Text Classification

A Preprint

Enwei Zhu

November 11, 2021

1 English Tasks

1.1 Settings

- Word embeddings are initialized with GloVe
- From scratch
  - Optimizer: Adadelta (lr=0.5)
  - Batch size: 64
  - Number of epochs: 50
- Fine tuning
  - Optimizer: AdamW (lr=1e-3/2e-3, ft_lr=1e-5)
  - Batch size: 32
  - Number of epochs: 10
  - Scheduler: Learning rate warmup at the first 20% steps followed by linear decay
  - PLMs are loaded with dropout rate of 0.2

1.2 Results

1.2.1 IMDb

Model	Paper	Reported Acc.	Our Imp. Acc.	Notes
LSTM + MaxPooling	–	–	91.58	num_layers=1
LSTM + Attention	McCann et al. [2017]	91.1	92.09	num_layers=1
BERT-base + Attention	Sun et al. [2019]	94.60	94.37	
RoBERTa-base + Attention	–	–	95.78	

Table 1: Results on IMDb.

1.2.2 Yelp Full

Model	Paper	Reported Acc.	Our Imp. Acc.	Notes
LSTM + MaxPooling	Zhang et al. [2015]	58.17	65.97	num_layers=2
LSTM + Attention	–	–	68.61	num_layers=2
BERT-base + Attention	Sun et al. [2019]	69.94	70.27	
RoBERTa-base + Attention	–	–	71.55	

Table 2: Results on Yelp Full.

1.2.3 Yelp 2013 (with User and Product IDs)

Model	Paper	Reported Acc.	Our Imp. Acc.	Notes
LSTM + MaxPooling	Chen et al. [2016]	62.7	64.96	num_layers=2
LSTM + Attention	Chen et al. [2016]	63.1	64.84	num_layers=2
BERT-base + Attention	–	–	68.76	
RoBERTa-base + Attention	–	–	70.80	

Table 3: Results on Yelp 2013.

2 Chinese Tasks

2.1 Settings

- Word-based (tokenized by jieba)
- Word embeddings are initialized by random or with Tencent embeddings [Song et al., 2018]
- From scratch
  - Optimizer: Adadelta (lr=1.0)
  - Batch size: 64
  - Number of epochs: 50
- Fine tuning
  - Optimizer: AdamW (lr=1e-3/2e-3, ft_lr=2e-5)
  - Batch size: 32
  - Number of epochs: 10
  - Scheduler: Learning rate warmup at the first 20% steps followed by linear decay
  - PLMs are loaded with dropout rate of 0.2
  - BERT refers to BERT-wwm [Cui et al., 2019]

2.2 Results

2.2.1 ChnSentiCorp

Model	Paper	Reported Acc.	Our Imp. Acc.	Notes
Multi-Channel CNN	Liu et al. [2018]	92.08		
LSTM + MaxPooling	–	–	92.25	num_layers=2
LSTM + Attention	–	–	92.42	num_layers=2
Tencent Embeddings + LSTM + MaxPooling	–	–	93.50	num_layers=2
Tencent Embeddings + LSTM + Attention	–	–	93.08	num_layers=2
BERT-base + Attention	Cui et al. [2019]	95.3	95.83	
RoBERTa-base + Attention	Cui et al. [2019]	95.8	95.08	

Table 4: Results on ChnSentiCorp.

2.2.2 THUCNews-10

Model	Paper	Reported Acc.	Our Imp. Acc.	Notes
LSTM + MaxPooling	–	–	97.66	num_layers=2
LSTM + Attention	–	–	97.24	num_layers=2
Tencent Embeddings + LSTM + MaxPooling	–	–	98.79	num_layers=2
Tencent Embeddings + LSTM + Attention	–	–	98.57	num_layers=2
BERT-base + Attention	Cui et al. [2019]	97.7	98.79	
RoBERTa-base + Attention	Cui et al. [2019]	97.8	98.98	

Table 5: Results on THUCNews-10.

References

Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin, and Zhiyuan Liu. Neural sentiment classification with user and product attention. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 1650–1659, 2016.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Ziqing Yang, Shijin Wang, and Guoping Hu. Pre-training with whole word masking for Chinese BERT. *arXiv preprint arXiv:1906.08101*, 2019.

Pengfei Liu, Ji Zhang, Cane Wing-Ki Leung, Chao He, and Thomas L Griffiths. Exploiting effective representations for Chinese sentiment analysis using a multi-channel convolutional neural network. *arXiv preprint arXiv:1808.02961*, 2018.

Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. Learned in translation: contextualized word vectors. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6297–6308, 2017.

Yan Song, Shuming Shi, Jing Li, and Haisong Zhang. Directional skip-gram: Explicitly distinguishing left and right context for word embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 175–180, 2018.

Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. How to fine-tune BERT for text classification? In *China National Conference on Chinese Computational Linguistics*, pages 194–206. Springer, 2019.

Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1*, pages 649–657, 2015.