Literature Review paper for CS224U, Spring 2020.

Abhishek Goswami
Microsoft
agoswami@microsoft.com

Abstract

This is a literature review of the Natural Language Inference (NLI) task, particularly as it relates to adversarial data. We also review structural properties of BERT, with an eye towards how they relate to the adversarial setting.

1 General Problem

The general problem we are trying to address is Natural Language Inference (NLI), the task of determining if a premise sentence entails a hypothesis statement. In recent years, Transformer-based models have been shown to be very effective in this task.

In particular, we study the problem of Adversarial NLI (Nie et al., 2019) which poses a challenge to the state-of-the-art models. We then study the structural properties of a popular deep learning model, BERT (Devlin et al., 2019) from the perspective of what it does (Michel et al., 2019; Rogers et al., 2020; Clark et al., 2019).

2 Article Summary

In this section we provide summaries of several papers.

2.1 (Devlin et al., 2019)

2.2 (Nie et al., 2019)

2.3 (Michel et al., 2019)

2.4 (Rogers et al., 2020)

2.5 (Clark et al., 2019)

2.6 (Vaswani et al., 2017)

2.7 (McCoy et al., 2019)

3 Compare And Contrast

Point out the similarities and differences of the papers. Do they agree with each other? Are results seemingly in conflict? If the papers address different subtasks, how are they related? (If they are not related, then you may have made poor choices for a lit review...). This section is probably the most valuable for the final project, as it can become the basis for a lit review section..

4 Future Work

Future work

References

Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. 2019. What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

R Thomas McCoy, Ellie Pavlick, and Tal Linzen. 2019. Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference. *arXiv preprint arXiv:1902.01007*.

Paul Michel, Omer Levy, and Graham Neubig. 2019. Are sixteen heads really better than one? In *Advances in Neural Information Processing Systems*, pages 14014–14024.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. 2019. Adversarial nli: A new benchmark for natural language understanding. *arXiv preprint arXiv:1910.14599*.

Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. A primer in bertology: What we know about how bert works. *arXiv preprint arXiv:2002.12327*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.