RELATION EXTRACTION

A PREPRINT

Enwei Zhu

November 11, 2021

1 English Tasks

1.1 Settings

- Word embeddings are initialized with GloVe
- From scratch
  - Optimizer: Adadelta (lr=1.0)
  - Batch size: 64
  - Number of epochs: 100
- Fine tuning
  - Optimizer: AdamW (lr=1e-3/2e-3, ft_lr=1e-4)
  - 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

1.2 Results

1.2.1 CoNLL 2004

Model	Paper	Reported F1 (Ent / Rel+)	Our Imp. F1 (Ent / Rel+)	Notes
SpERT (w/ CharLSTM + LSTM)	–	–	86.57 / 66.01	num_layers=2
SpERT (w/ BERT-base)	Eberts and Ulges [2020]	88.94† / 71.47†	88.93 / 70.82 (88.80 / 69.78)	
SpERT (w/ BERT-base + LSTM)	–	–	89.86 / 72.51 (89.89 / 69.68)	
SpERT (w/ RoBERTa-base)	–	–	90.18 / 72.64 (90.30 / 72.18)	
SpERT (w/ RoBERTa-base + LSTM)	–	–	89.17 / 75.03 (90.10 / 73.46)	

Table 1: Results on CoNLL 2004. Pipeline results are reported in parentheses. † means that both training and development splits are used for training (see SpERT repo).

1.2.2 SciERC

Model	Paper	Reported F1 (Ent / Rel)	Our Imp. F1 (Ent / Rel / Rel+)	Notes
SpERT (w/ CharLSTM + LSTM)	–	–	59.63 / 34.25 / 23.04	num_layers=2
SpERT (w/ BERT-base)	Eberts and Ulges [2020]	67.62† / 46.44†	66.71 / 46.07 / 33.94	
SpERT (w/ BERT-base + LSTM)	–	–	67.47 / 45.82 / 33.67	
SpERT (w/ RoBERTa-base)	–	–	69.29 / 48.93 / 36.65	
SpERT (w/ RoBERTa-base + LSTM)	–	–	68.89 / 47.52 / 34.65	

Table 2: Results on SciERC. † means that both training and development splits are used for training (see SpERT repo).

References

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`.