Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme

Suncong Zheng, Feng Wang, Hongyun Bao,
Yuexing Hao, Peng Zhou, Bo Xu

Institute of Automation, Chinese Academy of Sciences
Reporter: Suncong Zheng
Email: suncong.zheng@ia.ac.cn, zsctju@126.com
Conference: 2017 ACL

Background

Donald Trump

45th U.S. President

donaldjtrump.com

Donald John Trump is the 45th and current President of the United States. Before entering politics he was a businessman and television personality. Wikipedia

Born: June 14, 1946 (age 70 years), Jamaica Hospital Medical Center

Height: 1.88 m

Spouse: Melania Trump (m. 2005), Marla Maples (m. 1993–1999), Ivana Trump (m. 1977–1992)

Children: Ivanka Trump, Tiffany Trump, Eric Trump, Donald Trump Jr., Barron Trump

Education: Wharton School of the University of Pennsylvania (1968), more

Unstructured text

At age 30 in 1977, Trump married his first wife, Czech model Ivana Zelníčková, at the Marble Collegiate Church

....

Trump was born and raised in Queens, New York City, and earned an economics degree from the Wharton School of the University of Pennsylvania.

Structured Data

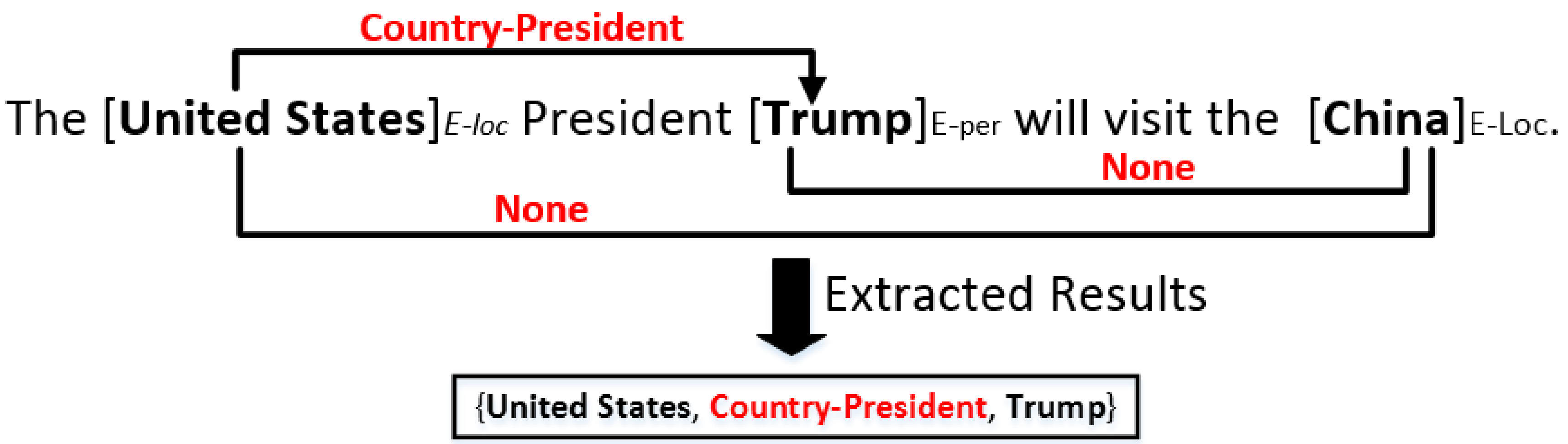
(Entity-1, Relation Type, Entity-2)

(Trump, Spouse, Ivana)…

(Trump, Education, Pennsylvania)…

Problem Description

- Our task is to recognize entity mentions and extract their semantic relations simultaneously from unstructured text. The relation words are extracted from a predefined relation set which may not appear in the given sentence.



A special triplet extraction: subject and object are both entities, predicate is the predefined relation type.

Existing Works

Pipelined method:

Input Sentence

The United States president Trump will visit China.

Named Entity Recognition

Location	Person	Location
United States	Trump	China

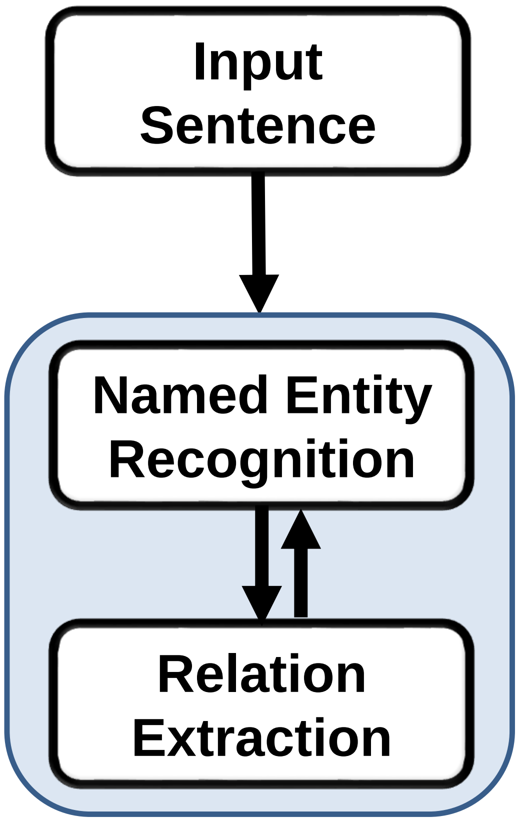
Relation Extraction

{United States, Country-President, Trump }
{China, None, Trump }
{United States, None, China}

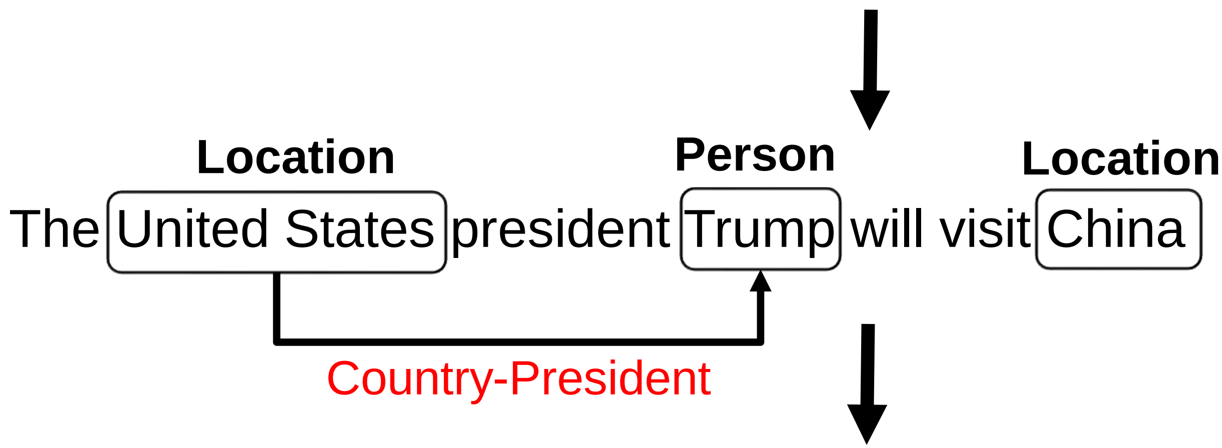
Error Propagation, Produce redundant information

Existing Works

Typical joint extraction method:



The United States president Trump will visit China .



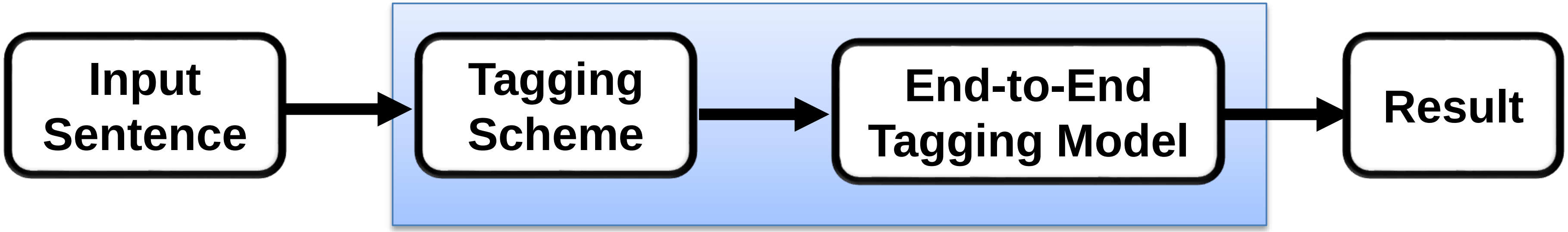
{United States, Country-President, Trump }
{China, None, Trump }
{United States, None, China}

Produce redundant information

Overview of Our Work

- Tagging scheme transforms the extraction problem into a tagging task.

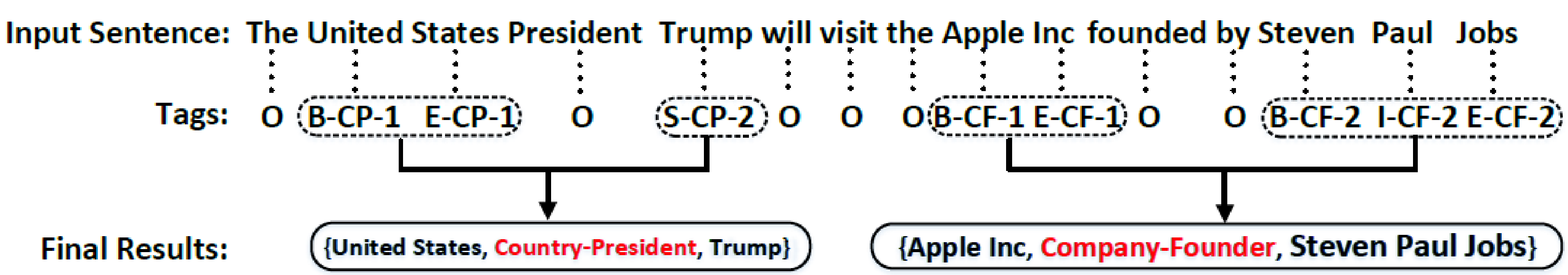
- An end-to-end tagging model is used to extract the results.



The Tagging Scheme

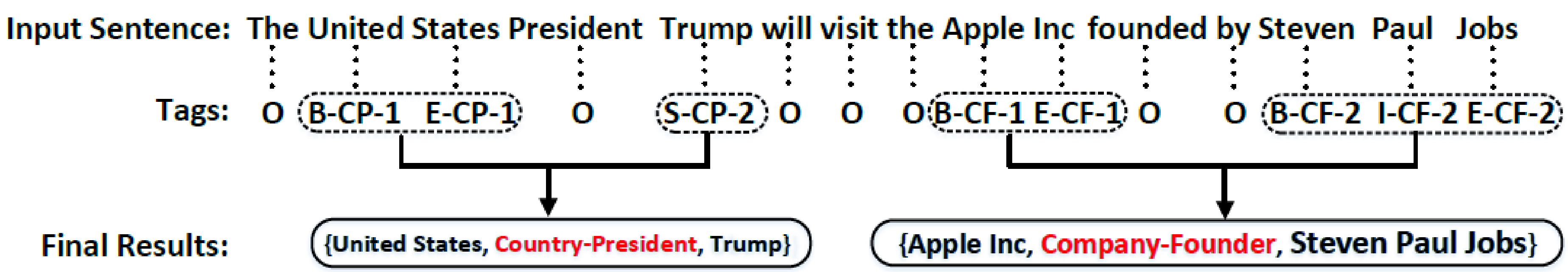
The special tags consist of three parts:

- Word position in the entity { B (begin), I (inside), E (end), S (single) }
- Relation type information { CF, CP, ….}
- Relation role information { 1 (entity 1), 2 (entity 2)}



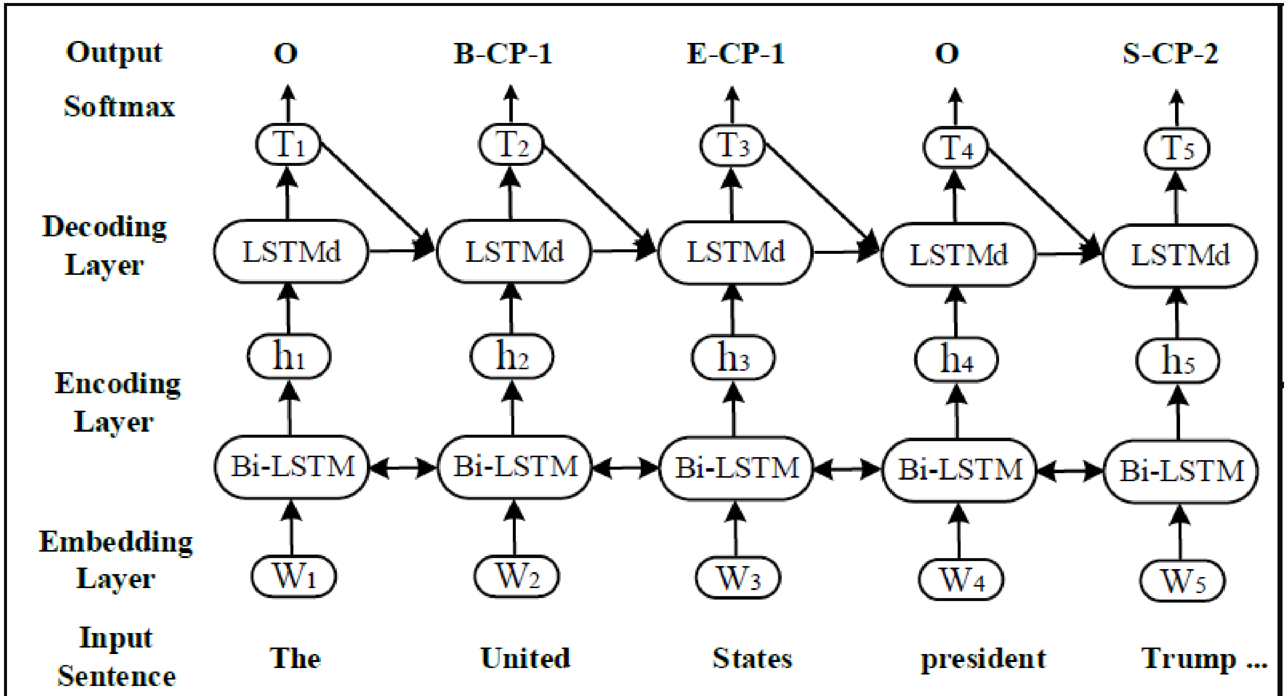
From Tag Sequence To Extracted Results

- Entities with the same relation type are combined into a triplet.
- If a sentence contains two or more triplets with the same relation type, we combine every two entities into a triplet based on the nearest principle.



The End-to-end Model for Tagging

- A Bi-LSTM Encoding Layer.
- A LSTM Decoding Layer.
- A Biased Objective Function.



(a) The End-to-End Model

The End-to-end Model for Tagging

- **The LSTM memory block in Bi-LSTM Encoding Layer.**

$$i_t = \delta(W_{wi}w_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i),$$

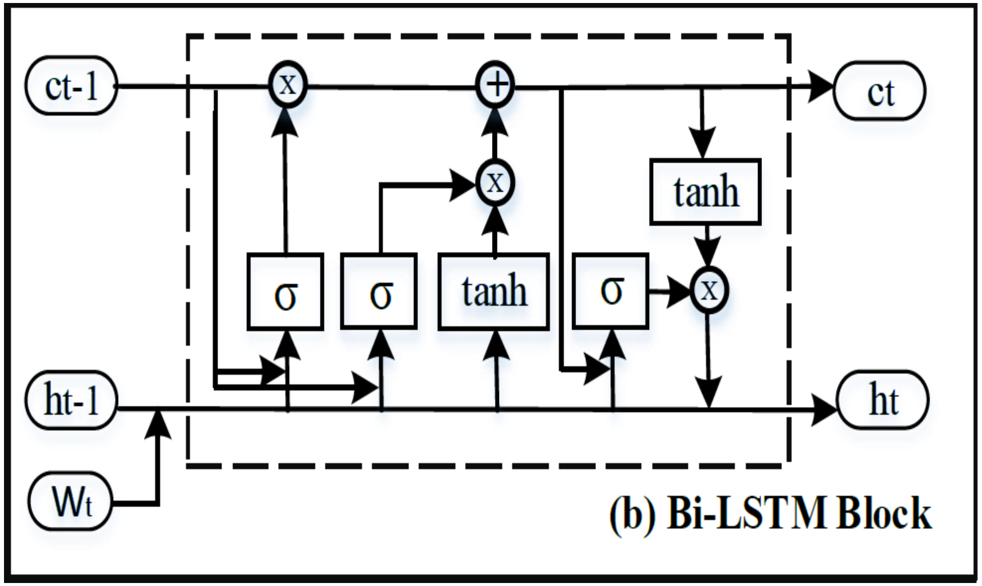
$$f_t = \delta(W_{wf}w_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f),$$

$$z_t = tanh(W_{wc}w_t + W_{hc}h_{t-1} + b_c),$$

$$c_t = f_t c_{t-1} + i_t z_t,$$

$$o_t = \delta(W_{wo}w_t + W_{ho}h_{t-1} + W_{co}c_t + b_o),$$

$$h_t = o_t tanh(c_t),$$



The End-to-end Model for Tagging

- **The LSTM memory block in LSTMd decoding Layer.**

$$i_t^{(2)} = \delta(W_{wi}^{(2)} h_t + W_{hi}^{(2)} h_{t-1}^{(2)} + W_{ti} T_{t-1} + b_i^{(2)}),$$

$$f_t^{(2)} = \delta(W_{wf}^{(2)} h_t + W_{hf}^{(2)} h_{t-1}^{(2)} + W_{tf} T_{t-1} + b_f^{(2)}),$$

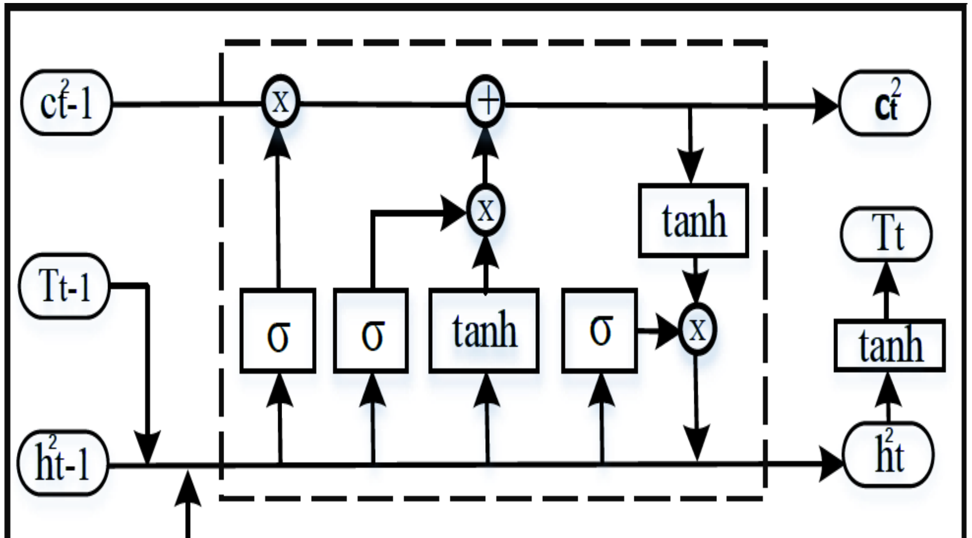
$$z_t^{(2)} = tanh(W_{wc}^{(2)} h_t + W_{hc}^{(2)} h_{t-1}^{(2)} + W_{tc} T_{t-1} + b_c^{(2)}),$$

$$c_t^{(2)} = f_t^{(2)} c_{t-1}^{(2)} + i_t^{(2)} z_t^{(2)},$$

$$o_t^{(2)} = \delta(W_{wo}^{(2)} h_t + W_{ho}^{(2)} h_{t-1}^{(2)} + W_{co}^{(2)} c_t + b_o^{(2)}),$$

$$h_t^{(2)} = o_t^{(2)} tanh(c_t^{(2)}),$$

$$T_t = W_{ts} h_t^{(2)} + b_{ts}.$$



(c) LSTMd Block

The End-to-end Model for Tagging

- **The Biased Objective Function.**

$$L = max \sum_{j=1}^{|\mathbb{D}|} \sum_{t=1}^{L_j} (log(p_t^{(j)} = y_t^{(j)} | x_j, \Theta) \cdot I(O)$$

$$+ \alpha \cdot log(p_t^{(j)} = y_t^{(j)} | x_j, \Theta) \cdot (1 - I(O))),$$

$$I(O) = \begin{cases} 1, \ if \ \ tag = 'O' \\ 0, \ if \ \ tag \neq 'O'. \end{cases}$$

Experimental setting

- **Dataset: The public dataset NYT (**New York Times news**)** *[1]*
  - The training corpus was heuristically labeled using distant supervision method without manually labeling.
  - The test set is manually labeled to ensure its quality.
  - The training data contains 353k triplets, and the test set contains 3,880 triplets. Besides, the size of relation set is 24.

- **Evaluation**
  - Precision/recall and f-measure for triplet (entity1; relation; entity2)
  - Head offsets of two entity mention + relation type

[1] Cotype: Joint extraction of typed entities and relations with knowledge bases (2017 WWW)

Experimental setting

- **Baselines:**
  - **The pipelined methods**: the NER results are obtained by CoType *[1]*
    - DS-logistic (Mintz et al., 2009)
    - LINE (Tang et al., 2015)
    - FCM (Gormley et al., 2015)
  - **The jointly extracting methods:**
    - DS-Joint (Li and Ji, 2014)
    - MultiR (Hoffmann et al., 2011)
    - CoType (Ren et al., 2017)
  - **The end-to-end tagging model**:
    - LSTM-CRF (Lample et al., 2016)
    - LSTM-LSTM (Vaswani et al., 2016)

[1] Cotype: Joint extraction of typed entities and relations with knowledge bases (2017 WWW)

Performance Comparison

Methods	*Prec.*	*Rec.*	*F1*
FCM	0.553	0.154	0.240
DS+logistic	0.258	0.393	0.311
LINE	0.335	0.329	0.332
MultiR	0.338	0.327	0.333
DS-Joint	0.574	0.256	0.354
CoType	0.423	**0.511**	0.463
LSTM-CRF	$\mathbf{0.693} \pm \mathbf{0.008}$	0.310 ± 0.007	0.428 ± 0.008
LSTM-LSTM	0.682 ± 0.007	0.320 ± 0.006	0.436 ± 0.006
LSTM-LSTM-Bias	0.615 ± 0.008	0.414 ± 0.005	$\mathbf{0.495} \pm \mathbf{0.006}$

- The jointly extracting methods are better than pipelined methods, and the tagging methods are better than most of the jointly extracting methods.
- The precisions of the end-to-end models are significantly improved. But only LSTM-LSTM-Bias can be better to balance the precision and recall.

Predicted Results on Triplet's Elements

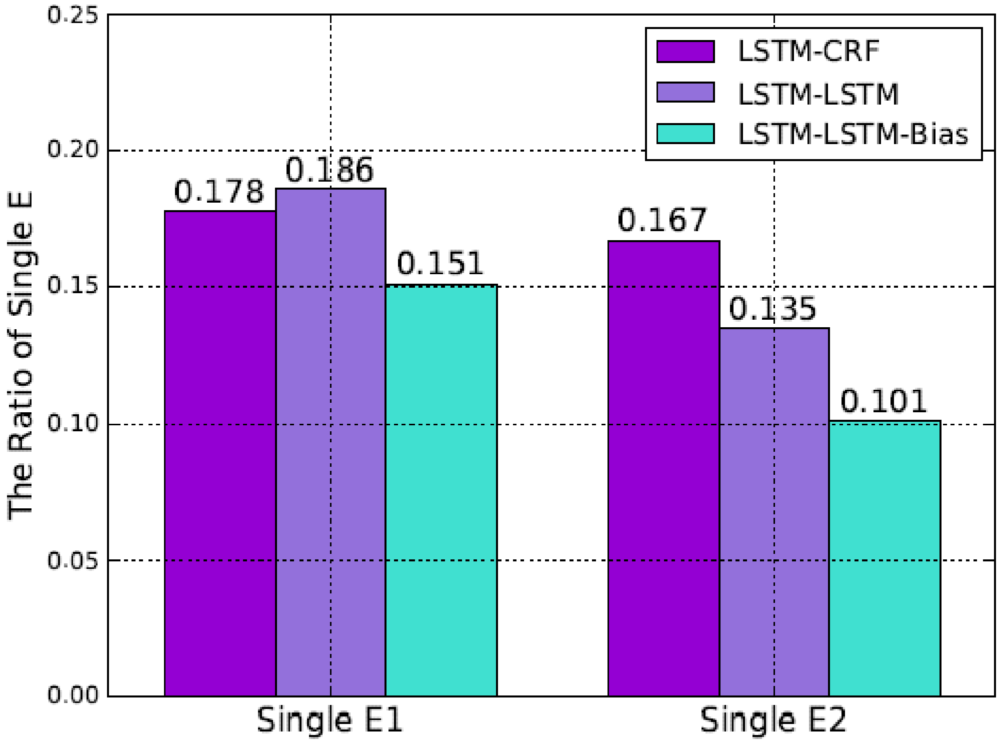
- E1 and E2 represent the performance on predicting each entity, respectively.
- Regardless of relation type, if the head offsets of two corresponding entities are both correct, the instance of (E1, E2) is correct.

Elements	E1			E2			(E1,E2)		
PRF	*Prec.*	*Rec.*	*F1*	*Prec.*	*Rec.*	*F1*	*Prec.*	*Rec.*	*F1*
LSTM-CRF	**0.596**	0.325	0.420	0.605	0.325	0.423	**0.724**	0.341	0.465
LSTM-LSTM	0.593	0.342	0.434	**0.619**	0.334	0.434	0.705	0.340	0.458
LSTM-LSTM-Bias	0.590	**0.479**	**0.529**	0.597	**0.451**	**0.514**	0.645	**0.437**	**0.520**

- Compare (E1,E2) with single E: some of the predicted entities do not form a pair.
- Compare (E1,E2) with Triplet : some of the test data is predicted to be wrong because the relation type is predicted to be wrong.

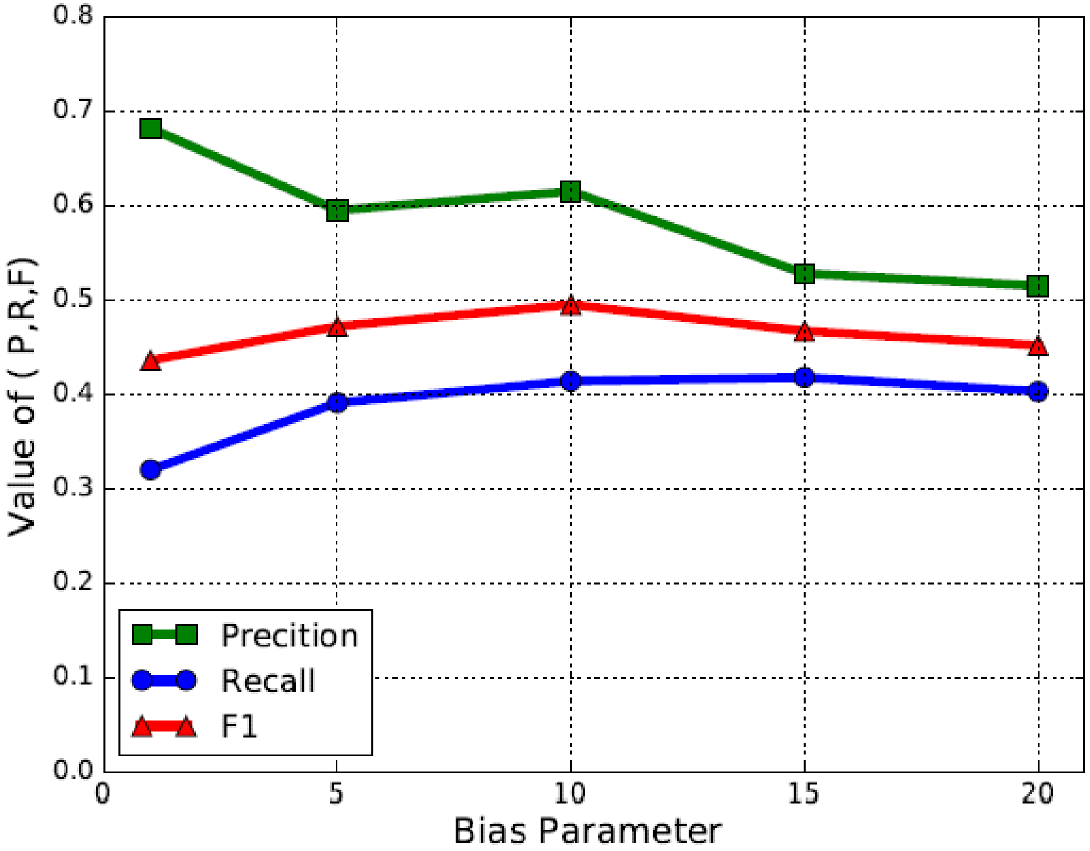
The Ratio of Single Entity

- The single entities refer to those who cannot find their corresponding entities.



- LSTM-LSTM-Bias can effectively associate two entities when compared LSTM-CRF and LSTM-LSTM.

The Effect of Bias Parameter



- If bias parameter is too large, it will affect the precision of prediction and if is too small, the recall will decline.

Case Study --- (1)

- S1 represents the situation that the distance between the two interrelated entities is far away from each other, which is more difficult to detect their relationships.

Standard S1	**[Panama City Beach]**$_{E2contain}$ has condos , but the area was one of only two in **[Florida]**$_{E1contain}$ where sales rose in March , compared with a year earlier.
LSTM-LSTM	Panama City Beach has condos , but the area was one of only two in **[Florida]**$_{E1contain}$ where sales rose in March , compared with a year earlier.
LSTM-LSTM-Bias	**[Panama City Beach]**$_{E2contain}$ has condos , but the area was one of only two in **[Florida]**$_{E1contain}$ where sales rose in March , compared with a year earlier.

Case Study --- (2)

- S2 is a negative example that shows these methods may mistakenly predict one of the entity.

Standard S2	All came from **[Nuremberg]**$_{E2contain}$, **[Germany]**$_{E1contain}$, a center of brass production since the Middle Ages.
LSTM-LSTM	All came from Nuremberg , **[Germany]**$_{E1contain}$, a center of brass production since the **[Middle Ages]**$_{E2contain}$.
LSTM-LSTM-Bias	All came from Nuremberg , **[Germany]**$_{E1contain}$, a center of brass production since the **[Middle Ages]**$_{E2contain}$.

Case Study --- (3)

- S3 is a case that models can predict the entities' head offset right, but the relational role is wrong.

Standard S3	**[Stephen A.]**$_{E2CF}$, the co-founder of the **[Blackstone Group]**$_{E1CF}$, which is in the process of going public , made $ 400 million last year.
LSTM-LSTM	**[Stephen A.]**$_{E1CF}$, the co-founder of the **[Blackstone Group]**$_{E1CF}$, which is in the process of going public , made $ 400 million last year.
LSTM-LSTM-Bias	**[Stephen A.]**$_{E1CF}$, the co-founder of the **[Blackstone Group]**$_{E2CF}$, which is in the process of going public , made $ 400 million last year.

Conclusions & Future Work

- A novel tagging scheme is proposed to jointly extract entities and relations, which can easily transform the extraction problem into a tagging task.
  - A jointly learning method: no error propagation,
  - Modeling triplet directly: no redundant information.
- An end-to-end tagging model is proposed to extract results.
  - Enhance the association between related entities.
- Future work: Based on the tagging scheme, then develop the end-to-multiple-end model to settle the triplet overlapping problem.

Thank you!

Q&A

suncong.zheng@ia.ac.cn , zsctju@126.com