BERT-based Contextual Semantic analysis for English Preposition Error Correction

Guimin Huang, Jiahao Wang, Hui Tang, Xuan Ye

School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541000, China
sendhuang@126.com
gmhuang@guet.edu.cn

Abstract. In the writing of English as a Second Language (ESL) learners, preposition related errors account for nearly 30 percent in the all grammatical errors (Bitchener et al. 2005)[1]. Unlike verbs and nouns, isolated preposition does not mean anything, but in different contexts, prepositions usually have different meanings so that ESL students hardly master how to use all prepositions correctly. In this publication, we present a preposition error correction (PEC) model based BERT (Devlin et al., 2018)[2] which contains rich contextual semantic information, fine-tuned BERT models to intensify sentence semantic in output, and calculate the preposition matching scores according to the formula.

1. Introduction

For ESL learners, learning English usually begins with learning the words and their main meanings. In this way ESL learners can basically understand the use of most words, but it is difficult for prepositions. One analysis has found that preposition errors are the second most error type in essays written by ESL learners (Leacock et al., 2010)[3]. The using of prepositions largely depends on the semantics of sentences, such as “I lived in London but I lived at New York.” There are many cases like this, only native English speakers are likely to be faultless in any cases, but it is hard also for them to explain the subtle differences between two prepositions about their usage. Obviously, mastering the usage is almost an impossible task for non-native (L2) English learners.

The model combining random forest classifier and neural network is used to detect preposition errors (Monaikul et al., 2020)[4]. The value of recall and F1-Score were higher than Bert model (16.81% and 17.6% respectively), but the accuracy was 14.36% lower[8]. It is a challenge to correct more preposition errors in the field of Grammar Error Correction(GEC).

This paper introduces an improved preposition error correction method based on Bert model, which detects and corrects preposition errors by strengthening sentence semantics, reaching an accuracy of 84.16% and an F0.5-Score of 43.29%.

2. Related Work

In the past, there have been few studies specific to L2 learners, especially for automatic grammar correction and PEC[5]. Most scholars use maximum entropy model to evaluate the appropriateness of a preposition in the text written by L2 learners. De Felice et al. (2008)[6] trained a maximum entropy classifier to detect preposition errors through context features, finally used the model to predict the input sentences and correct the existing usage errors. Gamon et al. (2008)[7] used decision tree and
language model, a complex system is designed to correct the errors of prepositions and determiner, achieving 64.93% accuracy. In a different way, Han et al. (2010) used annotated corpus to train the classifier instead of paying attention to the appropriateness of prepositions, which can make the result more suitable for ESL learners’ writing. Only in recent years has there been a great improvement on automatic grammar error correction, mainly due to the rapid development of deep learning. The system combining random forest classifiers and neural network analysed the features between prepositions, verbs and nouns to predict correct words, finally achieved 62.58% precision and 30.82% recall. Natawut Monaikul and Barbara Di Eugenio (2020) fine-tuned a pre-trained BERT model to predict preposition in the masked position achieving higher 76.94% precision, but only 14.01% recall.

3. Methods

3.1. Description of methods

The core of Bert model is transformer structure, in which the self-attention mechanism is used to judge the degree of dependence between words in a sentence. We first construct a confounding set of prepositions that alternate prepositions in sentences and obtain the weights of related words through the normalization layer.

The whole encoder is composed of 6 identical layers. Taking a single layer as an example, we focus on the weight of attention layer. Input embedding compute with three matrices Wq, Wk, and Wv in different dimensions, and then get three matrix of Queries, Keys and Values in multiple dimensions also. The importance of the current word to the other words in the sentence can be calculated by the next formula.

$$Z = \text{softmax} \left( \frac{q \times K^T}{\sqrt{d_k}} \right) V$$

The dk is the dimension of Keys and we use the original 64 dimensional vector in this paper. Scaling is done to prevent the probability distribution calculated by softmax from being too extreme and resulting in too small a gradient for parameter updates.

Concatenate all the attention heads and multiply with a weight matrix WO that was trained jointly with the model. Finally we get the result of Z matrix that captures information from all the attention heads.
We optimize the output $Z$ and focus on the relevance of prepositions in a sentence. A preposition in a sentence is usually influenced by some certain words or the meaning of whole sentences, so we use the output matrix to represent the probability of relevancy between words.

We use fluency to measure whether a sentence deviates from the correct meaning and calculate it by the following formula.

$$f(x) = \frac{1}{1 + H(x)}$$

We only focus on the words that are related to the preposition in the sentence and regard them as the context of the preposition.

$$H(x) = - \sum_{i=1}^{\left| x \right|} Z_i$$

It is need to note that we are not calculating the matrix $Z$ directly as a parameter, but processing the output matrix to make it more like a probability of context: $\log P(x_i \mid x_{<i})$.

![Fig.2 The computational process of the model](image)

### 3.2. Experimental data

The prepositions are replaced by a preposition confusion set, and the sentence with the highest degree of fluency is extracted as an alternative correction. In Table 1, we list the prepositions in the confusion set and their frequency in the training set and testing set respectively.

| Prep. | Freq. in training data | Freq. in test data |
|-------|------------------------|--------------------|
| to    | 40175                  | 836                |
| in    | 27762                  | 517                |
| at    | 26305                  | 436                |
| of    | 22872                  | 458                |
| on    | 19030                  | 402                |
| for   | 16563                  | 316                |
| with  | 5890                   | 168                |
| from  | 4631                   | 88                 |
| by    | 4065                   | 126                |
| about | 2310                   | 69                 |
3.3. Experimental results
In training the model we used 80% of the International Corpus Network of Asian Learners of English (ICNALE) corpus for training and selected the ten prepositions with the highest frequency to form a preposition confusion set. To evaluate our model on ESL text, we used 20% of the (ICNALE) corpus as testing dataset, which contains 1,120 essays written by ESL. Finally we set the sentence fluency range to (0, 1) and use a classifier to filter the possible corrections.

| Tab.2 Comparison with BERT model |
|----------------------------------|
|                                | Precision | Recall | F0.5 - Score |
| BERT                            | 78.54     | 15.32  | 43.03        |
| our                             | 84.16     | 14.71  | 43.29        |

Table 2 shows the comparison between our model and Bert model. We have a certain improvement over the original Bert model in precision and our F0.5-Score has a slight increase.

4. Conclusion
We propose a new method based on Bert model, which can improve the accuracy of preposition error correction. As for the preposition errors that L2 learners often make in writing, we correct these errors by semantic model. Although BERT model has state-of-the-art results in many aspects, it is not suitable for preposition error correction. Through our design, the preposition error correction model based on BERT can well correct preposition errors. The results of experiment show that our method has higher accuracy and F value in preposition correction than the original Bert model.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (No. 61662012, No.61662015).

References
[1] Bitchener J, Young S, Cameron D. The effect of different types of corrective feedback on ESL student writing[J]. Journal of second language writing, 2005, 14(3): 191-205.
[2] 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.
[3] ZHEN Yi, Cheng Shuwei, et al. The caculation of shear slippage of light-weight aggregate concrete composite floor slabs[J].Industrial Construction, 2007,37(s1):579-591(in Chinese).
[4] Leacock, C.; Chodorow, M.; Gamon, M.; and Tetreault, J. 2010. Automated grammatical error detection for language learners. Synthesis Lectures on Human Language Technologies 3(1):1–134.
[5] Monaikul N, Di Eugenio B. Detecting Preposition Errors to Target Interlingual Errors in Second Language Writing[C]/The Thirty-Third International Flairs Conference. 2020.
[6] Izumi, Emi, Kiyotaka Uchimoto, and Hitoshi Isahara. 2004. SST speech corpus of Japanese learners’ English and automatic detection of learners’ errors. ICAME, 28:31–48.
[7] De Felice, R., and Pulman, S. G. 2008. A classifier-based approach to preposition and determiner error correction in L2 English. In Proceedings of the 22nd International Conference on Computational Linguistics, 169–176.
[8] Gamon, M., J. Gao, C. Brockett, A. Klementiev, W. Dolan, D. Belenko, and L. Vanderwende. 2008. Using contextual speller techniques and language modeling for ESL error correction. In Proceedings of IJCNLP.