

Multi-labeled Relation Extraction with Attentive Capsule Network

Xinsong Zhang¹, Pengshuai Li¹, Weijia Jia^{2,1*}, and Hai Zhao^{1*}

¹Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China




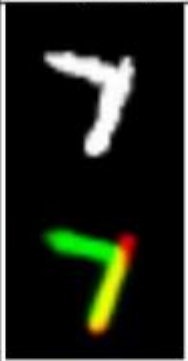








²State Key Lab of IoT for Smart City, University of Macau, Macau 999078, China

{xs Zhang0320, pengshuai.li}@sjtu.edu.cn and {jia-wj, zhaohai}@cs.sjtu.edu.cn

Presenter: Jian Li

Motivation

- Capsule Networks are capable of identify highly **overlapped** digits.
 - Multi-label classification

R:(2, 7) L:(2, 7)	R:(6, 0) L:(6, 0)	R:(6, 8) L:(6, 8)	R:(7, 1) L:(7, 1)	*R:(5, 7) L:(5, 0)	*R:(2, 3) L:(4, 3)
					
R:(8, 7) L:(8, 7)	R:(9, 4) L:(9, 4)	R:(9, 5) L:(9, 5)	R:(8, 4) L:(8, 4)	*R:(0, 8) L:(1, 8)	*R:(1, 6) L:(7, 6)
					

Overview

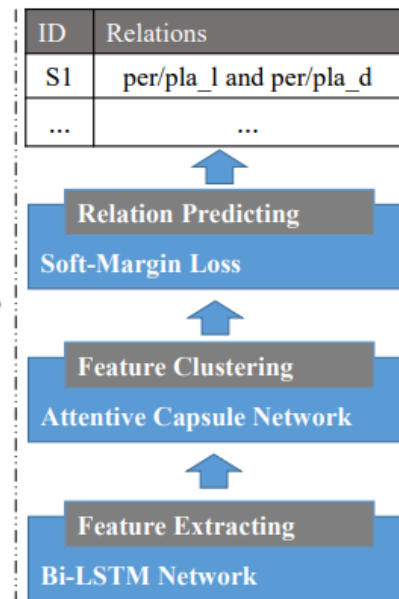
- Multi-labeled relation extraction:

ID	Instances	Relations
S1	<i>[Arthur Lee]</i> , the leader of Love, <i>died</i> on Thursday <i>in [Memphis]</i> .	person/place_death person/place_lived
S2	<i>[Arthur Lee]</i> was <i>born in [Memphis], Tennessee</i> , and <i>lived there</i> until 1952.	person/place_birth person/place_lived
S3	<i>[Abraham Lincoln]</i> was an <i>American statesman and lawyer</i> who served as the 16th <i>President of the [United States]</i> .	person/president_of person/nationality
S4	The politicians want to block a proposal for the <i>[Goshute] Indian reservation</i> to gain income by storing nuclear waste produced <i>in other states</i> , while <i>[Utah] has no</i> nuclear power <i>plants</i> .	location/contain
...

Problems



- 1, Overlapping Relations
- 2, Discrete Features



Our Solution

Figure 1: Problems, challenges and our solution for multi-labeled relation extraction. Words in brackets are entities and the italic red parts are key words that contain relation features. (The relation label in the right table is in abbreviation.)

Architecture

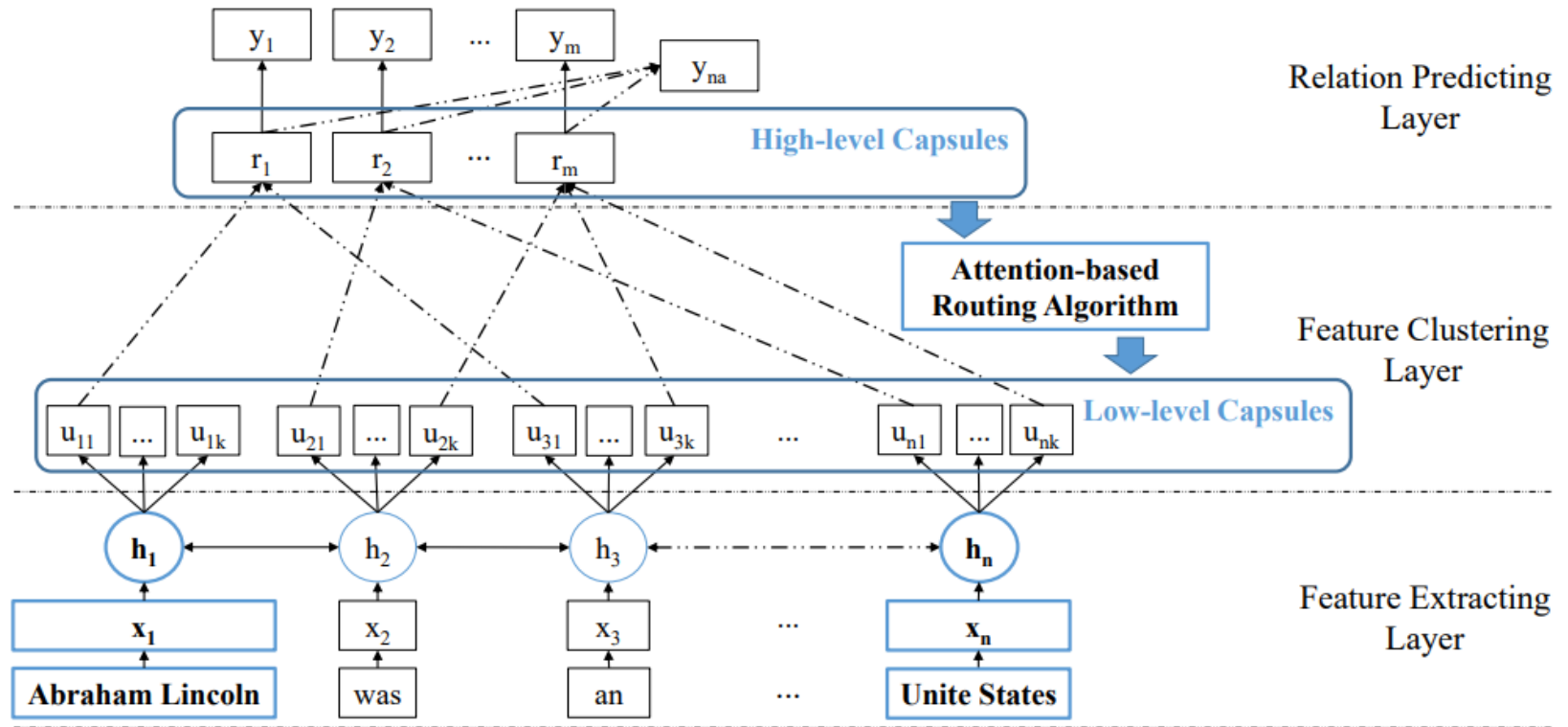


Figure 2: The architecture of our proposed relation extractor, illustrating the procedure for handling one sentence and predicting possible relations between [Abraham Lincoln] and [Unite States]. h is a set of hidden states of Bi-LSTM, u is a low-level capsule set and r represents high-level capsules. y indicates relation labels, and y_{na} expresses “no-relation”. The solid lines are determinate associations, and the dotted lines are possible ones.

Attentive Routing

Algorithm 1 Attention-based Routing Algorithm

Require: low-level capsules u , iterative number z , entity features h_e and hidden states h_t

Ensure: high-level capsules r

```
1: for all capsules  $u_i$  and capsules  $r_j$  do
2:   initialize the logits of coupling coefficients
3:    $b_{ij} = 0$ 
4: end for
5: for  $z$  iterations do
6:    $w_i = \text{softmax}(b_i), \forall u_i \in u$ 
7:    $\alpha_i = \sigma(h_e^T h_t^i), \forall u_i \in u$ 
8:    $r_j = g(\sum_i w_{ij} \alpha_i W_j u_i), \forall r_j \in r$ 
9:    $b_{ij} = b_{ij} + W_j u_i r_j, \forall u_i \in u \text{ and } \forall r_j \in r$ 
10: end for
```

coefficient amendment



$$w_{ij} = \frac{\exp(b_{ij})}{\sum_{j^*} \exp(b_{ij^*})}$$

$$\alpha_i = \sigma(h_e^T h_t^i),$$

similarity with entity tokens

Experiments

Methods	Precision(%)	Recall(%)	F1(%)	PR
Zeng et al. (2014)	28.5	56.3	37.8	0.35
Zhang and Wang (2015)	28.9	57.0	38.4	0.34
Zhou et al. (2016)	26.9	54.9	36.1	0.34
Avg+RNN	25.7	55.1	35.1	0.33
Att-CapNet (CNN-based)	29.9	55.0	38.8	0.36
Att-CapNet (RNN-based)	30.8	63.7	41.6	0.42

Table 3: Performance of all the baselines on NYT-10. PR represents precision-recall curve area.

Methods	Precision(%)	Recall(%)	F1(%)
Max-pooling+CNN	88.4	91.9	90.1
Max-pooling+RNN	89.3	91.8	90.5
Att+RNN	88.8	90.6	89.7
Avg+RNN	86.9	90.5	88.6
Att-CapNet (CNN-based)	87.3	93.0	90.1
Att-CapNet (RNN-based)	89.9	93.7	91.8

Table 6: Performance of all the baselines on selected 500 multi-labeled sentences from NYT-10.

Conclusion

- Apply new methods ([dynamic routing](#)) to existing problems ([relation extraction](#))
 - Adaptive innovation ([attentive routing](#))