BERT-based Lexical Substitution

Wangchunshu Zhou[1] * **Tao Ge**[2] **Ke Xu**[1] **Furu Wei**[2] **Ming Zhou**[2]
[1]Beihang University, Beijing, China
[2]Microsoft Research Asia, Beijing, China
zhouwangchunshu@buaa.edu.cn, kexu@nlsde.buaa.edu.cn
{tage, fuwei, mingzhou}@microsoft.com

Abstract

Previous studies on lexical substitution tend to obtain substitute candidates by finding the target word's synonyms from lexical resources (e.g., WordNet) and then rank the candidates based on its contexts. These approaches have two limitations: **(1)** They are likely to overlook good substitute candidates that are not the synonyms of the target words in the lexical resources; **(2)** They fail to take into account the substitution's influence on the global context of the sentence.

To address these issues, we propose an end-to-end BERT-based lexical substitution approach which can propose and validate substitute candidates without using any annotated data or manually curated resources. Our approach first applies dropout to the target word's embedding for partially masking the word, allowing BERT to take balanced consideration of the target word's semantics and contexts for proposing substitute candidates, and then validates the candidates based on their substitution's influence on the global contextualized representation of the sentence. Experiments show our approach performs well in both proposing and ranking substitute candidates, achieving the state-of-the-art results in both LS07 and LS14 benchmarks.

1 Introduction

Lexical substitution (McCarthy and Navigli, 2007) aims to replace a target word in a sentence with a substitute word without changing the meaning of the sentence, which is useful for many Natural Language Processing (NLP) tasks like text simplification and paraphrase generation.

One main challenge in this task is proposing substitutes that not only are semantically consistent with the original target word and fits in the context but also preserve the sentence's meaning. Most previous approaches to this challenge first obtain substitute candidates by picking synonyms from manually curated lexical resources as candidates, and then rank them based on their appropriateness in context, or instead ranking all words in the vocabulary to avoid the usage of lexical resources. For example, knowledge-based lexical substitution systems (Yuret, 2007; Hassan et al., 2007) use pre-defined rules to score substitute candidates; vector space modeling approach (Erk and Padó, 2008; Dinu and Lapata, 2010; Thater et al., 2010; Apidianaki, 2016) uses distributional sparse vector representations based on the syntactic context; substitute vector approach (Yuret, 2012; Melamud et al., 2015b) comprises the potential fillers for the target word slot in that context; word/context embedding similarity approach (Melamud et al., 2015a; Roller and Erk,

Sentence	The wine he sent to me as my birthday gift is too **strong** to drink.
WordNet	hard, solid, **stiff**, firm
BERT (keep target word)	stronger, strongly, hard, much
BERT (mask target word)	hot, thick, sweet, much
BERT (embedding dropout)	tough, **powerful**, potent, hard

(a)

Sentence	The wine he sent to me as my birthday gift is too **strong** to drink.
✗	The wine he sent to me as my birthday gift is too **hot** (0.81) to drink. (0.86)
✗	The wine he sent to me as my birthday gift is too **tough** (0.91) to drink. (0.92)
✓	The wine he sent to me as my birthday gift is too **powerful** (0.91) to drink. (**0.93**)

(b)

Figure 1: **(a)** WordNet and original BERT cannot propose the valid substitute *powerful* in their top-K results but applying target word embedding dropout enables BERT to propose it; **(b)** Undesirable substitutes (e.g., *hot*, *tough*) tend to change the contextualized representation of the sentence more than good substitutes (e.g., *powerful*). The numbers after the words are the cosine similarity of the words' contextualized vector to the original target words; while the numbers after the sentence are the similarity of the sentence's contextualized representation before and after the substitution, defined in Eq (2).

*This work was done during the first author's internship at Microsoft Research Asia.

Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3368–3373
Florence, Italy, July 28 - August 2, 2019.