Multi-Grained Named Entity Recognition

Congying Xia[1,5] Chenwei Zhang[1] Tao Yang[2] Yaliang Li[3] Nan Du[2] Xian Wu[2] Wei Fan[2] Fenglong Ma[4] Philip S. Yu[1,5]

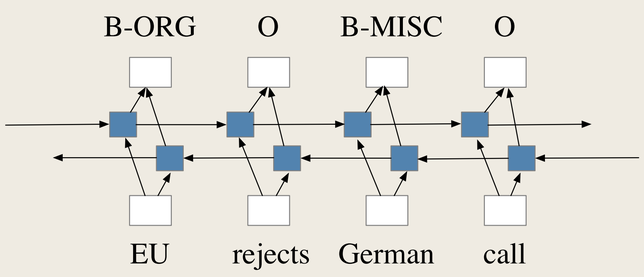
[1]University of Illinois at Chicago [2]Tencent Medical AI Lab [3]Alibaba Group [4]University of Baffulo [5]Zhijiang Lab

BACKGROUND

- ***Named Entity Recognition*** is a fundamental task in NLP
  - Input: an utterance
  - Output: identified entities, such as person names, locations, and organizations
  - Applications: syntactic parsing, question answering, relation extraction
- ***Non-overlapping Named Entity Recognition***

Beijing, previously known as Peking, is the capital city of China.
GPE GPE GPE

  - Sequence Tagging models (LSTM/CNN + CRF)



  - Fail to detect nested named entities which are embedded in longer entity mentions

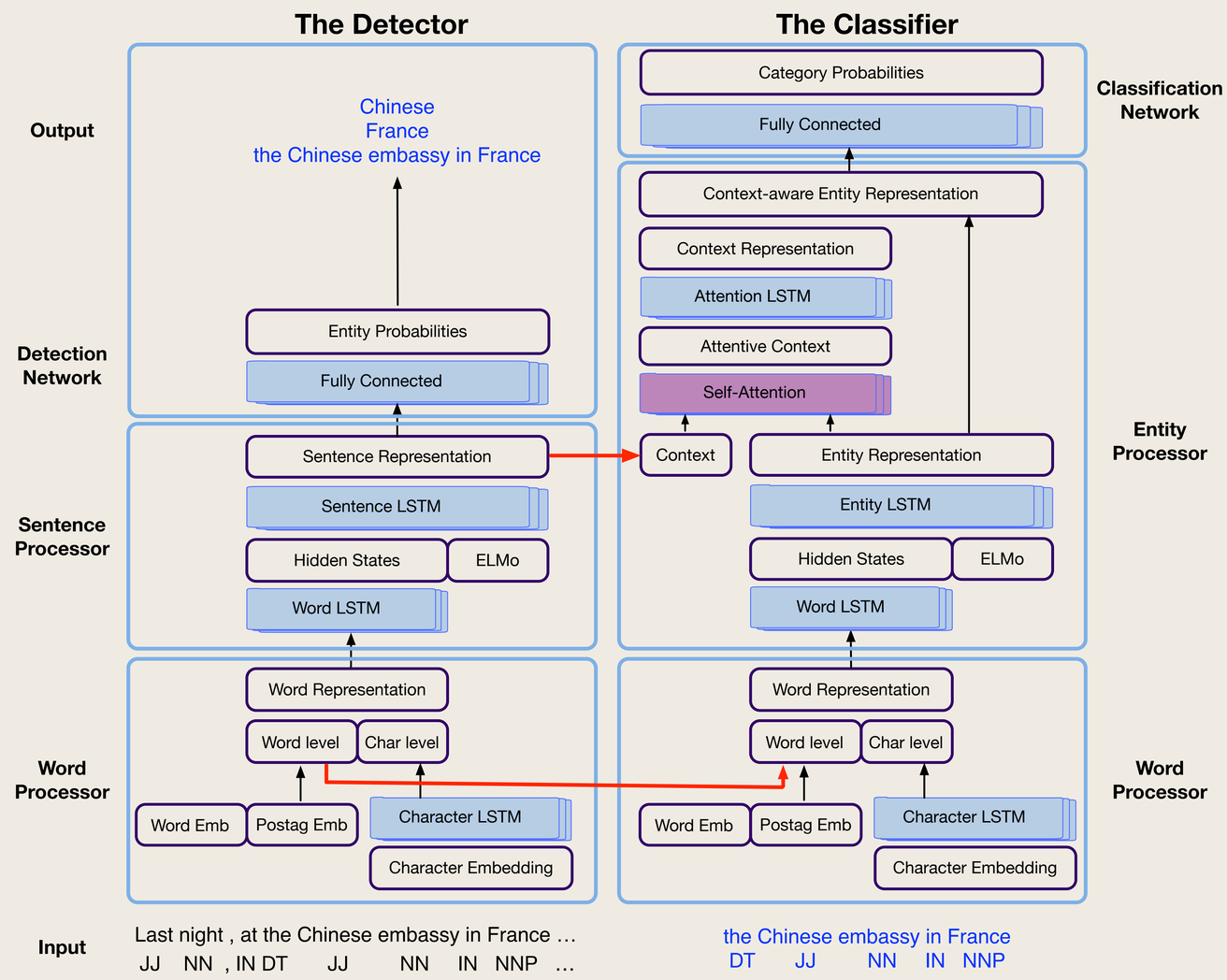
- ***Nested Named Entity Recognition***

Last night , at the Chinese embassy in France, there was a holiday atmosphere.
GPE GPE
Facility Entity

  - Nested NER models are designed explicitly for recognizing nested named entities
  - Nested NER models usually do not perform well on non-overlapping entities

MGNER

- ***Detector***: detect possible entities in various granularities
  - *Word Processor*: word embedding, postage embedding, character level embedding
  - *Sentence Processor*: sentence LSTM + Elmo embedding
  - *Detection Network*: generate all possible word segments and estimate the probability of each proposal as being an entity or not
- ***Classifier***: detect entity positions in various granularities
  - *Entity Processor*: context-aware entity representation with self-attention
  - *Classification Network*: classify candidates into pre-defined categories



RESULTS

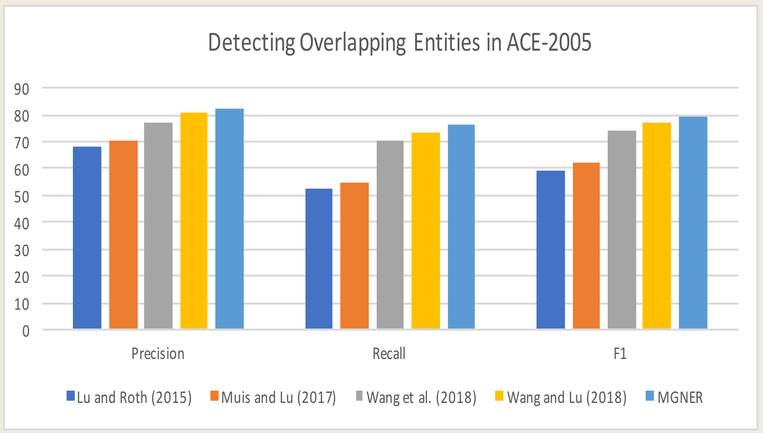
- Nested NER task
  - Datasets: ACE-2004, ACE-2005

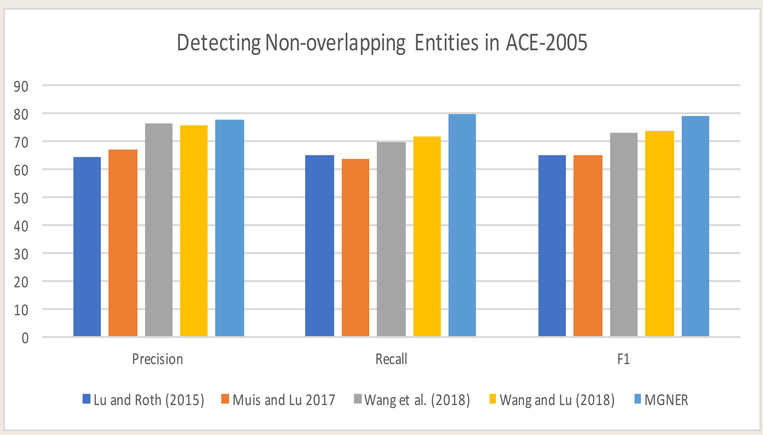
MODEL	ACE-2004			ACE-2005		
	P	R	F1	P	R	F1
Lu and Roth (2015)	70.0	56.9	62.8	66.3	59.2	62.5
Lample et al. (2016)	71.3	50.5	58.3	64.1	52.4	57.6
Muis and Lu (2017)	72.7	58.0	64.5	69.1	58.1	63.1
Xu et al. (2017)	68.2	54.3	60.5	67.4	55.1	60.6
Katiyar and Cardie (2018)	73.6	71.8	72.7	70.6	70.4	70.5
Ju et al. (2018)	-	-	-	74.2	70.3	72.2
Wang et al. (2018)	74.9	71.8	73.3	74.5	71.5	73.0
Wang and Lu (2018)	78.0	72.4	75.1	76.8	72.3	74.5
MGNER w/o context	79.8	76.3	78.0	**79.6**	75.6	77.5
MGNER w/o attention	81.5	76.5	78.9	79.4	76.0	77.7
MGNER	**81.7**	**77.4**	**79.5**	79.0	**77.3**	**78.2**

- Non-overlapping NER task
  - Dataset: CoNLL-2003

MODEL	CoNLL-2003	
	DEV	TEST
Lu and Roth (2015)	89.2	83.8
Muis and Lu (2017)	-	84.3
Xu et al. (2017)	-	90.85
Wang and Lu (2018)	-	90.2
Lample et al. (2016)	-	90.94
Ma and Hovy (2016)	94.74	91.21
Chiu and Nichols (2016)	94.03 ± 0.23	91.62 ± 0.33
Peters et al. (2017)	-	91.93 ± 0.19
Peters et al. (2018)	-	92.22 ± 0.10
MGNER w/o context	95.21 ± 0.12	92.23 ± 0.06
MGNER w/o attention	95.23 ± 0.06	92.26 ± 0.09
MGNER	$\mathbf{95.24 \pm 0.13}$	$\mathbf{92.28 \pm 0.12}$

- Performance on different types of sentences
  - Split the test data in ACE-2005 into two portions: sentences with/without overlapping entities





CONCLUSIONS

- MGNER: state-of-the-art performance on both Nested NER and Non-overlapping NER
- High modularity and each component in MGNER can adopt a wide range of neural networks

Code&Data
GitHub