ACL2020

Named Entity Recognition as Dependency Parsing

Juntao Yu
Queen Mary University
London, UK
juntao.yu@qmul.ac.uk

Bernd Bohnet
Google Research
Netherlands
bohnetbd@google.com

Massimo Poesio
Queen Mary University
London, UK
m.poesio@qmul.ac.uk

Main Work

- In this paper, we use ideas from graph-based dependency parsing to provide our model a global view on the input via a biaffine model
- The biaffine model scores pairs of start and end tokens in a sentence which we use to explore all spans， so that the model is able to predict named entities accurately
- We show that the model works well for both nested and flat NER through evaluation on 8 corpora and achieving SoTA performance on all of them, with accuracy gains of up to 2.2 percentage points.

CASE

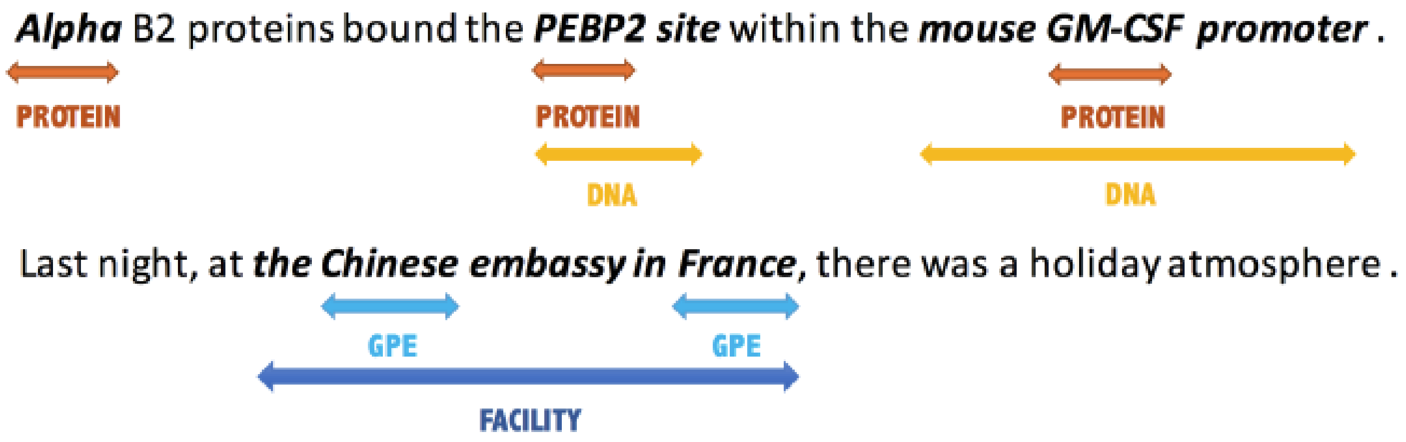
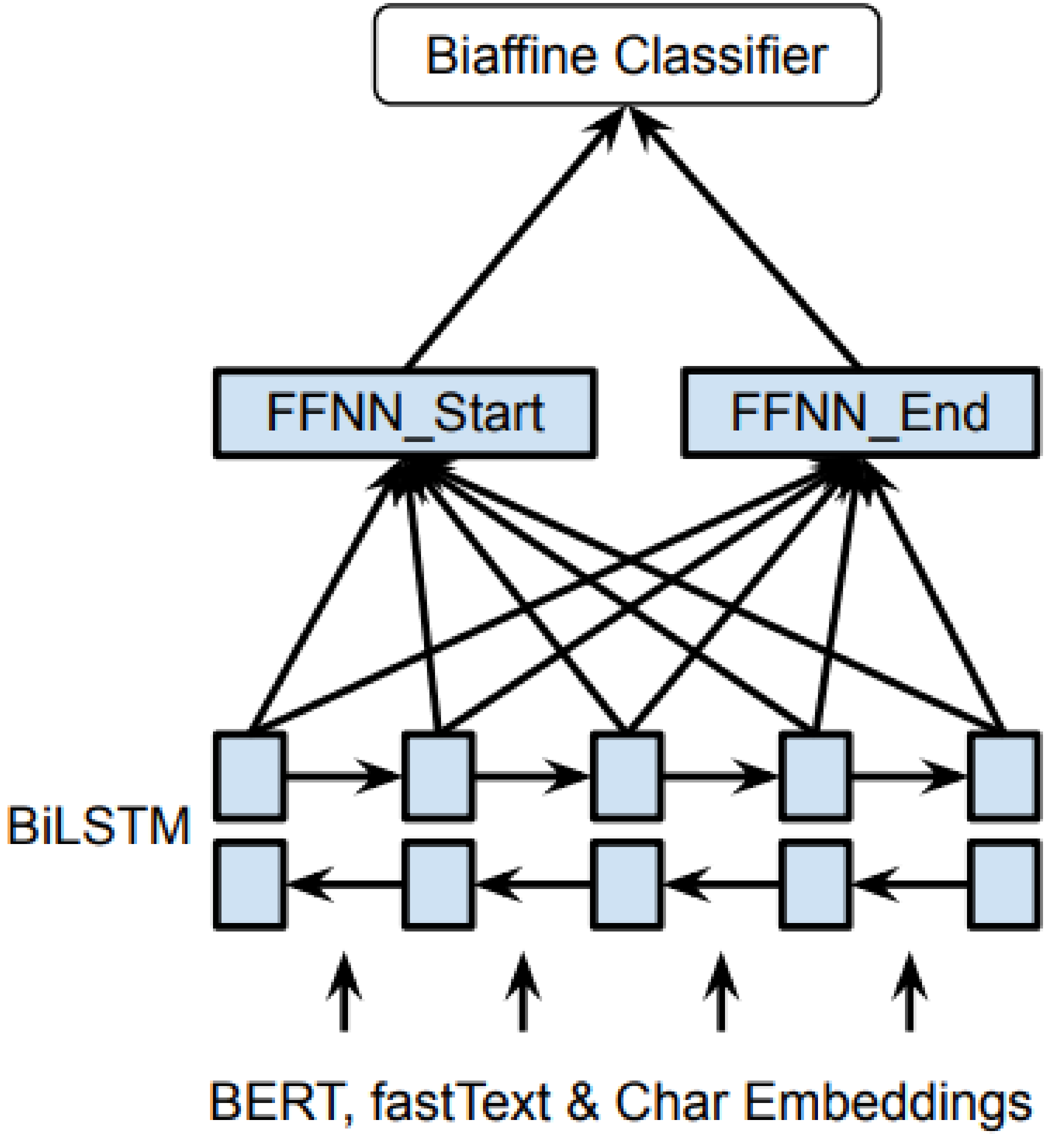


Figure 1: Examples for *nested* entities from GENIA and ACE04 corpora.

Model


Biaffine Classifier
FFNN_Start
FFNN_End
BiLSTM
BERT, fastText & Char Embeddings

Biaffine Classifier

$$
\begin{aligned}
h_s(i) &= \mathrm{FFNN}_s(x_{s_i}) \\
h_e(i) &= \mathrm{FFNN}_e(x_{e_i}) \\
r_m(i) &= h_s(i)^\top \mathrm{U}_m h_e(i) \\
&\quad + W_m(h_s(i) \oplus h_e(i)) + b_m
\end{aligned}
$$

where s_i and e_i are the start and end indices of the span i, U_m is a $d \times c \times d$ tensor, W_m is a $2d \times c$ matrix and b_m is the bias.

STRATEGY

- For nested NER, a entity is selected as long as it does **not clash** the boundaries of higher ranked entities.
- We denote a entity i to **clash** boundaries with another entity j if $si < sj \leqslant ei < ej$ or $sj < si \leqslant ej < ei$

  (the Bank of China: *the Bank of* clashs *Bank of China*)
- For flat NER, we apply one more constraint, in which any entity containing or is inside an entity ranked before it will not be selected

Result

Model	P	R	F1
ACE 2004			
Katiyar and Cardie (2018)	73.6	71.8	72.7
Wang et al. (2018)	-	-	73.3
Wang and Lu (2018)	78.0	72.4	75.1
Straková et al. (2019)	-	-	84.4
Luan et al. (2019)	-	-	84.7
Our model	87.3	86.0	**86.7**
ACE 2005			
Katiyar and Cardie (2018)	70.6	70.4	70.5
Wang et al. (2018)	-	-	73.0
Wang and Lu (2018)	76.8	72.3	74.5
Lin et al. (2019)	76.2	73.6	74.9
Fisher and Vlachos (2019)	82.7	82.1	82.4
Luan et al. (2019)	-	-	82.9
Straková et al. (2019)	-	-	84.3
Our model	85.2	85.6	**85.4**
GENIA			
Katiyar and Cardie (2018)	79.8	68.2	73.6
Wang et al. (2018)	-	-	73.9
Ju et al. (2018)	78.5	71.3	74.7
Wang and Lu (2018)	77.0	73.3	75.1
Sohrab and Miwa (2018)[5]	93.2	64.0	77.1
Lin et al. (2019)	75.8	73.9	74.8
Luan et al. (2019)	-	-	76.2
Straková et al. (2019)	-	-	78.3
Our model	81.8	79.3	**80.5**

Table 2: State of the art comparison on ACE 2004, ACE 2005 and GENIA corpora for nested NER.

Model	P	R	F1
ONTONOTES			
Chiu and Nichols (2016)	86.0	86.5	86.3
Strubell et al. (2017)	-	-	86.8
Clark et al. (2018)	-	-	88.8
Fisher and Vlachos (2019)	-	-	89.2
Our model	91.1	91.5	**91.3**
CONLL 2003 English			
Chiu and Nichols (2016)	91.4	91.9	91.6
Lample et al. (2016)	-	-	90.9
Strubell et al. (2017)	-	-	90.7
Devlin et al. (2019)	-	-	92.8
Straková et al. (2019)	-	-	93.4
Our model	93.7	93.3	**93.5**
CONLL 2003 German			
Lample et al. (2016)	-	-	78.8
Straková et al. (2019)	-	-	85.1
Our model	88.3	84.6	**86.4**
CONLL 2003 German revised[6]			
Akbik et al. (2018)	-	-	88.3
Our model	92.4	88.2	**90.3**
CONLL 2002 Spanish			
Lample et al. (2016)	-	-	85.8
Straková et al. (2019)	-	-	88.8
Our model	90.6	90.0	**90.3**
CONLL 2002 Dutch			
Lample et al. (2016)	-	-	81.7
Akbik et al. (2019)	-	-	90.4
Straková et al. (2019)	-	-	92.7
Our model	94.5	92.8	**93.7**

ACL2020

A Unified MRC Framework for Named Entity Recognition

Xiaoya Li[♣], Jingrong Feng[♣], Yuxian Meng[♣], Qinghong Han[♣], Fei Wu[♠] and Jiwei Li[♣]
[♠] Department of Computer Science and Technology, Zhejiang University
[♣] Shannon.AI
{xiaoya_li, jingrong_feng, yuxian_meng,qinghong_han}@shannonai.com
wufei@cs.zju.edu.cn, jiwei_li@shannonai.com

Main Work

- In this paper, we reformalize the NER task as a MRC question answering task.

- This formalization comes with two key advantages:
  - being capable of addressing overlapping or nested entities;
  - the query encodes significant prior knowledge about the entity category to extract.

- The proposed method obtains SOTA results on both nested and flat NER datasets, which indicates its effectiveness.

Dataset Construction

- Firstly we need to transform the tagging-style annotated NER dataset to a set of *(QUESTION, ANSWER, CONTEXT)* triples.
- The question generation procedure is important since queries encode prior knowledge about labels and have a significant influence on the final results.

How to Construct Queries

- **Position index of labels:** a query is constructed using the index of a tag to , i.e., "one", "two", "three".
- **Keyword:** a query is the keyword describing the tag, e.g., the question query for tag ORG is *"organization"*.
- **Rule-based template filling:** generates questions using templates. The query for tag ORG is *"which organization is mentioned in the text"*.
- **Wikipedia:** a query is constructed using its wikipedia definition. The query for tag ORG is *"an organization is an entity comprising multiple people, such as an institution or an association."*
- **Synonyms:** are words or phrases that mean exactly or nearly the same as the original keyword extracted using the Oxford Dictionary. The query for tag ORG is *"association"*.
- **Keyword+Synonyms:** the concatenation of a keyword and its synonym.
- **Annotation guideline notes:** is the method we use in this paper. The query for tag ORG is *"find organizations including companies, agencies and institutions"*.

Annotation Guideline Note

Entity	Natural Language Question
Location	Find locations in the text, including non-geographical locations, mountain ranges and bodies of water.
Facility	Find facilities in the text, including buildings, airports, highways and bridges.
Organization	Find organizations in the text, including companies, agencies and institutions.

Table 1: Examples for transforming different entity categories to question queries.

Model

- **Model Backbone**
  - To be in line with BERT, the question and the passage are concatenated, forming the combined string:

$$\{[\text{CLS}], q_1, q_2, ..., q_m, [\text{SEP}], x_1, x_2, ..., x_n\}$$

- **Span Selection**
  - the strategy is to have two binary classifiers, one to predict whether each token is the start index or not, the other to predict whether each token is the end index or not.

$$P_{\text{start}} = \text{softmax}_{\text{each row}}(E \cdot T_{\text{start}}) \in \mathbb{R}^{n \times 2}$$

$$\hat{I}_{\text{start}} = \{i \mid \text{argmax}(P_{\text{start}}^{(i)}) = 1, i = 1, \cdots, n\}$$
$$\hat{I}_{\text{end}} = \{j \mid \text{argmax}(P_{\text{end}}^{(j)}) = 1, j = 1, \cdots, n\}$$

$$P_{i_{\text{start}}, j_{\text{end}}} = \text{sigmoid}(m \cdot \text{concat}(E_{i_{\text{start}}}, E_{j_{\text{end}}}))$$

Train and Test

- Two losses for start and end index predictions:

$$\mathcal{L}_{\text{start}} = \text{CE}(P_{\text{start}}, Y_{\text{start}})$$
$$\mathcal{L}_{\text{end}} = \text{CE}(P_{\text{end}}, Y_{\text{end}})$$

- The start-end index matching loss is given as follows:

$$\mathcal{L} = \alpha\mathcal{L}_{\text{start}} + \beta\mathcal{L}_{\text{end}} + \gamma\mathcal{L}_{\text{span}}$$

Result

English ACE 2004

Model	Precision	Rrecall	F1
Hyper-Graph (Katiyar and Cardie, 2018)	73.6	71.8	72.7
Seg-Graph (Wang and Lu, 2018)	78.0	72.4	75.1
Seq2seq-BERT (Straková et al., 2019)	-	-	84.40
Path-BERT (Shibuya and Hovy, 2019)	83.73	81.91	82.81
DYGIE (Luan et al., 2019)	-	-	84.7
BERT-MRC	**85.05**	**86.32**	**85.98** **(+1.28)**

English ACE 2005

Model	Precision	Recall	F1
Hyper-Graph (Katiyar and Cardie, 2018)	70.6	70.4	70.5
Seg-Graph (Wang and Lu, 2018)	76.8	72.3	74.5
ARN (Lin et al., 2019a)	76.2	73.6	74.9
Path-BERT (Shibuya and Hovy, 2019)	82.98	82.42	82.70
Merge-BERT (Fisher and Vlachos, 2019)	82.7	82.1	82.4
DYGIE (Luan et al., 2019)	-	-	82.9
Seq2seq-BERT (Straková et al., 2019)	-	-	84.33
BERT-MRC	**87.16**	**86.59**	**86.88** **(+2.55)**

English GENIA

Model	Precision	Recall	F1
Hyper-Graph (Katiyar and Cardie, 2018)	77.7	71.8	74.6
ARN (Lin et al., 2019a)	75.8	73.9	74.8
Path-BERT (Shibuya and Hovy, 2019)	78.07	76.45	77.25
DYGIE (Luan et al., 2019)	-	-	76.2
Seq2seq-BERT (Straková et al., 2019)	-	-	78.31
BERT-MRC	**85.18**	**81.12**	**83.75** **(+5.44)**

English KBP 2017

Model	Precision	Recall	F1
KBP17-Best (Ji et al., 2017)	76.2	73.0	72.8
ARN (Lin et al., 2019a)	77.7	71.8	74.6
BERT-MRC	**82.33**	**77.61**	**80.97** **(+6.37)**

English CoNLL 2003

Model	Precision	Recall	F1
BiLSTM-CRF (Ma and Hovy, 2016)	-	-	91.03
ELMo (Peters et al., 2018b)	-	-	92.22
CVT (Clark et al., 2018)	-	-	92.6
BERT-Tagger (Devlin et al., 2018)	-	-	92.8
BERT-MRC	**92.33**	**94.61**	**93.04** **(+0.24)**

English OntoNotes 5.0

Model	Precision	Recall	F1
BiLSTM-CRF (Ma and Hovy, 2016)	86.04	86.53	86.28
Strubell et al. (2017)	-	-	86.84
CVT (Clark et al., 2018)	-	-	88.8
BERT-Tagger (Devlin et al., 2018)	90.01	88.35	89.16
BERT-MRC	**92.98**	**89.95**	**91.11** **(+1.95)**

Chinese MSRA

Model	Precision	Recall	F1
Lattice-LSTM (Zhang and Yang, 2018)	93.57	92.79	93.18
BERT-Tagger (Devlin et al., 2018)	94.97	94.62	94.80
Glyce-BERT (Wu et al., 2019)	95.57	95.51	95.54
BERT-MRC	**96.18**	**95.12**	**95.75** **(+0.21)**

Chinese OntoNotes 4.0

Model	Precision	Recall	F1
Lattice-LSTM (Zhang and Yang, 2018)	76.35	71.56	73.88
BERT-Tagger (Devlin et al., 2018)	78.01	80.35	79.16
Glyce-BERT (Wu et al., 2019)	81.87	81.40	81.63
BERT-MRC	**82.98**	**81.25**	**82.11** **(+0.48)**

Table 3: Results for *flat* NER tasks.

Result

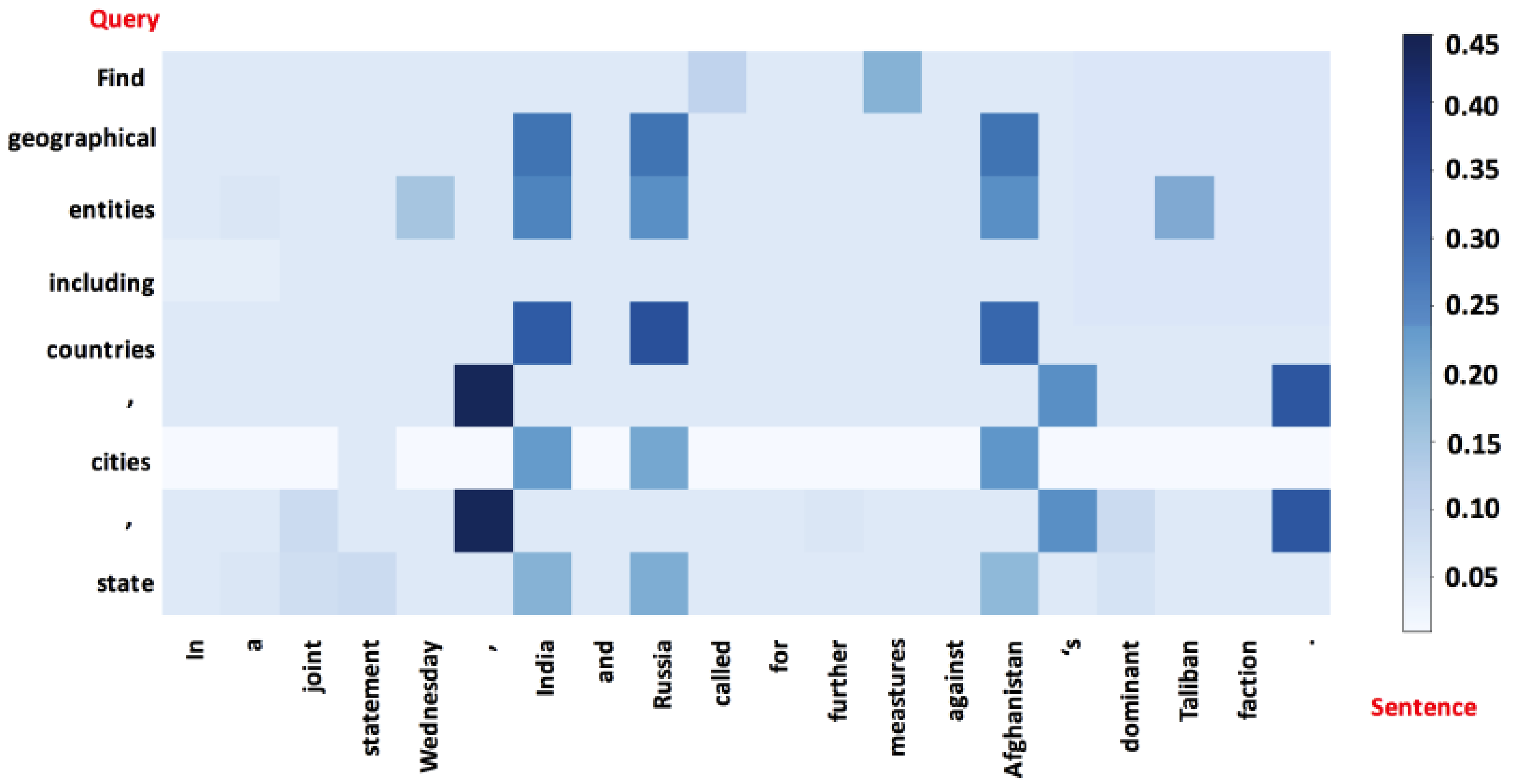
English OntoNotes 5.0	
Model	**F1**
LSTM tagger (Strubell et al., 2017)	86.84
BiDAF (Seo et al., 2017)	87.39 (+0.55)
QAnet (Yu et al., 2018)	87.98 (+1.14)
BERT-Tagger	89.16
BERT-MRC	**91.11** (+1.95)

Table 4: Results of different MRC models on English OntoNotes5.0.

English OntoNotes 5.0	
Model	**F1**
BERT-Tagger	89.16
Position index of labels	88.29 (-0.87)
Keywords	89.74 (+0.58)
Wikipedia	89.66 (+0.59)
Rule-based template filling	89.30 (+0.14)
Synonyms	89.92 (+0.76)
Keywords+Synonyms	90.23 (+1.07)
Annotation guideline notes	91.11 (+1.95)

Table 5: Results of different types of queries.

Result



In a joint statement Wednesday , India and Russia called for further measures against Afghanistan ‘s dominant Taliban faction .

Result

Models	Train	Test	F1
BERT-tagger	OntoNotes5.0	OntoNotes5.0	89.16
BERT-MRC	OntoNotes5.0	OntoNotes5.0	91.11
BERT-tagger	CoNLL03	OntoNotes5.0	31.87
BERT-MRC	CoNLL03	OntoNotes5.0	72.34

Table 6: Zero-shot evaluation on OntoNotes5.0. BERT-MRC can achieve better zero-shot performances.