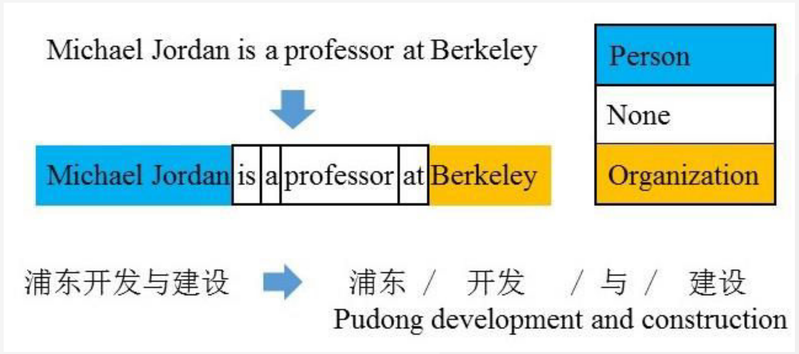
Exploring Segment Representations for Neural Segmentation Models

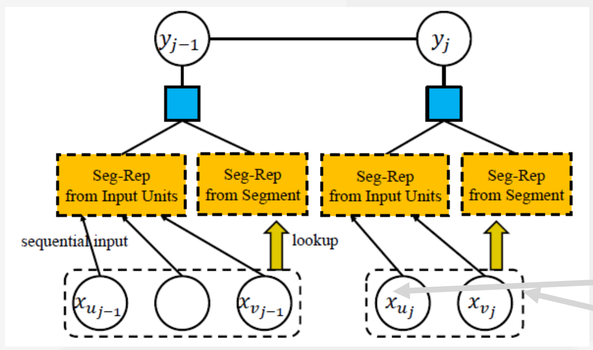
Yijia Liu, Wanxiang Che, Jiang Guo, Bing Qin, and Ting Liu

哈爾濱工業大學

NLP Segmentation Problem



- ***input***: a sequence of elements
- ***segmentation***: a sequence of segment $\mathbf{S} = (s_1, s_2, \dots, s_p)$
- ***segment****:* a tuple $s = (u, v, y)$
  - u: the beginning position
  - v: the ending position
  - y : the label associated with the segment (optional)
- constrained on $v_i + 1 = u_{i+1}$

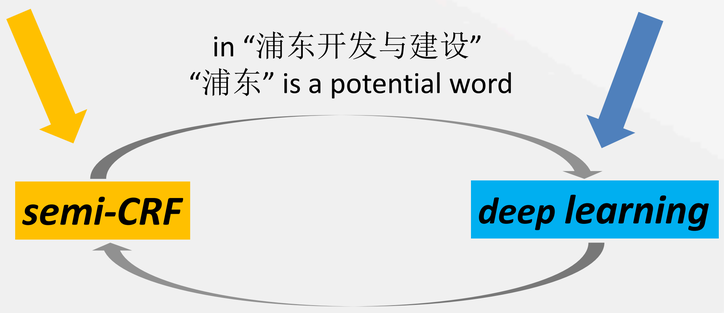


Model Overview

Use word embedding in CWS.

Challenges and Solutions

□access the segment □representing the segment



“浦东”([0.5, 0.3, 0.6, …]) and “虹桥”([0.5, 0.2, 0.5, …]) have similar syntactic/semantic function.

Representing $\phi(s_i, X)$

Old-school $\phi(s_i, X)$ representation

- *crf styled features:*
  - input unit level information: e.g. words, postags
- *semi-crf styled features:*
  - segment-level information: e.g. segment length

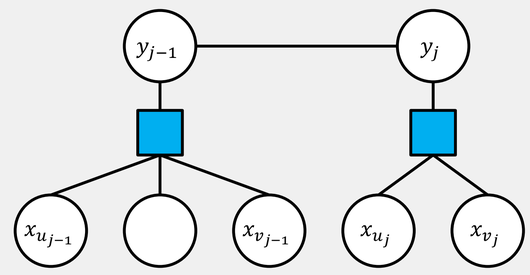
Neuralized $\phi(s_i, X)$ representation

- *crf styled features:*
  - composing input units representation into a vector
- *semi-crf styled features:*
  - embedding the entire segment

Semi-CRF vs. CRF

Modeling Segmentation

- Markov assumption
  - labeling individual inputs with B, I, E, S
  - one input, one state
- Semi-Markov assumption
  - labeling contiguous inputs
  - several inputs, one state

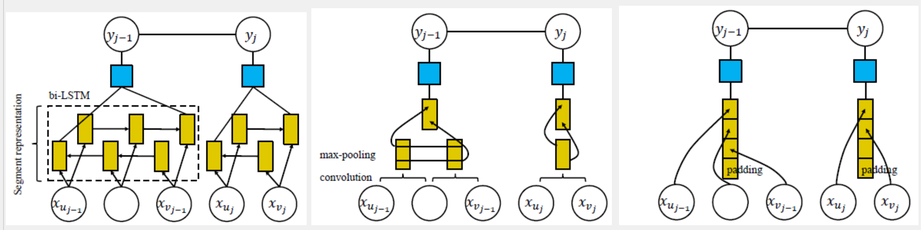


Semi-CRF

- follows semi-Markov assumption
- conditional probability of segmentation S over input x
  - $p(S|X) = \frac{1}{Z} \exp W\Phi(S, X)$
  - $\Phi(S, X)$: decomposed as $\sum_i^p \phi(s_i, X)$

Core Problem
Representing $\phi(s_i, X)$

Seg-Rep from Input Units



SRNN SCNN SCONCATE

Seg-Rep from Segment

Where did the embedding come from?

learning from unlabeled but auto-segmented data

- auto-data: homogeneous or heterogeneous models?
- embedding: tune or fixed?

Experiments

Results of ***Seg-Rep from Input Units*** only: comparable result with baseline

	model	NER CoNLL03 dev	NER CoNLL03 test	CWS CTB6 dev	CWS CTB6 test	CWS PKU dev	CWS PKU test	CWS MSR dev	CWS MSR test	spd
baseline	NN-LABELER	93.03	88.62	93.70	93.06	93.57	92.99	93.22	93.79	**3.30**
	NN-CRF	**93.06**	**89.08**	94.33	93.65	94.09	93.28	93.81	94.17	2.72
	SPARSE-CRF	88.87	83.43	**95.68**	**95.08**	**95.85**	**95.06**	**96.09**	**96.54**	
neural semi-CRF	SRNN	92.97	88.63	94.56	94.06	94.86	93.91	94.38	95.21	0.62
	SCONCATE	92.96	89.07	94.34	93.96	94.41	93.57	94.05	94.53	1.08
	SCNN	91.53	87.68	87.82	87.51	79.64	80.75	85.04	85.79	1.46

Results of ***Seg-Rep from Input Units and Segment***: Segment embedding significant helps!

model	CoNLL03	CTB6	PKU	MSR
NN-LABELER	88.62	93.06	92.99	93.79
NN-CRF	89.08	93.65	93.28	94.17
SPARSE-CRF	83.43	95.08	95.06	96.54
SRNN	88.63	94.06	93.91	95.21
+SEMB-HETERO	89.59	**95.48**	95.60	97.39
	+0.96	*+1.42*	*+1.69*	*+2.18*
SCONCATE	89.07	93.96	93.57	94.53
+SEMB-HETERO	**89.77**	95.42	**95.67**	**97.58**
	+0.70	*+1.43*	*+2.10*	*+3.05*

genre	model	CTB6	PKU	MSR
NN	[Zheng *et al.*, 2013]	-	92.4	93.3
	[Pei *et al.*, 2014]		94.0	94.9
	[Pei *et al.*, 2014] w/bigram	-	95.2	97.2
	[Kong *et al.*, 2015]		90.6	90.7
non-NN	[Tseng, 2005]	-	95.0	96.4
	[Zhang and Clark, 2007]	-	95.1	97.2
	[Sun *et al.*, 2009]	-	95.2	97.3
	[Wang *et al.*, 2011]	**95.7**	-	-
	our best	95.48	**95.67**	**97.58**

Conclusion: we thoroughly study representing the segment in neural semi-CRF. Segment embedding greatly improve the performance