KDABert: Knowledge Distillation with Generative Adversarial Networks for Bert Model Compression

闫森

2017111497

摘要

这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。这里是摘要。

1 介绍

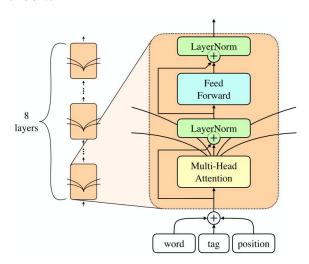这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。这里是介绍。

图 1: Transformer 结构。

2 实验

这里是实验。这里是实验。这里是实验。这里是实验。这里是实验。这里是实验。这里是实验。这里是实验。这里是实验。这里是实验。

这里是实验。

模型	LR	LP	F1
单模型			
Vinyals et al. (2015)	-	-	88.3
Dyer et al. (2016)	-	-	89.8
Zhu et al. (2013)	90.2	90.7	90.4
Cross and Huang (2016)	90.5	92.1	91.3
Liu and Zhang (2017b)	91.3	92.1	91.7
Stern et al. (2017a)	90.3	93.2	91.8
Liu and Zhang (2017a)	-	-	91.8
Shen et al. (2018)	92.0	91.7	91.8
Hong and Huang (2018)	91.5	92.5	92.0
Teng and Zhang (2018)	92.2	92.5	92.4
Stern et al. (2017b)	92.5	92.5	92.5
Kitaev and Klein (2018a)	93.2	93.9	**93.6**
非单模型			
Choe and Charniak (2016)	-	-	93.8
Liu and Zhang (2017a)	-	-	94.2
Fried et al. (2017)	-	-	94.7
Kitaev and Klein (2018a)	94.9	95.4	95.1
Kitaev and Klein (2018b)	95.6	96.0	**95.8**

表 1: 不同模型在 PTB 测试集上的最终结果。

致谢

这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。这里是致谢。

参考文献

Do Kook Choe and Eugene Charniak. 2016. Parsing as language modeling. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 2331–2336.

James Cross and Liang Huang. 2016. Span-based constituency parsing with a structure-label system and provably optimal dynamic oracles. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 1–11.

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. 2016. Recurrent neural network grammars. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 199–209.

Daniel Fried, Mitchell Stern, and Dan Klein. 2017. Improving neural parsing by disentangling model combination and reranking effects. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers*, pages 161–166.

Juneki Hong and Liang Huang. 2018. Linear-time constituency parsing with rnns and dynamic programming. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 477–483.

Nikita Kitaev and Dan Klein. 2018a. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 2675–2685.

Nikita Kitaev and Dan Klein. 2018b. Multilingual constituency parsing with self-attention and pre-training. *CoRR*, abs/1812.11760.

Jiangming Liu and Yue Zhang. 2017a. In-order transition-based constituent parsing. *Transactions of the Association for Computational Linguistics*, 5:413–424.

Jiangming Liu and Yue Zhang. 2017b. Shift-reduce constituent parsing with neural lookahead features. *Transactions of the Association for Computational Linguistics*, 5:45–58.

Yikang Shen, Zhouhan Lin, Athul Paul Jacob, Alessandro Sordoni, Aaron C. Courville, and

Yoshua Bengio. 2018. Straight to the tree: Constituency parsing with neural syntactic distance. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 1171–1180.

Mitchell Stern, Jacob Andreas, and Dan Klein. 2017a. A minimal span-based neural constituency parser. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 818–827.

Mitchell Stern, Daniel Fried, and Dan Klein. 2017b. Effective inference for generative neural parsing. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 1695–1700.

Zhiyang Teng and Yue Zhang. 2018. Two local models for neural constituent parsing. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 119–132.

Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey E. Hinton. 2015. Grammar as a foreign language. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 2773–2781.

Muhua Zhu, Yue Zhang, Wenliang Chen, Min Zhang, and Jingbo Zhu. 2013. Fast and accurate shift-reduce constituent parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*, pages 434–443.