Named Entity Recognition

A Preprint

Enwei Zhu

November 11, 2021

1 English Tasks

1.1 Settings

- Tagging scheme: BIOES
- Word embeddings are initialized with GloVe
- From scratch (sequence tagging)
  - Optimizer: SGD (lr=0.1)
  - Batch size: 32
  - Number of epochs: 100
- From scratch (span classification)
  - Optimizer: Adadelta (lr=1.0)
  - Batch size: 64
  - Number of epochs: 100
- Fine tuning
  - Optimizer: AdamW (lr=1e-3/2e-3, ft_lr=1e-5)
  - Batch size: 48
  - Number of epochs: 50
  - Scheduler: Learning rate warmup at the first 20% steps followed by linear decay
  - PLMs are loaded with dropout rate of 0.2
  - BERT-uncased models inputs are converted to truecase

1.2 Results

1.2.1 CoNLL 2003

Model	Paper	Reported F1	Our Imp. F1	Notes
CharLSTM + LSTM + CRF	Lample et al. [2016]	90.94	91.28	num_layers=1
CharCNN + LSTM + CRF	Ma and Hovy [2016]	91.21	90.70	num_layers=1
ELMo + Char + LSTM + CRF	Peters et al. [2018]	92.22 (0.10)	92.60	num_layers=1
Flair + Char + LSTM + CRF	Akbik et al. [2018]	93.09† (0.12)	92.60	num_layers=1
ELMo + Flair + Char + LSTM + CRF	–	–	92.67	num_layers=1
BERT-base + Softmax	Devlin et al. [2019]	92.4	92.02	
BERT-base + CRF	–	–	92.38	
BERT-base + LSTM + CRF	–	–	92.40	
BERT-large + Softmax	Devlin et al. [2019]	92.8	92.34	
BERT-large + CRF	–	–	92.64	
BERT-large + LSTM + CRF	–	–	92.80	
RoBERTa-base + Softmax	Liu et al. [2019]	–	92.39	
RoBERTa-base + CRF	–	–	92.59 (93.31‡)	
RoBERTa-base + LSTM + CRF	–	–	92.71 (93.39‡)	
RoBERTa-large + Softmax	Liu et al. [2019]	–	92.81	
RoBERTa-large + CRF	–	–	93.20 (93.37‡)	
RoBERTa-large + LSTM + CRF	–	–	93.26 (93.31‡)	
BERT-large-wwm + CRF	Devlin et al. [2019]	–	92.60	
BERT-large-wwm + LSTM + CRF	–	–	92.68	
ALBERT-base + CRF	Lan et al. [2019]	–	90.19	
ALBERT-base + LSTM + CRF	–	–	90.39	
ALBERT-xxlarge + CRF	Lan et al. [2019]	–	92.30	
ALBERT-xxlarge + LSTM + CRF	–	–	92.46	
SpanBERT-base + CRF	Joshi et al. [2020]	–	92.29	
SpanBERT-base + LSTM + CRF	–	–	92.27	
SpanBERT-large + CRF	Joshi et al. [2020]	–	93.07	
SpanBERT-large + LSTM + CRF	–	–	93.04	
SpERT (w/ CharLSTM + LSTM)	–	–	91.22	num_layers=2
SpERT (w/ BERT-base)	Eberts and Ulges [2020]	–	91.97	
SpERT (w/ BERT-base + LSTM)	–	–	92.62	
SpERT (w/ RoBERTa-base)	–	–	92.36	
SpERT (w/ RoBERTa-base + LSTM)	–	–	92.50	
Biaffine (w/ CharLSTM + LSTM)	–	–	91.05	num_layers=2
Biaffine (w/ BERT-base)	–	–	92.47	
Biaffine (w/ BERT-base + LSTM)	–	–	92.74	
Biaffine (w/ BERT-large)	Yu et al. [2020]	93.5†‡	92.67	
Biaffine (w/ RoBERTa-base)	–	–	92.56	
Biaffine (w/ RoBERTa-base + LSTM)	–	–	92.77	
Biaffine (w/ RoBERTa-large)	–	–	93.26	

Table 1: Results on CoNLL 2003. † means that both training and development splits are used for training (see Biaffine repo); ‡ means that document-level (cross-sentence) context is used.

1.2.2 OntoNotes 5 (CoNLL 2012)

Model	Paper	Reported F1	Our Imp. F1	Notes
CharLSTM + LSTM + CRF	Lample et al. [2016]	–	87.68	num_layers=2
CharCNN + LSTM + CRF	Chiu and Nichols [2016]	86.17 (0.22)	87.43	num_layers=2
ELMo + Char + LSTM + CRF	Peters et al. [2018]	–	89.71	num_layers=2
Flair + Char + LSTM + CRF	Akbik et al. [2018]	89.3†	89.02	num_layers=2
ELMo + Flair + Char + LSTM + CRF	–	–	89.55	num_layers=2
BERT-base + Softmax	Devlin et al. [2019]	–	89.35	
BERT-base + CRF	–	–	90.14	
BERT-base + LSTM + CRF	–	–	89.89	
Biaffine (w/ BERT-large)	Yu et al. [2020]	91.3†‡		
RoBERTa-base + Softmax	Liu et al. [2019]	–	90.22	
RoBERTa-base + CRF	–	–	90.83	
RoBERTa-base + LSTM + CRF	–	–	91.05	

Table 2: Results on OntoNotes 5. † means that both training and development splits are used for training (see Biaffine repo); ‡ means that document-level (cross-sentence) context is used.

2 Chinese Tasks

2.1 Settings

- Character-based
- Tagging scheme: BIOES
- From scratch (sequence tagging)
  - Optimizer: AdamW (lr=1e-3)
  - Batch size: 32
  - Number of epochs: 100
- Fine tuning
  - Optimizer: AdamW (lr=1e-3/2e-3, ft_lr=1e-5)
  - Batch size: 48
  - Number of epochs: 50
  - 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 MSRA (SIGHAN 2006)

Model	Paper	Reported F1	Our Imp. F1	Notes
LSTM + CRF	Zhang and Yang [2018]	88.81	89.49	num_layers=2
Bichar + LSTM + CRF	Zhang and Yang [2018]	91.87	92.02	num_layers=2
Lattice-LSTM + CRF	Zhang and Yang [2018]	93.18		
FLAT + CRF	Li et al. [2020]	94.35		
SoftLexicon + LSTM + CRF	Ma et al. [2020]	93.66	93.64	num_layers=2; Adamax (lr=1e-3)
BERT + CRF	Ma et al. [2020]	93.76	95.92	
BERT + LSTM + CRF	Ma et al. [2020]	94.83	96.18	
FLAT + BERT + CRF	Li et al. [2020]	96.09		
SoftLexicon + BERT + CRF	Ma et al. [2020]	95.42		
ERNIEv1 + CRF	Sun et al. [2019]	93.8*	95.87	
ERNIEv1 + LSTM + CRF	Sun et al. [2019]	–	96.24	
MacBERT-base + CRF	Cui et al. [2020]	–	95.72	
MacBERT-base + LSTM + CRF	Cui et al. [2020]	–	96.13	

Table 3: Results on MSRA (SIGHAN 2006). All experiments use testing split as development split (see SoftLexicon repo).

2.2.2 Weibo NER

Model	Paper	Reported F1	Our Imp. F1	Notes
LSTM + CRF	Zhang and Yang [2018]	52.77	50.19	num_layers=2
Bichar + LSTM + CRF	Zhang and Yang [2018]	56.75	57.18	num_layers=2
Lattice-LSTM + CRF	Zhang and Yang [2018]	58.79		
FLAT + CRF	Li et al. [2020]	63.42		
SoftLexicon + LSTM + CRF	Ma et al. [2020]	61.42	61.17	num_layers=2; Adamax (lr=5e-3)
BERT + CRF	Ma et al. [2020]	63.80	68.79	
BERT + LSTM + CRF	Ma et al. [2020]	67.33	70.48	
FLAT + BERT + CRF	Li et al. [2020]	68.55		
SoftLexicon + BERT + CRF	Ma et al. [2020]	70.50		
ERNIEv1 + CRF	Sun et al. [2019]	–	66.59	
ERNIEv1 + LSTM + CRF	Sun et al. [2019]	–	70.81	
MacBERT-base + CRF	Cui et al. [2020]	–	67.73	
MacBERT-base + LSTM + CRF	Cui et al. [2020]	–	70.71	
MacBERT-large + CRF	Cui et al. [2020]	–	70.01	
MacBERT-large + LSTM + CRF	Cui et al. [2020]	–	70.24	

Table 4: Results on Weibo NER v2.

2.2.3 Resume NER

Model	Paper	Reported F1	Our Imp. F1	Notes
LSTM + CRF	Zhang and Yang [2018]	93.48	94.93	num_layers=2
Bichar + LSTM + CRF	Zhang and Yang [2018]	94.41	94.51	num_layers=2
Lattice-LSTM + CRF	Zhang and Yang [2018]	94.46		
FLAT + CRF	Li et al. [2020]	94.93		
SoftLexicon + LSTM + CRF	Ma et al. [2020]	95.53	95.48	num_layers=2; Adamax (lr=2e-3)
BERT + CRF	Ma et al. [2020]	95.68	95.68	
BERT + LSTM + CRF	Ma et al. [2020]	95.51	95.97	
FLAT + BERT + CRF	Li et al. [2020]	95.86		
SoftLexicon + BERT + CRF	Ma et al. [2020]	96.11		
ERNIEv1 + CRF	Sun et al. [2019]	–	95.95	
ERNIEv1 + LSTM + CRF	Sun et al. [2019]	–	96.25	
MacBERT-base + CRF	Cui et al. [2020]	–	95.80	
MacBERT-base + LSTM + CRF	Cui et al. [2020]	–	96.32	
MacBERT-large + CRF	Cui et al. [2020]	–	95.60	
MacBERT-large + LSTM + CRF	Cui et al. [2020]	–	95.63	

Table 5: Results on Resume NER.

2.2.4 OntoNotes 4

Model	Paper	Reported F1	Our Imp. F1	Notes
LSTM + CRF	Zhang and Yang [2018]	64.30	65.92	num_layers=2
Bichar + LSTM + CRF	Zhang and Yang [2018]	71.89	70.40	num_layers=2
Lattice-LSTM + CRF	Zhang and Yang [2018]	73.88		
FLAT + CRF	Li et al. [2020]	76.45		
SoftLexicon + LSTM + CRF	Ma et al. [2020]	75.64	74.43	num_layers=2; Adamax (lr=1e-3)
BERT + CRF	Ma et al. [2020]	77.93	82.43	
BERT + LSTM + CRF	Ma et al. [2020]	81.82	82.29	
FLAT + BERT + CRF	Li et al. [2020]	81.82		
SoftLexicon + BERT + CRF	Ma et al. [2020]	82.81		
ERNIEv1 + CRF	Sun et al. [2019]	–	81.63	
ERNIEv1 + LSTM + CRF	Sun et al. [2019]	–	82.04	
MacBERT-base + CRF	Cui et al. [2020]	–	82.04	
MacBERT-base + LSTM + CRF	Cui et al. [2020]	–	82.31	

Table 6: Results on OntoNotes 4. Data split follow Che et al. [2013].

2.2.5 OntoNotes 5 (CoNLL 2012)

Model	Paper	Reported F1	Our Imp. F1	Notes
LSTM + CRF	–	–	73.30	num_layers=2
Bichar + LSTM + CRF	–	–	75.36	num_layers=2
Lattice-LSTM + CRF	Jie and Lu [2019]	76.67		
SoftLexicon + LSTM + CRF	Ma et al. [2020]	–	76.13	num_layers=2; Adamax (lr=2e-3)
BERT + CRF	–	–	80.34	
BERT + LSTM + CRF	–	–	80.31	

Table 7: Results on OntoNotes 5.

2.2.6 Yidu S4K (CCKS 2019)

Model	Paper	Reported F1	Our Imp. F1	Notes
LSTM + CRF	–	–	80.43	num_layers=2
Bichar + LSTM + CRF	DeepIE	81.76	81.04	num_layers=2
SoftLexicon + LSTM + CRF	DeepIE	82.76	82.70	num_layers=2; Adamax (lr=2e-3)
BERT + CRF	DeepIE	83.49	82.97	
BERT + LSTM + CRF	–	–	82.94	

Table 8: Results on Yidu S4K.

References

Alan Akbik, Duncan Blythe, and Roland Vollgraf. Contextual string embeddings for sequence labeling. In *Proceedings of the 27th international conference on computational linguistics*, pages 1638–1649, 2018.

Wanxiang Che, Mengqiu Wang, Christopher D Manning, and Ting Liu. Named entity recognition with bilingual constraints. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 52–62, 2013.

Jason PC Chiu and Eric Nichols. Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, 4:357–370, 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.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. Revisiting pre-trained models for Chinese natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 657–668, 2020.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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, 2019.

Markus Eberts and Adrian Ulges. Span-based joint entity and relation extraction with Transformer pre-training. In *Proceedings of the 24th European Conference on Artificial Intelligence*, Santiago de Compostela, Spain, 2020. URL `https://ecai2020.eu/papers/1283_paper.pdf`.

Zhanming Jie and Wei Lu. Dependency-guided LSTM-CRF for named entity recognition. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3862–3872, 2019.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77, 2020.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, 2016.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*, 2019.

Xiaonan Li, Hang Yan, Xipeng Qiu, and Xuan-Jing Huang. FLAT: Chinese NER using flat-lattice transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6836–6842, 2020.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Ruotian Ma, Minlong Peng, Qi Zhang, Zhongyu Wei, and Xuan-Jing Huang. Simplify the usage of lexicon in Chinese NER. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5951–5960, 2020.

Xuezhe Ma and Eduard Hovy. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1064–1074, 2016.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, 2018.

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

Juntao Yu, Bernd Bohnet, and Massimo Poesio. Named entity recognition as dependency parsing. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6470–6476, 2020.

Yue Zhang and Jie Yang. Chinese NER using lattice LSTM. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1554–1564, 2018.