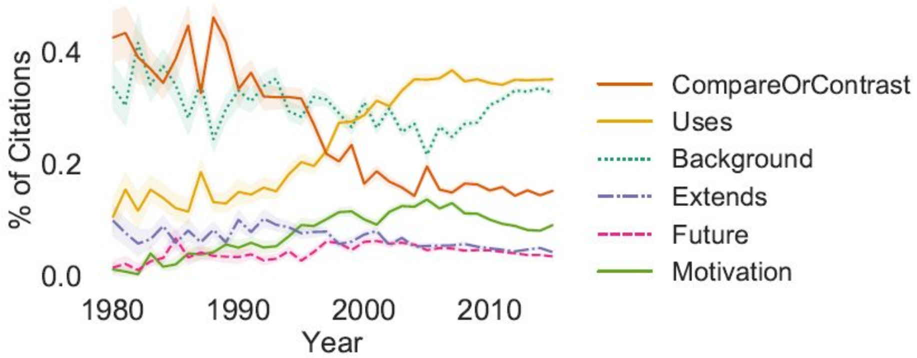
- Goal: Analyze how to frame citation, the functionality of this framing
- Classification Scheme: Uses, Motivation, Future,, Extends, Compare or Contrast, Background
- Features: (i) Structural, (ii) Lexical, Morphological, and Grammatical, (iii) Field, (iv) Usage



[1] Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the Evolution of a Scientific Field through Citation Frames. Transactions of the Association for Computational Linguistics, 6, 391-406.

Measuring the Evolution of a Scientific Field through Citation Frames

- Model: Random Forest classifier
- Contribution: (i) New corpus of citation function, (ii) Venue influence the citation significantly (iii) Vitation framing have a significant impact on future citations (iv) NLP community has evolved by how authors frame their work



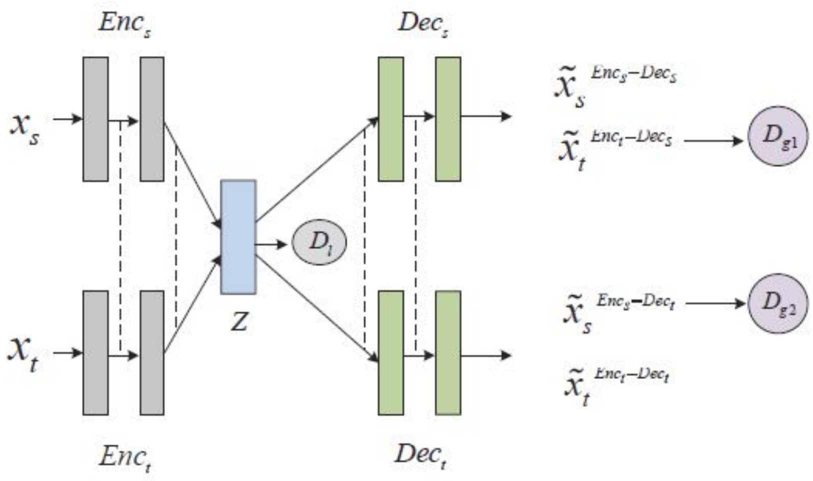
[1] Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the Evolution of a Scientific Field through Citation Frames. Transactions of the Association for Computational Linguistics, 6, 391-406.

Unsupervised Neural Machine Translation with Weight Sharing

- Problem: Losing unique and internal characteristics of each language; Shared encoder may be a factor limiting the potential translation performance
- Contribution: Independent encoder for each language, Two different GAN, Directional self attention
- Local GAN: Constrain the source and target latent representations to have the same distribution
- Global GAN: Finetune the composition of encoder and decoder on different language

Overview

- Dashlines represents weight sharing constraint on last few layer of encoder and first few layer of decoder
- D_l is utilized to assess whether the hidden representation of the encoder is from the source or target language.
- D_{g1} and D_{g2} are used to evaluate whether the translated sentences are realistic for each language respectively.
- Z represents the shared-latent space



Networks	Roles
$\{Enc_s, Dec_s\}$	AE for source language
$\{Enc_t, Dec_t\}$	AE for target language
$\{Enc_s, Dec_t\}$	translation $source \rightarrow target$
$\{Enc_t, Dec_s\}$	translation $target \rightarrow source$
$\{Enc_s, D_l\}$	1st local GAN (GAN_{l1})
$\{Enc_t, D_l\}$	2nd local GAN (GAN_{l2})
$\{Enc_t, Dec_s, D_{g1}\}$	1st global GAN (GAN_{g1})
$\{Enc_s, Dec_t, D_{g2}\}$	2nd global GAN (GAN_{g2})

[2] Yang, Z., Chen, W., Wang, F., & Xu, B. (2018). Unsupervised Neural Machine Translation with Weight Sharing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)

Details and analysis

Directional self-attention: Forward and backward positional masks

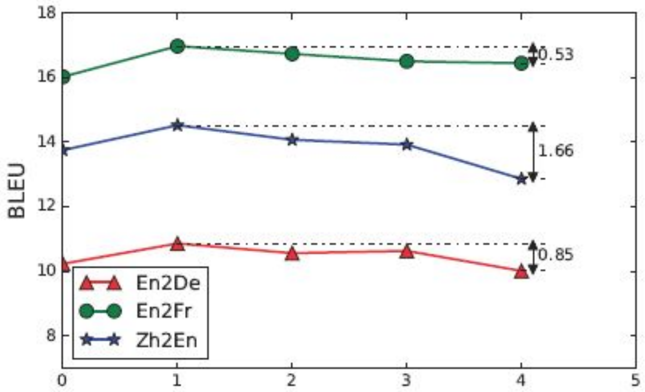
Embedding reinforced encoder: Combine the initial output sequence with the fixed cross lingual embeddings

Denoising auto-encoding: shuffle input sentences

Best translation on one layer sharing

More distance between language pair, more different characteristics

Shared layers are vital

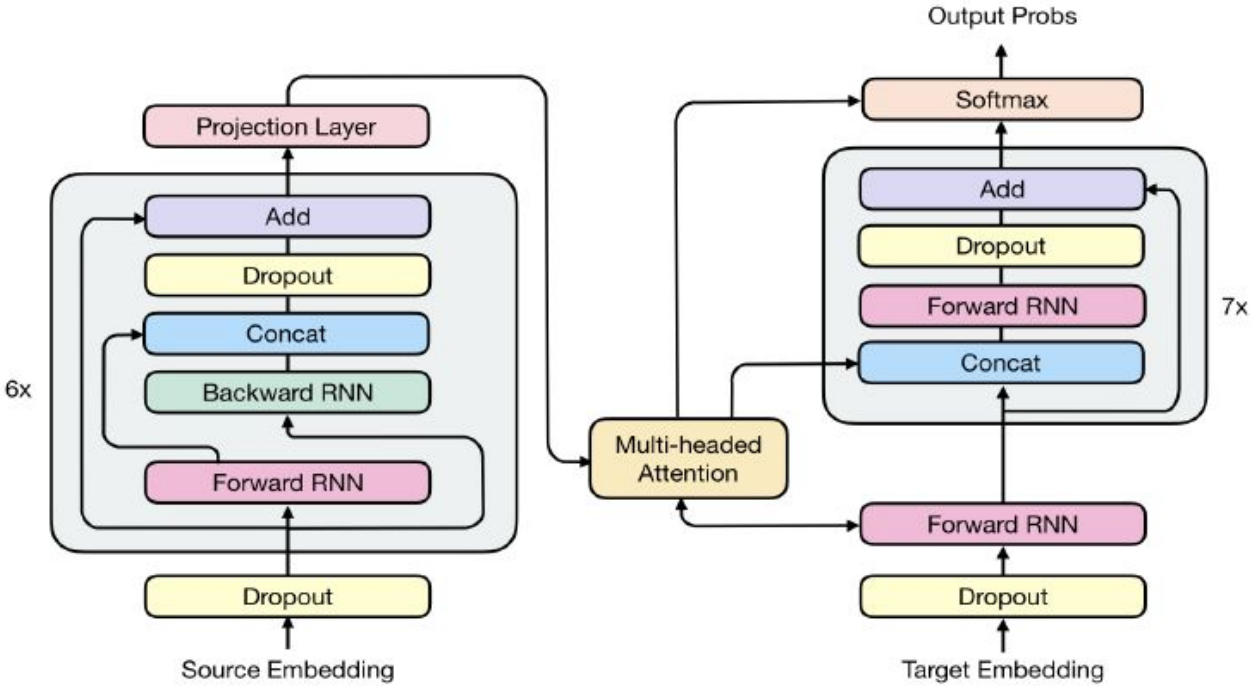


[2] Yang, Z., Chen, W., Wang, F., & Xu, B. (2018). Unsupervised Neural Machine Translation with Weight Sharing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)

The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation

- Goal: Identify key modeling and training techniques; Devise new hybrid architectures to combine strengths
- RNMT: Pros: Sequential; Cons: Cannot parallelize, dilemma of trainability vs expressivity
- ConvS2S: Pros: Parallelize; Cons: Fixed and narrow receptive field
- Transformer: Pros: Extended receptive fields of features from entire sequence, strict computation sequence, final output normalized to prevent blow up; Cons: Lack a memory component

Overview



[3] Chen, M. X., Firat, O., Bapna, A., Johnson, M., Macherey, W., Foster, G., ... & Wu, Y. (2018). The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)

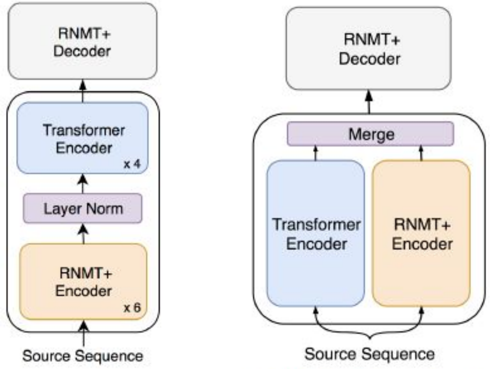
Hybrid NMT model

Assessing Individual Encoders and Decoders

Encoder	Decoder	En→Fr Test BLEU
Trans. Big	Trans. Big	40.73 ± 0.19
RNMT+	RNMT+	41.00 ± 0.05
Trans. Big	RNMT+	**41.12 ± 0.16**
RNMT+	Trans. Big	39.92 ± 0.21

Assessing Encoder Combinations

Model	En→Fr BLEU	En→De BLEU
Trans. Big	40.73 ± 0.19	27.94 ± 0.18
RNMT+	41.00 ± 0.05	28.49 ± 0.05
Cascaded	**41.67 ± 0.11**	28.62 ± 0.06
MultiCol	41.66 ± 0.11	**28.84 ± 0.06**



(a) Cascaded Encoder (b) Multi-Column Encoder

[3] Chen, M. X., Firat, O., Bapna, A., Johnson, M., Macherey, W., Foster, G., ... & Wu, Y. (2018). The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)

Thank you