

How Domain Adaptation of BERT Improves Syntactic Parsing of Math Text

Runa Yoshida¹ Takuya Matsuzaki²

¹ Department of Applied Mathematics, Graduate School of Science, Tokyo University of Science

² The Faculty of Science Division I, Tokyo University of Science

1423532@ed.tus.ac.jp matuzaki@rs.tus.ac.jp

掲載号の情報

31 巻 4 号 pp. 1691-1716.

doi: <https://doi.org/10.5715/jnlp.31.1691>

概要

This study clarifies how the domain adaptation of bidirectional encoder representations from transformers (BERT) contributes to the syntactic analysis of mathematical texts and their limitations. Experimental results show that the domain adaptation of BERT is highly effective, even with a relatively small amount of raw in-domain data. This improves the accuracy of the syntactic dependency analysis by up to four points without any annotated in-domain data. By analyzing the improvement, we found that numerous errors involving the mathematical expressions have been corrected. Errors related to structures that are not frequent in the out-domain fine-tuning data were difficult to improve by only the domain adaptation of BERT. This study also revealed that the effectiveness of BERT depends on the representation of the mathematical expressions in the input. Among several different representations of mathematical expressions, the highest dependency accuracy was achieved using a simple method where an entire mathematical expression is replaced with a dedicated special token. ¹⁾

参考文献

- [1] Runa Yoshida and Takuya Matsuzaki. Bert の教師なし分野適応による数学問題テキスト構文解析の精度向上要因の分析. In **Proceedings of the 29th Annual Meeting of the Association for Natural Language Processing**, pp. 64–69, 2023. (in Japanese).

1) This paper is a revised version of [1], which has been submitted for Proceedings of the 29th Annual Conference of the Association for Natural Language Processing, 2023.