Reference papers

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for lan- guage understanding. arXiv preprint arXiv:1810.04805..

[2] YinhanLiu,MyleOtt,NamanGoyal,JingfeiDu,MandarJoshi,Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoy- anov. 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692..

[3] JeremyHowardandSebastianRuder.2018.Universallanguagemodel fine-tuning for text classification. arXiv preprint arXiv:1801.06146.

[4] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

[5] AlexisConneau,KartikayKhandelwal,NamanGoyal,VishravChaud- hary, Guillaume Wenzek, Francisco Guzman, À Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsuper- vised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.

[6] Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019a. Ernie: Enhanced representation through knowledge integration. arXiv preprint arXiv:1904.09223.

[7] http://research.baidu.com/Blog/index-view?id=128.