

Bidirectional Transformer Reranker for Grammatical Error Correction

Ying Zhang¹ Hidetaka Kamigaito² Manabu Okumura^{1,3}

¹RIKEN Center for Advanced Intelligence Project

²NARA Institute of Science and Technology (NAIST)

³Institute of Science Tokyo

ying.zhang@riken.jp kamigaito.h@is.naist.jp oku@lr.pi.titech.ac.jp

掲載号の情報

31 巻 1 号 pp. 3-46.

doi: <https://doi.org/10.5715/jnlp.31.3>

Abstract

Pre-trained sequence-to-sequence (seq2seq) models have achieved state-of-the-art results in the grammatical error correction tasks. However, these models are plagued by prediction bias owing to their unidirectional decoding (Figure 1a). Thus, this study proposed a bidirectional transformer reranker (BTR) that re-estimates the probability of each candidate sentence generated by the pre-trained seq2seq model (Figure 2). The BTR preserves the seq2seq-style transformer architecture but utilizes a BERT-style self-attention mechanism in the decoder to compute the probability of each target token using masked language modeling to capture bidirectional representations from the target context (Figure 1b). To guide the reranking process, the BTR adopted negative sampling in the objective function to minimize the unlikelihood. During inference, the BTR yielded the final results after comparing the reranked top-1 results with the original ones using an acceptance threshold λ . Experimental results showed that, when reranking candidates from a pre-trained seq2seq model, the T5-base, the BTR on top of T5-base yielded scores of 65.47 and 71.27 $F_{0.5}$ on the CoNLL-14 and BEA test sets, respectively, and yielded 59.52 GLEU score on the JFLEG corpus, with improvements of 0.36, 0.76, and 0.48 points compared with the original T5-base. Furthermore, when reranking candidates from T5-large, the BTR on top of T5-base improved the original T5-large by 0.26 on the BEA test set.



(a) Unidirectional

(b) Bidirectional

Figure 1: Decoding patterns for generating a sequence $\mathbf{x} = (x_1, x_2, x_3)$.

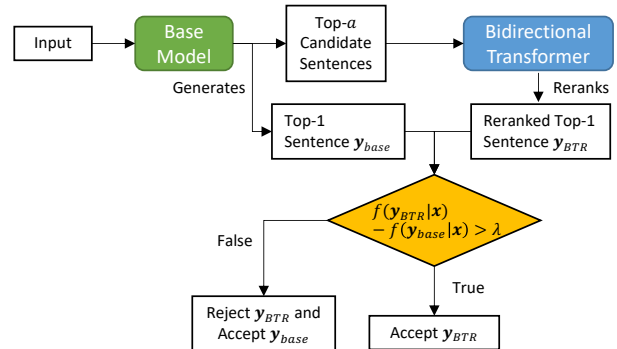


Figure 2: Overview of the reranking procedure by using the bidirectional transformer reranker (BTR).

Candidate	Accept	Reject	Equal
Proportion(%)	12.50	21.11	66.39
y_{base}	61.67	61.66[†]	68.78
y_{BTR}	63.97[†]	57.28	68.78

Table 1: Results on the CoNLL-14 dataset. The first block shows the proportions (%) of final decisions regarding the suggestions from y_{BTR} : Accept, Reject, or Equal (indicating the suggestion y_{BTR} matches the original selection y_{base}). The second block reports the $F_{0.5}$ scores for the sentences selected by the base model (y_{base}) and the BTR (y_{BTR}) for each decision type.