World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:16, No:11, 2022

Exploring Bidirectional Encoder Representations from the Transformers' Capabilities to Detect English Preposition Errors

Authors : Dylan Elliott, Katya Pertsova

Abstract : Preposition errors are some of the most common errors created by L2 speakers. In addition, improving error correction and detection methods remains an open issue in the realm of Natural Language Processing (NLP). This research investigates whether the bidirectional encoder representations from the transformers model (BERT) have the potential to correct preposition errors accurately enough to be useful in error correction software. This research finds that BERT performs strongly when the scope of its error correction is limited to preposition choice. The researchers used an open-source BERT model and over three hundred thousand edited sentences from Wikipedia, tagged for part of speech, where only a preposition edit had occurred. To test BERT's ability to detect errors, a technique known as multi-level masking was used to generate suggestions based on sentence context for every prepositional environment in the test data. These suggestions were compared with the original errors in the data and their known corrections to evaluate BERT's performance. The suggestions were further analyzed to determine if BERT more often agreed with the judgements of the Wikipedia editors. Both the untrained and fined-tuned models were compared. Finetuning led to a greater rate of error-detection which significantly improved recall, but lowered precision due to an increase in false positives or falsely flagged errors. However, in most cases, these false positives were not errors in preposition usage but merely cases where more than one preposition was possible. Furthermore, when BERT correctly identified an error, the model largely agreed with the Wikipedia editors, suggesting that BERT's ability to detect misused prepositions is better than previously believed. To evaluate to what extent BERT's false positives were grammatical suggestions, we plan to do a further crowd-sourcing study to test the grammaticality of BERT's suggested sentence corrections against native speakers' judgments.

Keywords : BERT, grammatical error correction, preposition error detection, prepositions
Conference Title : ICNLPCC 2022 : International Conference on Natural Language Processing and Cognitive Computing
Conference Location : San Francisco, United States
Conference Dates : November 03-04, 2022