Multi-labeled Relation Extraction with Attentive Capsule Network

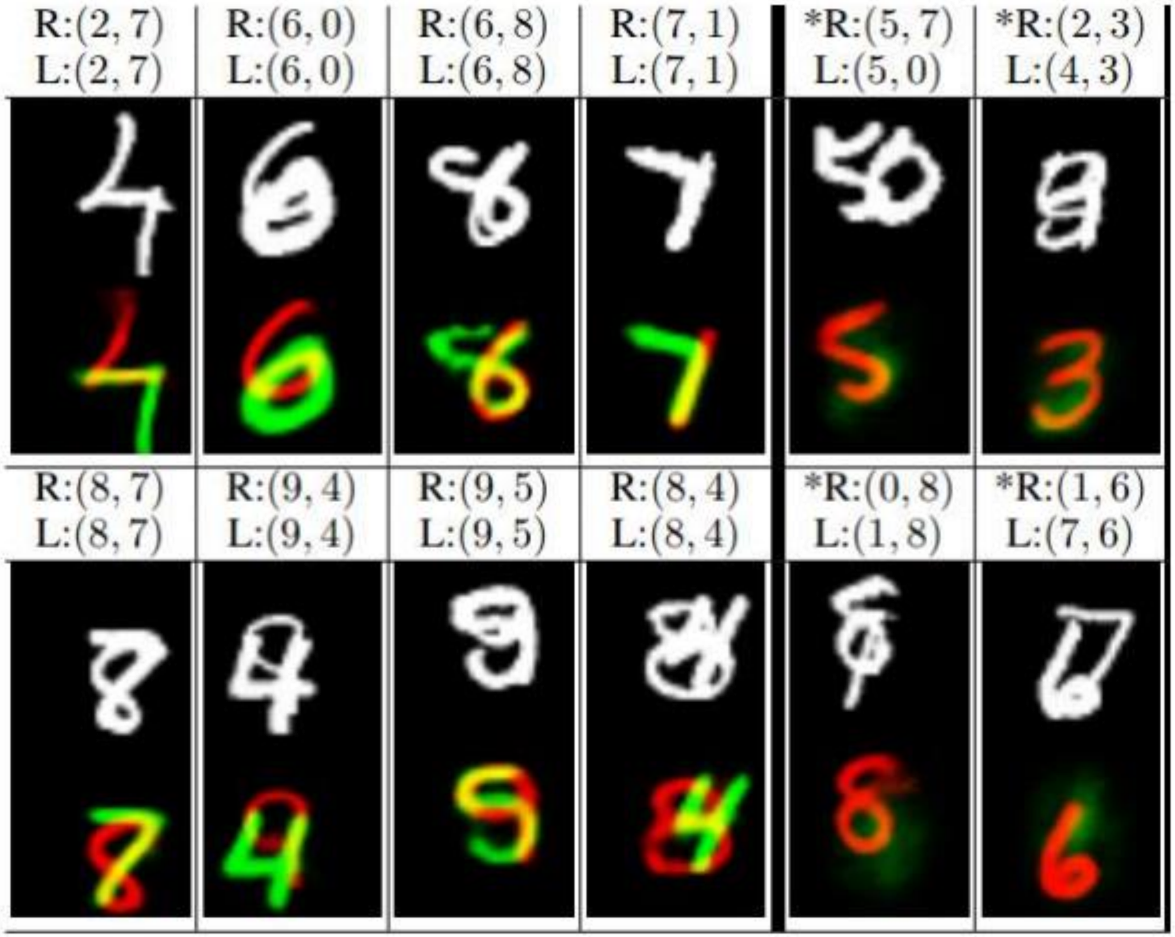
Xinsong Zhang[1], Pengshuai Li[1], Weijia Jia[2,1*], and Hai Zhao[1*]
[1]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]State Key Lab of IoT for Smart City, University of Macau, Macau 999078, China
{xszhang0320, pengshuai.li}@sjtu.edu.cn and {jia-wj, zhaohai}@cs.sjtu.edu.cn

Presenter: Jian LI

*in AAAI 2019

Motivation

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



Overview

- Multi-labeled relation extraction:

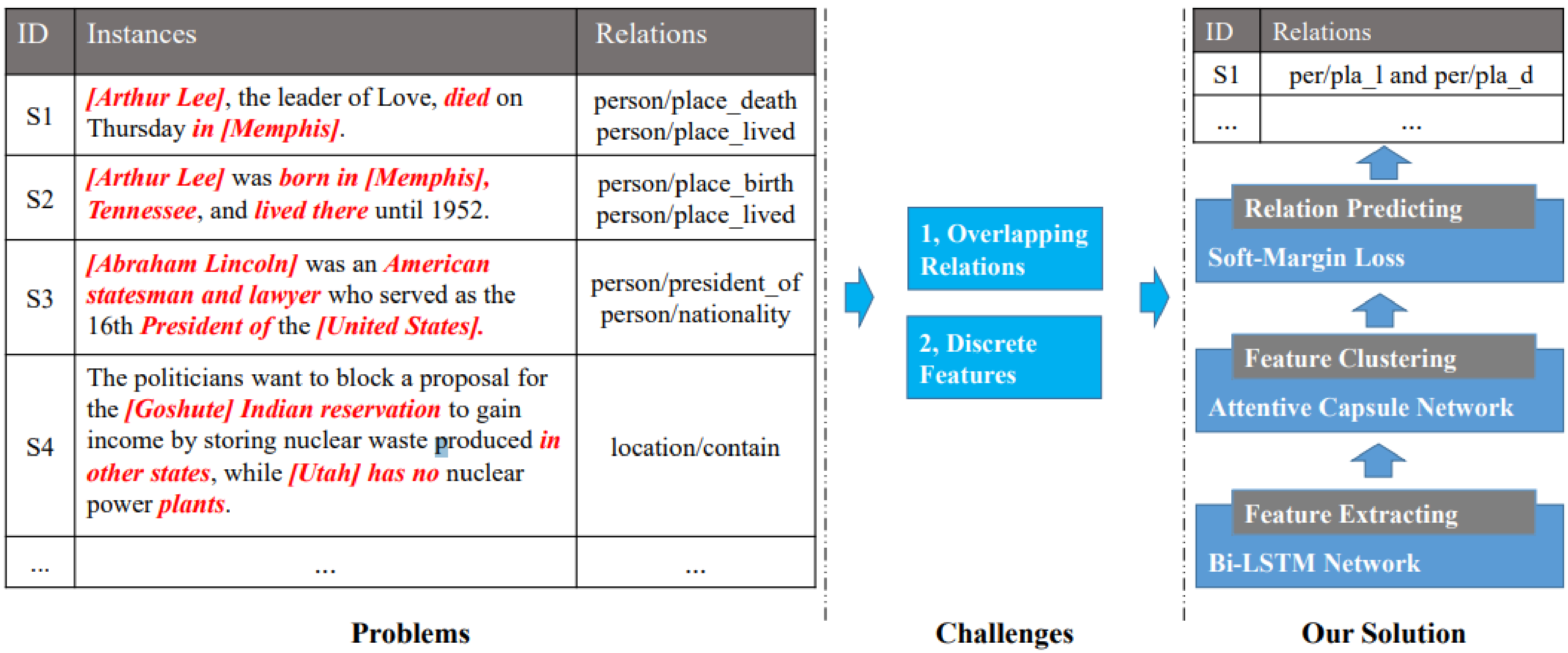


Figure 1: Problems, challenges and our solution for multi-labeled relation extraction. Words in brackets are entities and the italic red parts are key words that contain relation features. (The relation label in the right table is in abbreviation.)

Architecture

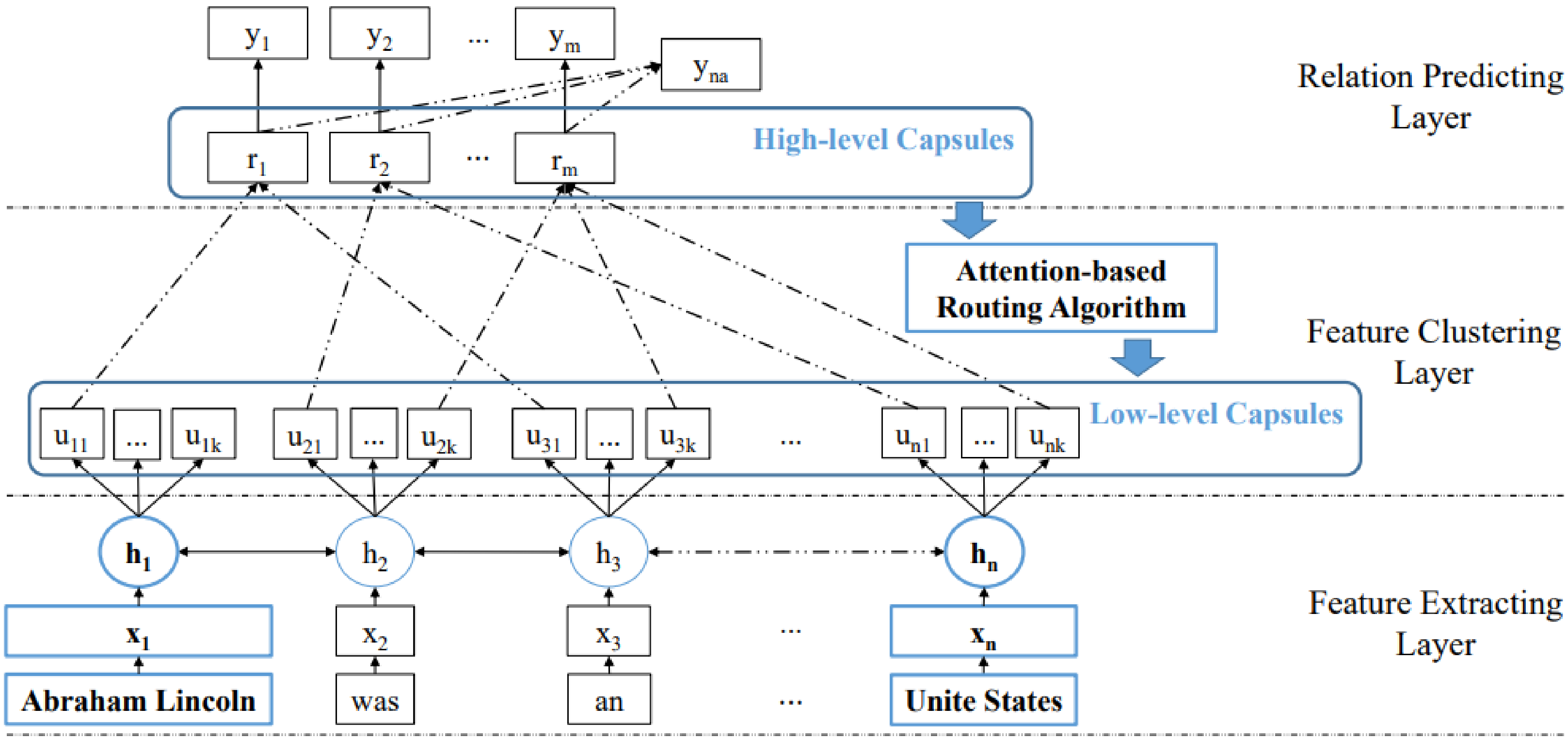


Figure 2: The architecture of our proposed relation extractor, illustrating the procedure for handling one sentence and predicting possible relations between [*Abraham Lincoln*] and [*Unite States*]. h is a set of hidden states of Bi-LSTM, u is a low-level capsule set and r represents high-level capsules. 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

for all capsules u_i and capsules r_j **do**
    initialize the logits of coupling coefficients
    $b_{ij} = 0$
end for
for z iterations **do**
    $w_i = \text{softmax}(b_i), \forall u_i \in u$
    $\alpha_i = \sigma(h_e^T h_t^i), \forall u_i \in u$
    $r_j = g(\sum_i w_{ij} \alpha_i W_j u_i), \forall r_j \in r$
    $b_{ij} = b_{ij} + W_j u_i r_j, \forall u_i \in u$ and $\forall r_j \in r$
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)