AAAI2020

Leveraging Multi-token Entities in Document-level Named Entity Recognition

Anwen Hu,[2,3] Zhicheng Dou,[1,2] Jian-Yun Nie,[5] Ji-Rong Wen[3,4]
[1]Gaoling School of Artificial Intelligence, Renmin University of China
[2]School of Information, Renmin University of China
[3]Beijing Key Laboratory of Big Data Management and Analysis Methods
[4]Key Laboratory of Data Engineering and Knowledge Engineering, MOE
[5]Department of Computer Science and Operations Research, University of Montreal
{anwenhu, dou, jrwen}@ruc.edu.cn, nie@iro.umontreal.ca

Main Work

- We propose a novel attention-based document-level NER model that leverages global context features across sentences as supplements to local context features.
- We take advantage of **multi-token entities** in the document to guide NER. Multi-token entities are detected by **an auxiliary sequence tagging task**.
- Experimental results confirm the effectiveness of the proposed method over **the state-of-the-art** sentence-level and document-level NER models.

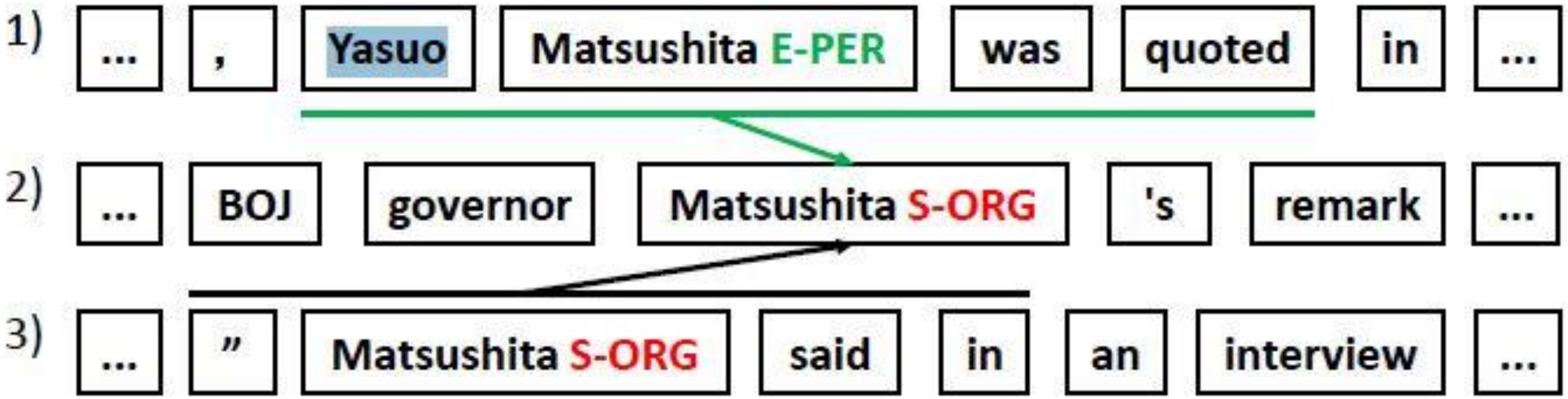


Figure 1: An example of the label inconsistency problem within a document in the CoNLL-2003 English dataset. Green and red tags indicate respectively correct and incorrect tags predicted by a sentence-level model. Green and black arrows refer to useful and less useful contextual information for the second 'Matsushita' token.

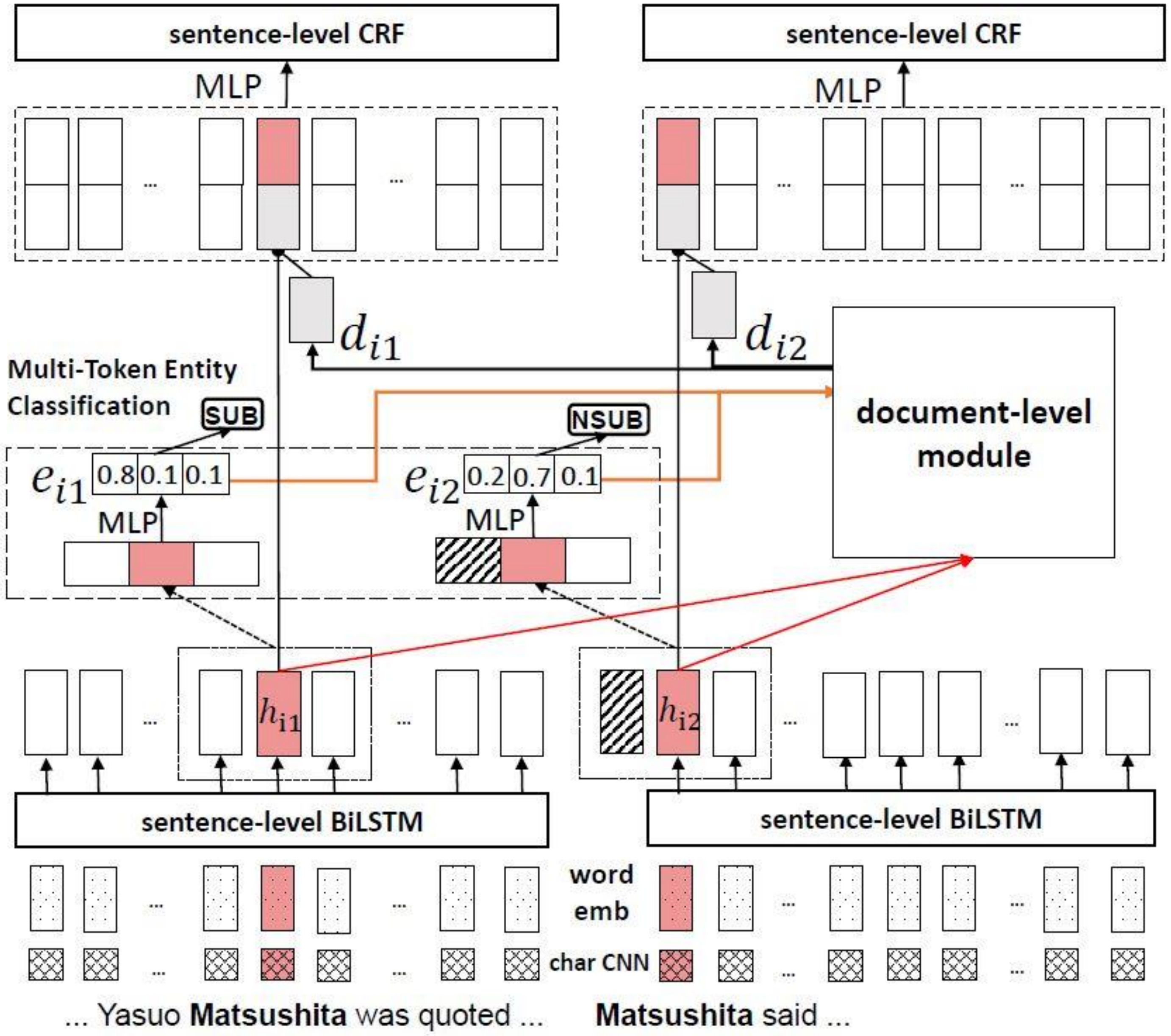
Motivation

- it is common that a multi-token entity such as '**Yasuo Matsushita**' is fully spelled out at the beginning of the article, and then referred to by one of its token (e.g '**Matsushita**') later

- we found **26.62%** of the single-token entities are constituents of multi-token entities in the same document.

- we propose to pay more attention to the token occurrences within multi-token entities

Model



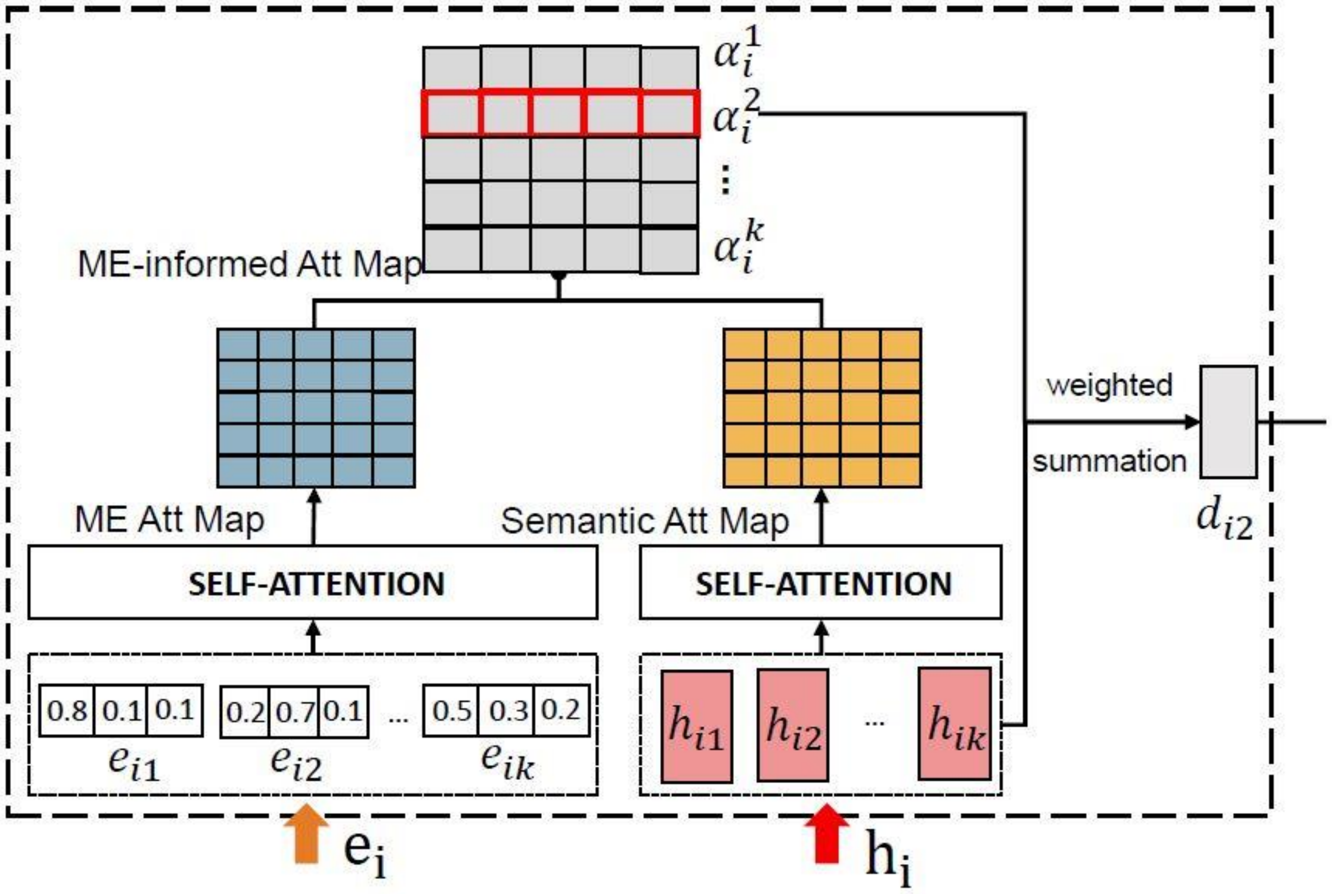
sentence-level CRF
MLP
sentence-level CRF
MLP
d_{i1}
d_{i2}
document-level module
Multi-Token Entity Classification
SUB
NSUB
e_{i1} 0.8 0.1 0.1
e_{i2} 0.2 0.7 0.1
MLP
MLP
h_{i1}
h_{i2}
sentence-level BiLSTM
sentence-level BiLSTM
word emb
char CNN
... Yasuo **Matsushita** was quoted ... **Matsushita** said ...

ME-informed Attention

$$s_{in}^m = v_s^\top \tanh(W_{s_1} h_{im} + W_{s_2} h_{in} + b_s)(h_{im}, h_{in} \in \mathbf{h_i}),$$

$$u_{in}^m = v_u^\top \tanh(W_{u_1} e_{im} + W_{u_2} e_{in} + b_u)(e_{im}, e_{in} \in \mathbf{e_i}),$$

$$\alpha_{in}^m = \mathrm{Softmax}(s_{in}^m + u_{in}^m),$$
$$\alpha_{\mathbf{i}}^{\mathbf{m}} = (\alpha_{i1}^m, \alpha_{i2}^m, ..., \alpha_{ik}^m),$$
$$d_{im} = \sum_{n=1}^{k} \alpha_{in}^m h_{in}, (\alpha_{in}^m \in \alpha_{\mathbf{i}}^{\mathbf{m}}, h_{in} \in \mathbf{h_i})$$



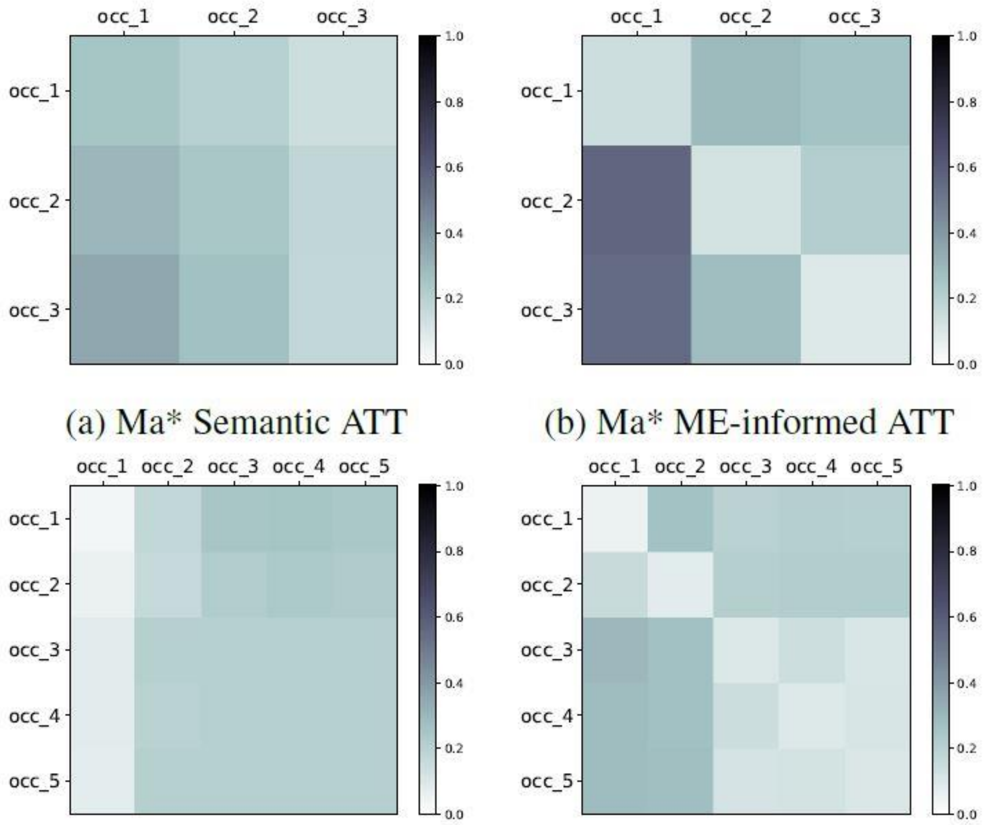
$\mathbf{e_i} = (e_{i1}, e_{i2}, ..., e_{ik}),$ $\mathbf{h_i} = (h_{i1}, h_{i2}, ..., h_{ik}),$

Result

Model	**CoNLL-2003**	**OntoNotes$_{\text{nbm}}$**
LSTM+CRF	90.94	87.57
BiLSTM+CNNS+CRF	91.21	88.42
ParallelRNNs	91.48	85.54
HSCRFs(JNT)	91.38	87.74
Att+BiLSTM+CRF	90.49	88.88
IDCNN	90.65	85.24
GlobalAtt	91.43	88.78
SENT	90.92	88.64
MEID-SEM	91.71	88.71
MEID-ME	91.78	88.84
MEID	**91.92**	**89.16**

Dataset	**Model**	**glove**	**bert-base**	**flair**
CoNLL-2003	SENT	90.92	90.66	92.59
	MEID	**91.92**	**91.47**	**93.09**
OntoNotes$_{\text{nbm}}$	SENT	88.64	88.41	89.89
	MEID	**89.16**	**88.96**	**90.29**

Case Analysis



(a) Ma* Semantic ATT (b) Ma* ME-informed ATT

(c) ZI* Semantic ATT (d) ZI* ME-informed ATT

Sentences in a Document

1) In a rare expression of a view on currencies by the Bank of Japan (BOJ) governor , Yasuo **Matsushita** *E-PER* was quoted in Japan 's leading economic daily on Friday as ...
2) ... was BOJ governor **Matsushita** *S-ORG* *S-PER* ‘s remark .
3) ..." **Matsushita** *S-PER* said in an interview with the ...

1) GLF – **ZIMBABWE** *B-LOC* *B-MISC* OPEN SECND RUND SCRES .
2) Leading second round scores in the **Zimbabwe** *B-MISC* Open at the par-72 Chapman Golf Club on Friday : 132 Des Terblanche 65 67 133 Mark McNulty (**Zimbabwe** *S-LOC*) 72 61 134 Steve van Vuuren 65 69 136 Nick Price (**Zimbabwe** *S-LOC*) 68 68 , Justin Hobday 71 61
3) Andrew Pitts (U.S.) 69 67 138 Mark Cayeux (**Zimbabwe** *S-LOC*) 69 69 , ...

Figure 4: Sample NER results. Blue tags mean correct tags predicted by both MEID-ME and MEID. Red tags mean wrong tags predicted by MEID-ME. Green tags mean correct tags predicted by MEID.