"JOURNAL OF SCIENCE-INNOVATIVE RESEARCH IN UZBEKISTAN" JURNALI
VOLUME 1, ISSUE 8, 2023. NOVEMBER
ResearchBib Impact Factor: 8.654/2023 ISSN 2992-8869

Lexical Substitution: Enhancing Language Generation with Word Replacement

Jurayeva Zarnigor Boʻstonovna
Qosimboyeva Durdona Qobiljon qizi
Students, Tashkent State Pedagogical University, Uzbekistan
E-mail: zarnigorjurayeva6@gmail.com
durdona.qosimboyeva@icloud.com
Astanova Dilafruz Murodovna,
Professor, Tashkent State Pedagogical University, Uzbekistan
Email: dilafruza19870714@mail.com

Abstract

Lexical substitution (LS) is a technique used in natural language processing (NLP) to replace words or phrases in a sentence while preserving the original meaning. Recent advancements in LS based on pretrained language models have shown promising results in suggesting suitable replacements for target words by considering their context. This article explores LS approaches using neural language models (LMs) and masked language models (MLMs) such as context2vec, ELMo, BERT, and XLNet. The study demonstrates the effectiveness of injecting target word information into these models and analyzes the semantic links between targets and substitutes. Lexical substitution plays a vital role in enhancing language generation models and is widely used in NLP tasks such as data augmentation, paraphrase generation, word sense induction, and text simplification. While earlier methods relied on manual lexical resources, the emergence of contextual language models like BERT, ELMo, and XLNet has revolutionized LS by incorporating contextual information and achieving state-of-the-art results. Ongoing research aims to develop more context-aware approaches to address the challenges of lexical substitution, ultimately advancing the capabilities of NLP systems.

Keywords:Lexical substitution, Contextual language models, BERT (Bidirectional Encoder Representations from Transformers), ELMo (Embeddings from Language Models), XLNet, Context-sensitive modeling, State-of-the-art (SOTA) results, Lexical resources, Paraphrasing,Sentiment analysis, Data augmentation,Rule-based methods, Word embedding-based methods, Machine learning-based methods, Context preservation, Grammatical accuracy, Syntactic organization,Natural language processing (NLP)

The goal of lexical substitution (LS) is to identify suitable replacements for a target word in a sentence. Recent advances in LS approaches based on pretrained language models have been impressive, suggesting plausible replacements for a target word through examination of its context. The process of finding a word that can be substituted in a clause's context is known as lexical substitution. A replacement game may be offered, for instance, if the language was "After the match, replace any remaining fluid deficit to prevent chronic dehydration throughout the tournament."

Comparative study of Lexical Substitution Approaches based on Neural Language Models

Lexical substitution, a technique used in natural language processing (NLP), is the process of changing words or phrases in a sentence with new words or phrases while maintaining the sense of the original text. Word sense induction, lexical relation extraction, data augmentation, and other NLP applications can all be supported by the incredibly potent technology known as lexical substitution in context. The task of lexical substitution was applied to popular neural language and masked language models (LMs and MLMs), including context2vec, ELMo, BERT, and XLNet. In this research, we give a thorough comparative study of these models. We demonstrate how appropriately injecting target word information can enhance the already competitive outcomes produced by SOTA LMs/MLMs, and we contrast several target injection techniques. The types of semantic links between the target and substitutes generated by various models are also analyzed, offering insights into the kinds of words that are really generated or provided as substitutes by annotators.Lexical substitution is an important method for enhancing language generation models since it significantly contributes to the quality and diversity of generated text. . Lexical substitution is used in a wide range of NLP tasks like data augmentation, paraphrase generation, word sense induction, or text simplification (Shardlow, 2014; Amrami and Goldberg, 2018). Earlier, methods typically relied entirely on manually curated lexical resources like WordNet (Miller, 1995). The synonyms obtained from such resources were then ranked based on their suitability evaluated by a similarity metric and predefined rules. Some approaches used vector-based modelling and distributional vectors based on syntactic context to obtain the most suitable synonyms (Melamud et al., 2015b). Recent advances in contextual language models like Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019), Embeddings from Language Models (ELMo) (Peters et al., 2018), and XLNet (Yang et al., 2019) have resulted in major breakthroughs in NLP. Because these models carry contextual information and have the ability of context-sensitive modelling of word probabilities, they have achieved the state-of-the-art (SOTA) results in lexical substitution as well. Some recent research efforts have improved lexical substitution by modifying the architecture of contextual embedding models (Zhou et al., 2019) whereas others integrated lexical resources to contextual embeddings to obtain the most suitable set of substitutes (Michalopoulos et al., 2022).

Understanding Lexical Substitution

Lexical substitution is the process of changing a word within a sentence while keeping the overall meaning and context. The word being changed should be a good replacement that flows seamlessly with the rest of the phrase. Take the following example, for instance:

Original Sentence: "The cat is sitting on the mat."

Lexical Substitution: "The feline is resting on the rug."

In this example, the words "cat" and "mat" are substituted with "feline" and "rug," respectively. The substituted words maintain the original meaning but add variety and expressiveness to the sentence.

Applications of Lexical Substitution

1. Text Generation: Lexical replacement is a common technique for varying and enhancing generated text. Language models can generate different iterations of the same material by swapping out words in a sentence, improving the output and avoiding repeating patterns.

2. Paraphrasing: When paraphrasing, lexical substitution is used to create alternate sentences while maintaining the sense of the original. It can be used to

improve the readability and clarity of text as well as create different training data for machine learning models.

3. Sentiment Analysis: To examine how various words affect the sentiment of a sentence, lexical substitution can be used in sentiment analysis. Sentiment analysis algorithms can learn more about the impact of particular vocabulary selections on the overall sentiment by replacing words with synonyms or antonyms.

4. Data Augmentation: In data augmentation, which involves producing more training data by transforming existing samples, lexical substitution is a useful strategy. Data augmentation can increase the generalization and robustness of NLP models by changing the words in phrases. Methods and Challenges:

Lexical replacement can be accomplished using a variety of techniques, including as rule-based methods, word embedding-based methods, and machine learning-based methods. To locate appropriate replacements, rule-based approaches rely on established lexical resources like dictionaries and thesauri. Word vector representations are used by word embedding-based algorithms to identify alternatives that are semantically equivalent. Machine learning techniques train models that can anticipate optimal substitutions using datasets that have been annotated. Lexical replacement, however, often presents difficulties. It might be difficult to find appropriate substitutions that preserve the context's meaning because words frequently have several meanings and depend on their surroundings. Additionally, when replacing words, significant thought must be given to maintaining grammatical accuracy and syntactic organization.

Lexical substitution is a valuable technique in natural language processing (NLP) that involves replacing words or phrases in a sentence while preserving the original meaning. Recent advancements in LS using neural language models (LMs) and masked language models (MLMs) have shown promising results, particularly with models like BERT, ELMo, and XLNet that incorporate contextual information. These models have revolutionized LS by achieving state-of-the-art results and enhancing the capabilities of NLP systems.

The study conducted a comparative analysis of LS approaches based on various LMs and MLMs, highlighting the importance of injecting target word information into these models. It also analyzed the semantic links between target words and substitutes, providing insights into the types of generated substitutes

Reference:

1) Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171-4186.

2)Melamud, O., Goldberger, J., & Dagan, I. (2015). Context2vec: Learning Generic Context Embedding with Bidirectional LSTM. Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, 51-61.

3)Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2227-2237.

4)Shardlow, M. (2014). A Survey of Automated Lexical Substitution. Journal of Artificial Intelligence Research, 51, 457-503.

5)Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. Advances in Neural Information Processing Systems, 32, 5753-5763.

6).https://aclanthology.org/2022.coling-1.362.pdf