Text Revision by On-the-Fly Representation Optimization

Jingjing Li[1], **Zichao Li**[2], **Tao Ge**[3], **Irwin King**[1], **Michael R. Lyu**[1]
[1] The Chinese University of Hong Kong
[2] McGill University [3] Microsoft Research Asia
llee.jingjing@gmail.com zichao.li@mail.mcgill.ca
tage@microsoft.com {king, lyu}@cse.cuhk.edu.hk

Abstract

Text revision refers to a family of natural language generation tasks, where the source and target sequences share moderate resemblance in surface form but differentiate in attributes, such as text style transfer (Shen et al., 2017), text simplification (Xu et al., 2016), counterfactual debiasing (Zmigrod et al., 2019), grammar error correction (Sun et al., 2022) and sentence fusion (Malmi et al., 2019).

As the most popular solution, sequence-to-sequence (seq2seq) learning achieves state-of-the-art results on many text revision tasks today. However, it becomes less applicable when there is no large-scale annotated parallel data for training.

With recent breakthroughs in self-supervised learning have enabled the pre-trained Transformer models (Vaswani et al., 2017), such as BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019) and GPT (Radford et al., 2020), to learn sufficient distributed representation of natural language, which is universally transferable to a wide range of downstream tasks even without labeled data (Tenney et al., 2019; Zhang et al., 2019; Wu et al., 2020). In this work, we borrow the power of a pre-trained Transformer for text revision without any parallel data.

In this paper, we propose OREO, a method of On-the-fly REpresentation Optimization for text revision. Instead of generating an entire sequence of tokens from scratch, OREO first detects partial text span to be edited, then conducts in-place span revision:

Step 1: Representation optimization Given an input sentence $X^{(i)}$ at the i-th iteration, RoBERTa parameterized by θ transforms it to a sequence of hidden states $H^{(i)}$, conditioned on which the attribute head estimates the probability of target attribute $P_{W_{\text{Att}}}(z^*|H^{(i)})$. Then, for each revision, we find a small local perturbation on $H^{(i)}$ that maximally increases the likelihood of target attribute. As such, the update rule of hidden states is:

$$H^{(i+1)} = H^{(i)} - \lambda \frac{\nabla_{H^{(i)}}\mathcal{L}}{\|\nabla_{H^{(i)}}\mathcal{L}\|_2}, \qquad (1)$$

where λ is a hyper-parameter that controls the norm of perturbation, and

$$\mathcal{L} = -\log P_{W_{\text{Att}}}(z^*|H^{(i)}). \qquad (2)$$

Step 2: Span replacement After hidden states are updated, OREO conducts span replacement. We calculate magnitude of $\nabla_{H^{(i)}}\mathcal{L}$ for i-th token, where $\mathcal{L}$ is calculated with (2), and select the span with largest magnitude. The selected span $X^{(i)}_{t:t+N}$ of length N is replaced by [LM-MASK] tokens. RoBERTa takes as input the masked sequence, and predicts a new span autoregressively with the previously updated hidden states.

The training for OREO is simple: we fine-tune the RoBERTa model with masked language modeling and attribute classification jointly. The first objective forces RoBERTa to infill a span consistent with the semantics and attributes represented by hidden states, while the latter one steers the hidden states towards a desired attribute.

We experiment with two fundamental revision tasks, text simplification and formalization. In text simplification, our method surpassed the supervised baseline by 4.2 SARI score and unsupervised baseline 5.3 SARI score on Newsela-turk (Maddela et al., 2020). In text formalization, our approach outperforms all of the unsupervised baseline models in terms of content preservation and formality on GYAFC-fr (Rao and Tetreault, 2018). Ablation study is conducted to validate the design of each component in the model, through which we have following key findings: (1) representation optimization is essential to formality metrics; (2) infilling conditioned on hidden states helps preserve content; (3) our gradient-guided span selection contributes to both of them.[1]

[1]This paper was originally published at AAAI 2022. Access the full version here.

Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022), pages 58 - 59
May 26, 2022 ©2022 Association for Computational Linguistics

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Mounica Maddela, Fernando Alva-Manchego, and Wei Xu. 2020. Controllable text simplification with explicit paraphrasing. *arXiv preprint arXiv:2010.11004*.

Eric Malmi, Sebastian Krause, Sascha Rothe, Daniil Mirylenka, and Aliaksei Severyn. 2019. Encode, tag, realize: High-precision text editing. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5057–5068.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2020. Improving language understanding by generative pre-training.

Sudha Rao and Joel R. Tetreault. 2018. Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer. In *NAACL-HLT*.

Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. In *NIPS*, pages 6830–6841.

Xin Sun, Tao Ge, Shuming Ma, Jingjing Li, Furu Wei, and Houfeng Wang. 2022. A unified strategy for multilingual grammatical error correction with pre-trained cross-lingual language model. *arXiv preprint arXiv:2201.10707*.

Ian Tenney, Dipanjan Das, and Ellie Pavlick. 2019. Bert rediscovers the classical nlp pipeline. In *ACL*, pages 4593–4601.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*.

Zhiyong Wu, Yun Chen, Ben Kao, and Qun Liu. 2020. Perturbed masking: Parameter-free probing for analyzing and interpreting bert. In *ACL*, pages 4166–4176.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *TACL*, 4:401–415.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *ArXiv*, abs/1904.09675.

Ran Zmigrod, Sabrina J. Mielke, Hanna Wallach, and Ryan Cotterell. 2019. Counterfactual data augmentation for mitigating gender stereotypes in languages with rich morphology. In *ACL*, pages 1651–1661.