End-to-End Neural Relation Extraction with Global Optimization

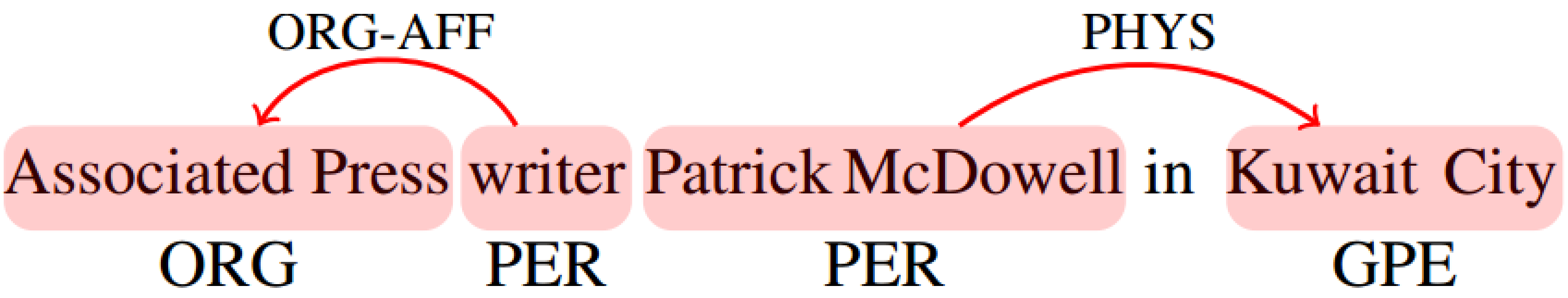
Meishan Zhang, Yue Zhang, Guohong Fu

Heilongjiang University, Harbin, China

& SUTD

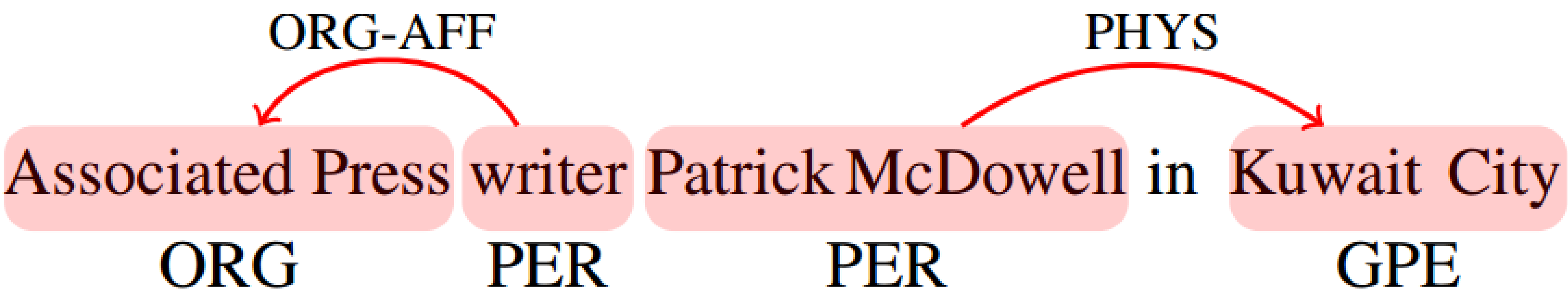
Background

Relation Extraction



Background

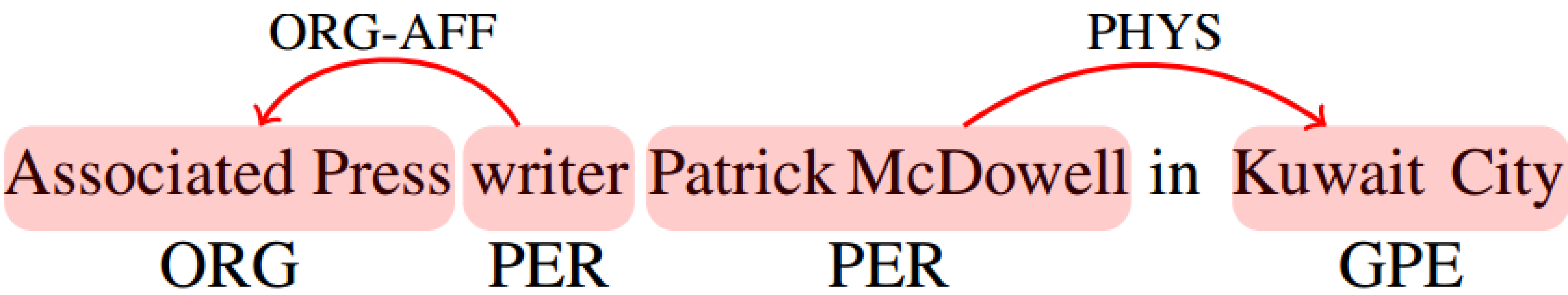
Relation Extraction



- Entity Recognition
- Relation Classification

Background

Relation Extraction



- Entity Recognition
- Relation Classification

Single Model
Joint & End to End

Background

Relation Extraction

Single Model (Joint & End to End)
Approach: Table Filling
Related work:

- **Miwa and Sasaki (2014)**
- **Miwa and Bansal (2016)**

Background

Relation Extraction

- **Table-Filling Sequence**

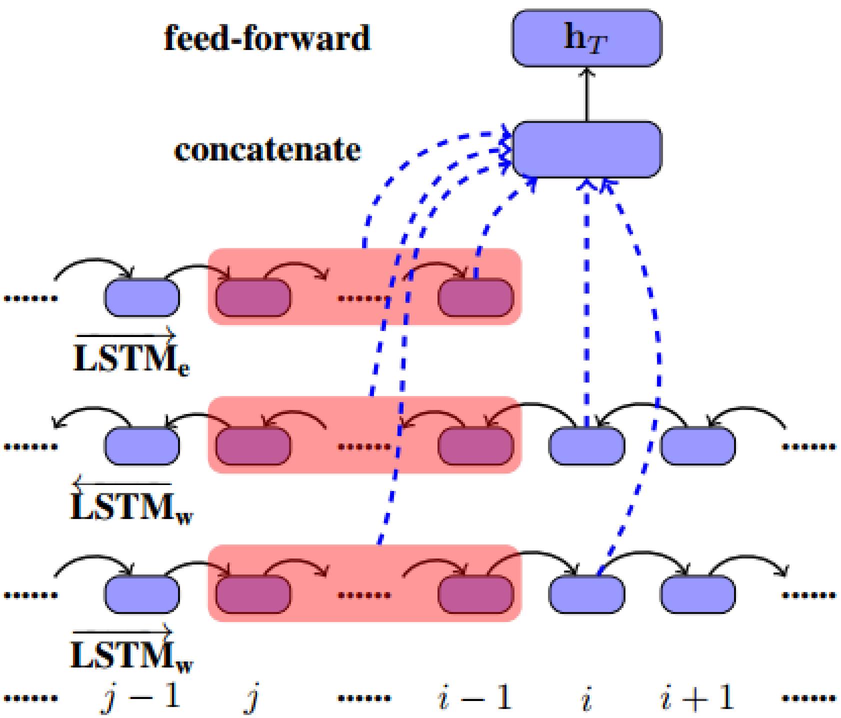
	Associated	Press	writer	Patrick	McDowell	in	Kuwait	City
Associated	1 B-ORG	9 ⊥	16 ⊥	22 ⊥	27 ⊥	31 ⊥	34 ⊥	36 ⊥
Press		2 L-ORG	10 $\overleftarrow{\text{ORG-AFF}}$	17 ⊥	23 ⊥	28 ⊥	32 ⊥	35 ⊥
writer			3 U-PER	11 ⊥	18 ⊥	24 ⊥	29 ⊥	33 ⊥
Patrick				4 B-PER	12 ⊥	19 ⊥	25 ⊥	30 ⊥
McDowell					5 L-PER	13 ⊥	20 ⊥	26 $\overrightarrow{\text{PHYS}}$
in						6 O	14 ⊥	21 ⊥
Kuwait							7 B-GPE	15 ⊥
City								8 L-GPE

Our Contributions

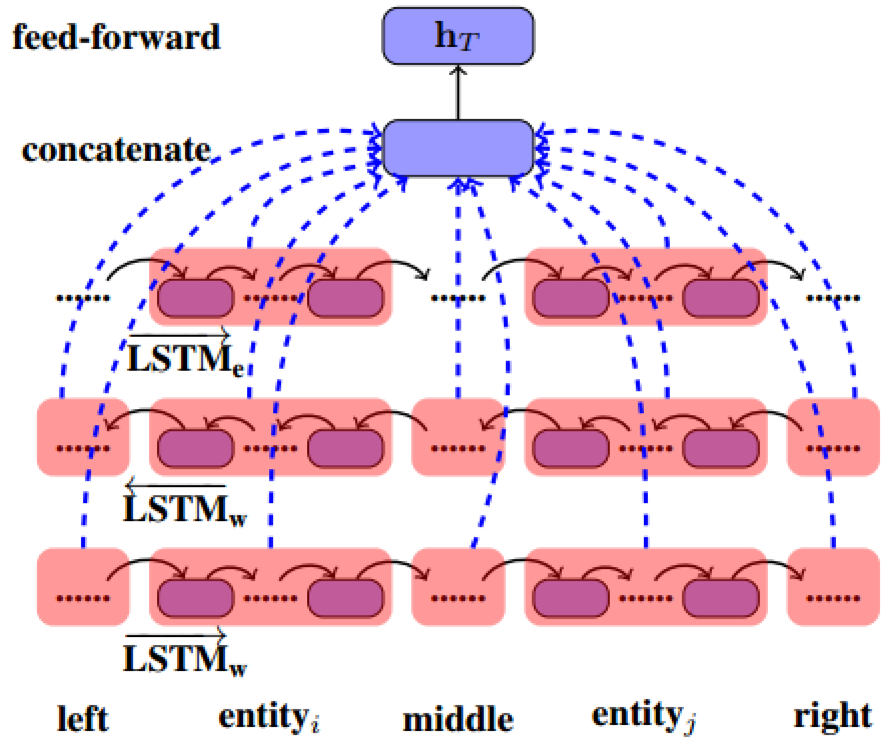
- **Beam Search with Global Learning**

- **Novel Syntactic Features**
  - Without any background on syntactic grammars

Baseline



6 features

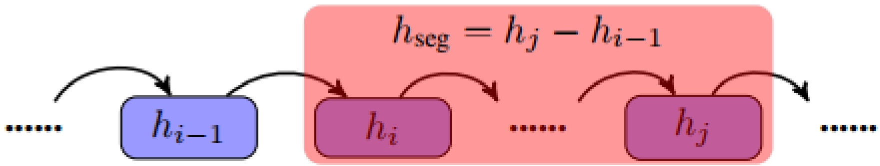


12 features

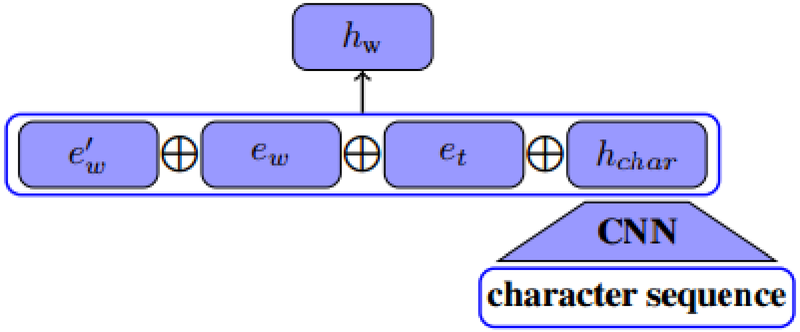
Baseline

Details

segment representation



word representation



Baseline

Classification

- **Greedy Search**

- **Objective**

$$\mathrm{loss}(T, l_i^g, \Theta) = -\log p_{l_i^g}$$

Beam Search

Algorithm 1 Beam-search.

```
agenda ← { (empty table, score=0.0) }
for i in 1 ··· max-step
    next_scored_tables ← { }
    for scored_table in agenda
        labels ← NEXTLABELS(scored_table)
        for next_label in labels
            new ← FILL(scored_table, next_label)
            ADDITEM(next_scored_tables, new)
    agenda ← TOP-B(next_scored_tables, B)
```

Beam Search

Local: classification

$$\text{loss}(T, l_i^g, \Theta) = -\log p_{l_i^g}$$

Global: beam search

$$\text{loss}(x, T_i^g, \Theta) = -\log p_{T_i^g} = -\log \frac{\text{score}(T_i^g)}{\sum_{T_i'} \text{score}(T_i')}$$

$$\text{score}(T_i) = \sum_{j=0}^{i} \text{score}(T_{j-1}, l_j)$$

Beam Search

实验对比(ACE05语料，开发集)

Model	Beam	Relation F1
Local	1	50.9
Local(+SS)	1	51.2
Global	1	51.4
	3	51.8
	5	**52.6**

Syntactic features

Why not dependency path?

- many paths caused dynamic outputting entities
- requiring background on dependency grammar

Syntactic features

Encoder-Decoder Framework

- Encoder：Sentence Representation.
  - usually Bi-LSTMs (multi-layer)

- Decoder: parsing decoding
  - transition-based, graph-based or other

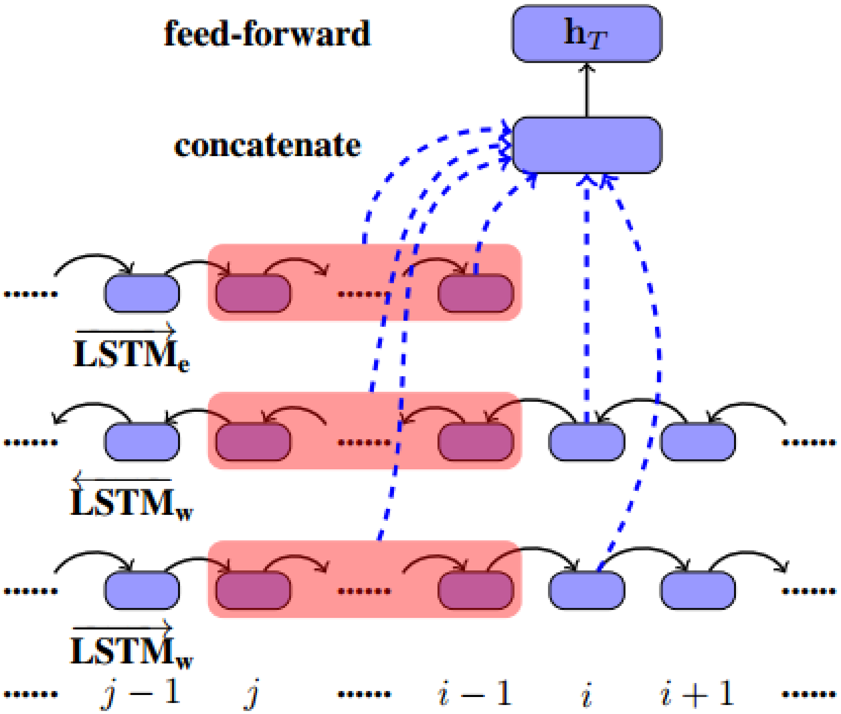
Syntactic features

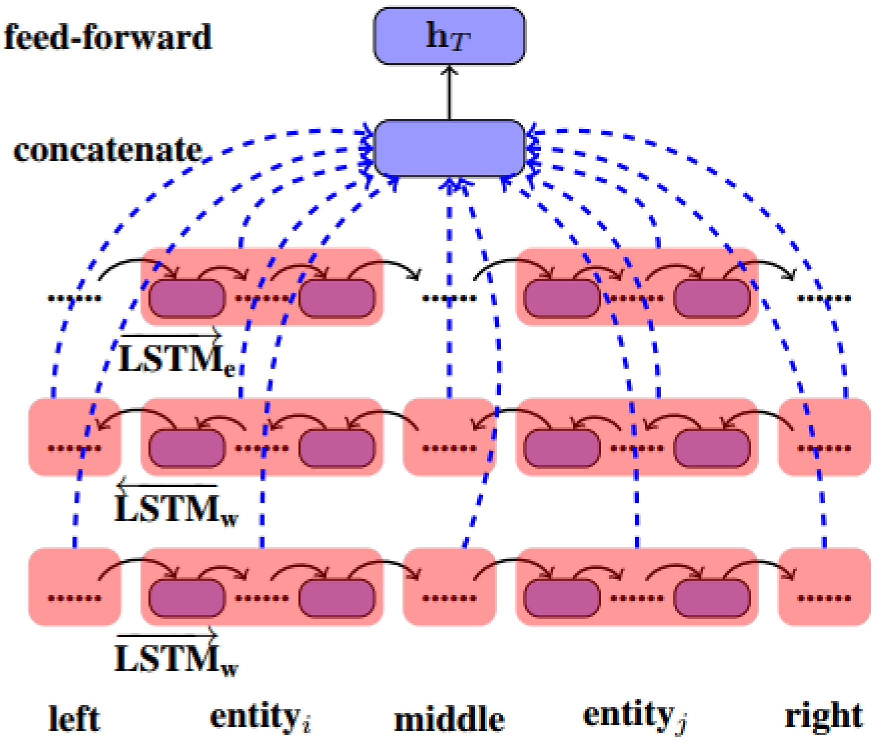
Encoder-Decoder Framework

- **Encoder：Sentence Representation.**
  - **usually Bi-LSTMs (multi-layer)**
- **Decoder: parsing decoding**
  - **transition-based, graph-based or other**

Simply dumping and build lstms based on the output!

Syntactic features





6 features → 10 features

12 features → 22 features

Syntactic features

实验对比(ACE05语料，开发集)

Model	Features	Entity F1	Relation F1
Local	all	**81.6**	**53.0**
	-syn	81.5	50.9
Global	all	**81.9**	**54.2**
	-syn	81.6	52.6

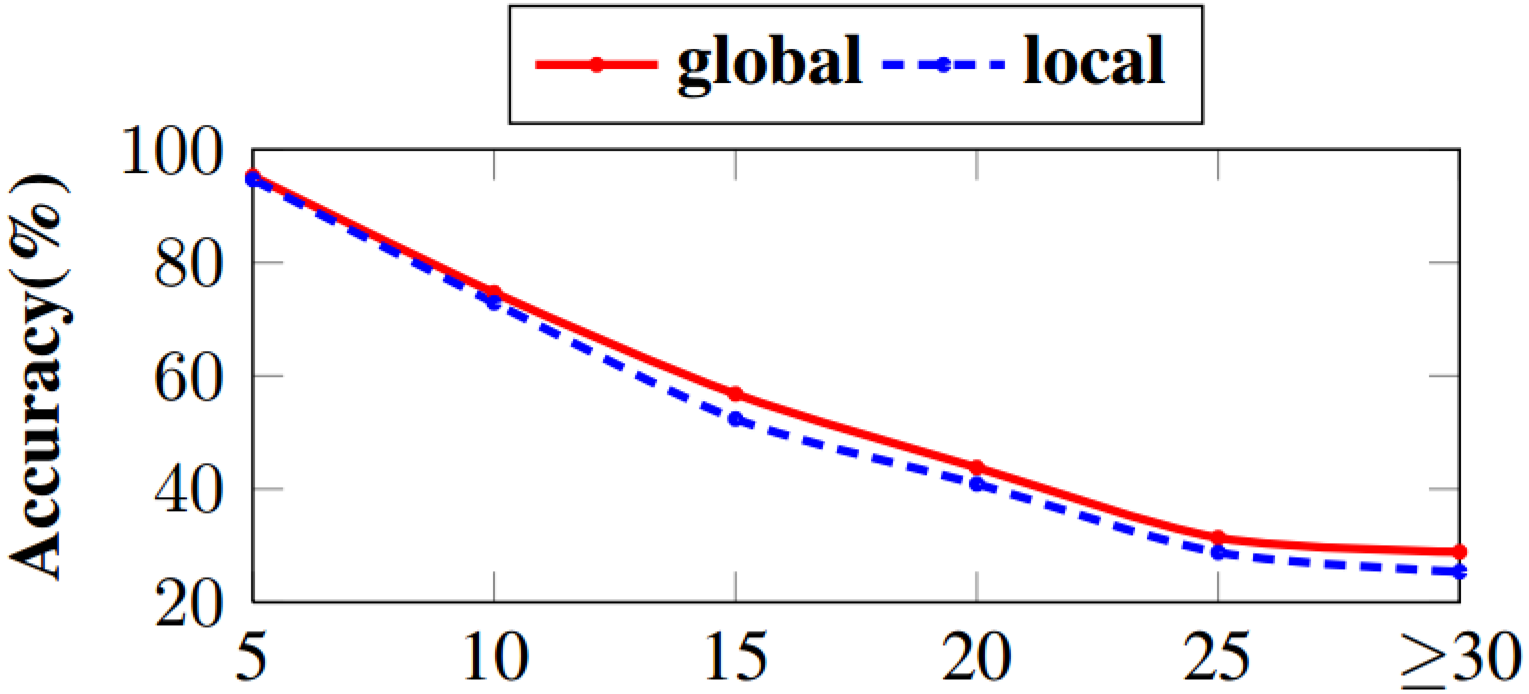
Final Results

实验对比(测试集)

model	ACE05		CONLL04	
	Entity	Relation	Entity	Relation
Our Model	**83.6**	**57.5**	**85.6**	**67.8**
M&B (2016)	83.4	55.6	—	—
L&J (2014)	80.8	49.5	—	—
M&S (2014)	—	—	80.7	61.0

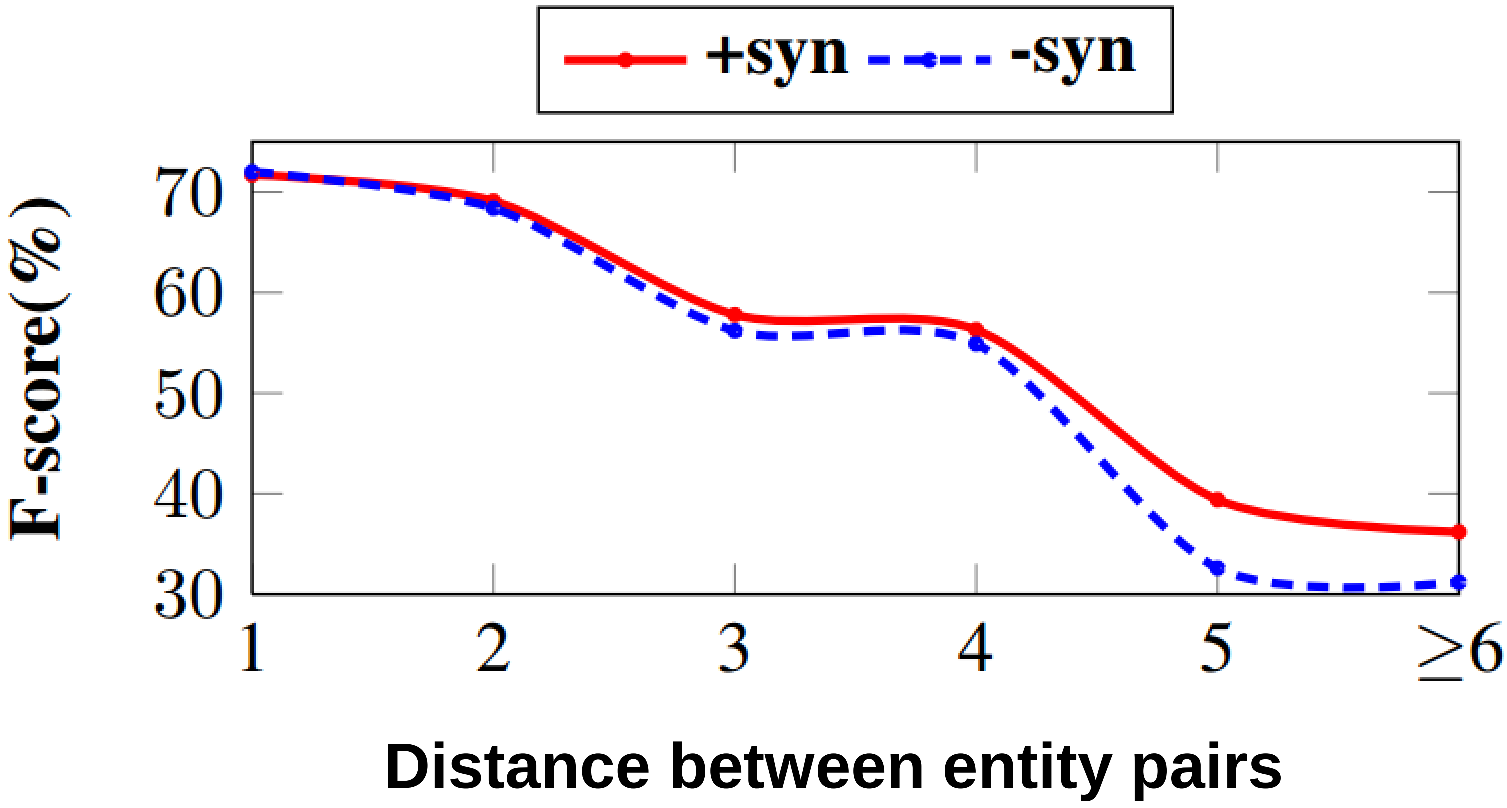
Analysis

Global Learning (整句正确)



Analysis

Syntactic Feature (Relation)



Code

https://github.com/zhangmeishan/NNRelationExtraction

N3LDG

https://github.com/zhangmeishan/N3LDG

- direct goal (seq-to-seq with beam search)
- other models (much easier, e.g., biaffine parser)
- experimental, research goal
- cpu&gpu supporting (gpu part is implemented with the help of zhenghua Li)

Q/A?

Thanks!