

NSGC - Neural Spell & Grammar Checker (en/pt)

Rafael Ito

June 2020

Abstract

The abstract is a short text to let people understand what is the document about. It should give the mains highlights of the documents. This document is a template to be use in the Final Project of the EA376E class. The recommendation is to use LaTeX by the application Overleaf.com, but it is also possible to use Google Docs or MS-Word.

<https://github.com/dl4nlp-rg/PF06-RIto>

1 Introduction

This is an example of paper citation: BERT [7] is a transformer. T5 [17] also uses transformers.

2 Methodology

This is an example of inserting a figure.

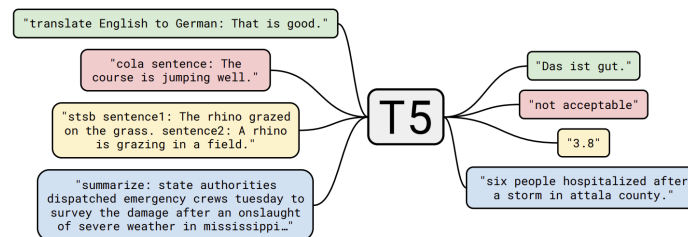


Figure 1: Figure example.

Figure 1 is an example of figure citation.

2.1 Metrics

2.1.1 M^2 (MaxMatch)

2.1.2 GLEU

2.1.3 Edit distance

3 Data set

3.1 CoNLL-2013

3.2 CoNLL-2014

3.3 JFLEG

3.4 BEA

3.5 ReGRA

4 Experiments

5 Conclusion

6 Future Work

6.1

BERT [7] T5 [17] Transformer [18] HOO [5] HOO-2011 [6] HOO-2012 [4] CoNLL-2013 [14] CoNLL-2014 [13] JFLEG [12] BEA [1] Encoder-Decoder [9] Soft-Masked BERT [20] GECToR [16] ReGRA [15] CoNLL-2013 [14] CoNLL-2014 [13] JFLEG [12] BEA [1] ReGRA [15] Write & Improve [19] LOCNESS [8] GLEU [10] GLEU improved [11] ERRANT [2] MaxMatch [3]

References

- [1] Christopher Bryant, Mariano Felice, Øistein E Andersen, and Ted Briscoe. The bea-2019 shared task on grammatical error correction. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 52–75, 2019.
- [2] Christopher Bryant, Mariano Felice, and Ted Briscoe. Automatic annotation and evaluation of error types for grammatical error correction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 793–805, Vancouver, Canada, July 2017. Association for Computational Linguistics.

- [3] Daniel Dahlmeier and Hwee Tou Ng. Better evaluation for grammatical error correction. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 568–572, 2012.
- [4] Robert Dale, Ilya Anisimoff, and George Narroway. Hoo 2012: A report on the preposition and determiner error correction shared task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 54–62, 2012.
- [5] Robert Dale and Adam Kilgarriff. Helping our own: Text massaging for computational linguistics as a new shared task. In *Proceedings of the 6th International Natural Language Generation Conference*, 2010.
- [6] Robert Dale and Adam Kilgarriff. Helping our own: The hoo 2011 pilot shared task. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 242–249, 2011.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [8] Sylviane Granger. *The computer learner corpus: a versatile new source of data for SLA research*. na, 1998.
- [9] Masahiro Kaneko, Masato Mita, Shun Kiyono, Jun Suzuki, and Kentaro Inui. Encoder-decoder models can benefit from pre-trained masked language models in grammatical error correction. *arXiv preprint arXiv:2005.00987*, 2020.
- [10] Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel Tetreault. Ground truth for grammatical error correction metrics. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 588–593, 2015.
- [11] Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel Tetreault. Gleu without tuning. *arXiv preprint arXiv:1605.02592*, 2016.
- [12] Courtney Napoles, Keisuke Sakaguchi, and Joel Tetreault. Jfleg: A fluency corpus and benchmark for grammatical error correction. *arXiv preprint arXiv:1702.04066*, 2017.
- [13] Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. The conll-2014 shared task on grammatical error correction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–14, 2014.

- [14] Hwee Tou Ng, Siew Mei Wu, Yuanbin Wu, Christian Hadiwinoto, and Joel Tetreault. The CoNLL-2013 shared task on grammatical error correction. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–12, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.
- [15] Maria das Graças Volpe Nunes and ON Oliveira Jr. O processo de desenvolvimento do revisor gramatical regra. In *Anais do XXVII SEMISH (XX Congresso Nacional da Sociedade Brasileira de Computação)*, volume 1, page 6, 2000.
- [16] Kostiantyn Omelianchuk, Vitaliy Atrasevych, Artem Chernodub, and Oleksandr Skurzhashkyi. Gector–grammatical error correction: Tag, not rewrite. *arXiv preprint arXiv:2005.12592*, 2020.
- [17] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*, 2019.
- [18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [19] Helen Yannakoudakis, Øistein E Andersen, Ardeshir Geranpayeh, Ted Briscoe, and Diane Nicholls. Developing an automated writing placement system for esl learners. *Applied Measurement in Education*, 31(3):251–267, 2018.
- [20] Shaohua Zhang, Haoran Huang, Jicong Liu, and Hang Li. Spelling error correction with soft-masked bert. *arXiv preprint arXiv:2005.07421*, 2020.