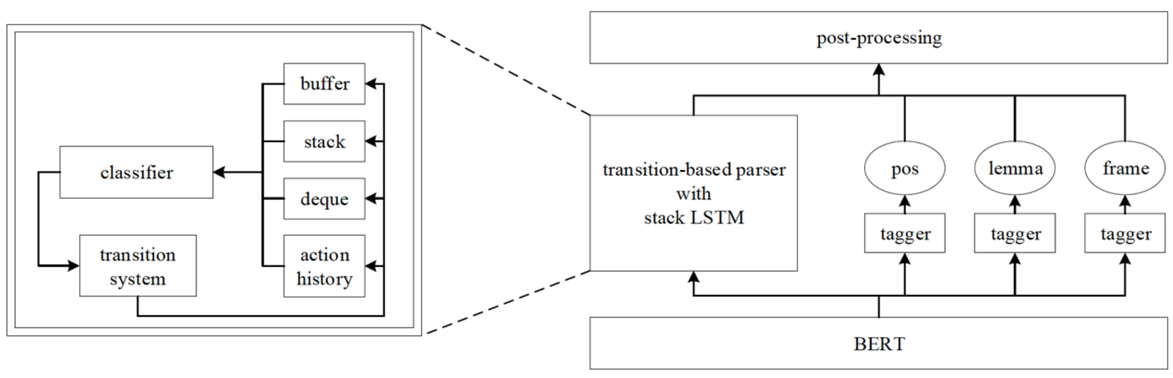
HIT-SCIR at MRP 2019: A Unified Pipeline for Meaning Representation Parsing via Efficient Training and Effective Encoding

Wanxiang Che · Longxu Dou· Yang Xu · Yuxuan Wang · Yijia Liu · Ting Liu
Harbin Institute of Technology

Released Code @ https://github.com/HIT-SCIR/HIT-SCIR-CoNLL2019

Our contribution

- We propose a unified pipeline for MR parsing.
- Efficient Training via Stack LSTM parallel training.
- Effective Encoding through adopting BERT.



Adopting BERT

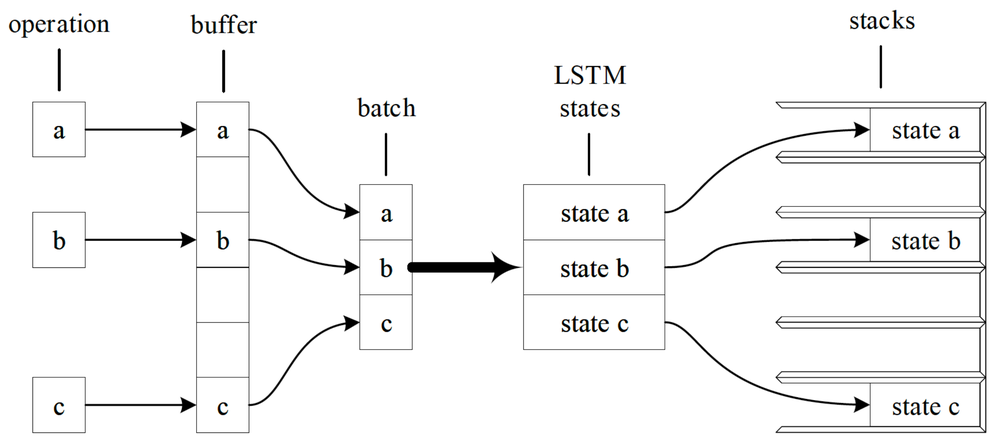
- We fine-tune the BERT with our parser.
- Layer-wise weighed BERT is adopted.

Feature	DM	PSD	EDS	UCCA	AMR	Avg
GloVe	87.1	74.1	88.2	87.5	65.3	**80.4**
BERT(base)	94.3	83.6	91.5	92.8	71.4	**86.7**

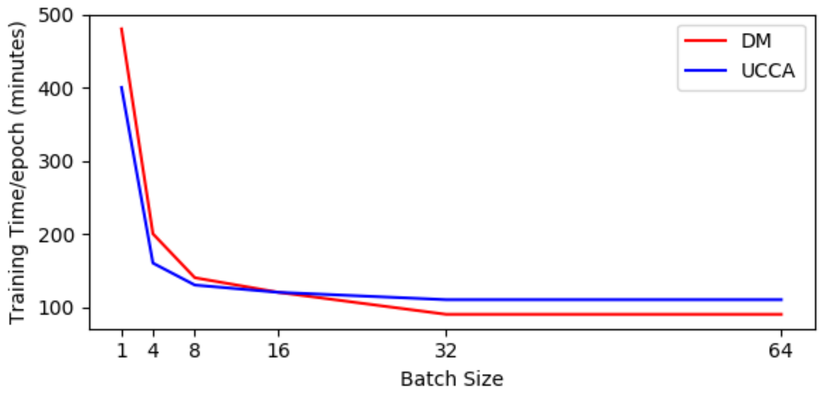
- Metric: ALL-F1 based on mtool.
- Dataset: Splited training data on 8:1:1 proportion.

Stack-LSTM Parallel Training

- Aligning the homogeneous operation in Stack-LSTM within a batch. Then compute them simultaneously.



- The training time within different batch-size.
- Conduct experiments with GloVe.
- 5.3x on DM. 2.7x on UCCA.



Final Evaluation of Shared Task

- Rank 1st according to ALL-F1 and 1st on UCCA.

System	DM	PSD	EDS	UCCA	AMR	ALL-F1
HIT-SCIR	95.08	90.55	90.75	**81.67**	72.94	**86.2**
SJTU-NICT	**95.50**	91.19	89.90	77.80	71.97	85.3
Suda-Alibaba	92.26	85.56	**91.85**	78.43	71.72	84.0
Saarland	94.69	**91.28**	89.10	67.55	66.72	81.9
Hitachi	91.02	91.21	83.74	70.36	43.86	76.0
Amazon	93.26	89.98	-	-	**73.38**	-

Structure vs Representation

Model	Feature	DM		PAS		PSD	
		id	ood	id	ood	id	ood
Wang et al	word2vec	89.3	83.2	91.4	87.2	76.1	73.2
Dozat et al	Glove+char	92.7	87.8	94.0	90.6	80.5	78.6
Transition	GloVe+char	86.1	79.2	89.8	85.2	72.8	68.5
Graph	GloVe+char	91.6	86.1	93.1	89.6	77.4	73.0
Transition	BERT	92.9	89.2	94.4	92.4	**81.6**	**81.0**
Graph	BERT	**94.1**	**90.8**	**94.8**	**92.9**	80.7	79.5

- Dataset: SemEval-2015 Task 18 dataset.
- The gap between Graph and Transition is almost eliminated under BERT.
- Kulmizev et al. found similar conclusion in PTB.
- Transition/Graph model is our implementation of Wang/Dozat.

Wang et al:
<A Neural Transition-Based Approach for Semantic Dependency Graph Parsing>
Dozat et al:
<Simpler but More Accurate Semantic Dependency Parsing>
Kulmizev et al:
<Deep Contextualized Word Embeddings in Transition-Based and Graph-Based Dependency Parsing – A Tale of Two Parsers Revisited>

Ensemble

System	DM	PSD	EDS	UCCA	AMR	Avg
Single	93.98	87.41	89.83	82.61	69.03	**84.57**
Ensemble	94.00	87.79	89.57	83.41	71.35	**85.16**

- Dataset: official-lpps dataset.
- Setting: ensemble model consists of 5 single model.

Conclusions

- BERT is powerful in transition-based parser.
- Stack LSTM parallel training achieves efficient training.