

BIOES tagging, Conditional Random Field

BIOES tagging, CRF

Geonil Yun,
rjsdlf45@gmail.com
2021. 08. 03.



Introduction

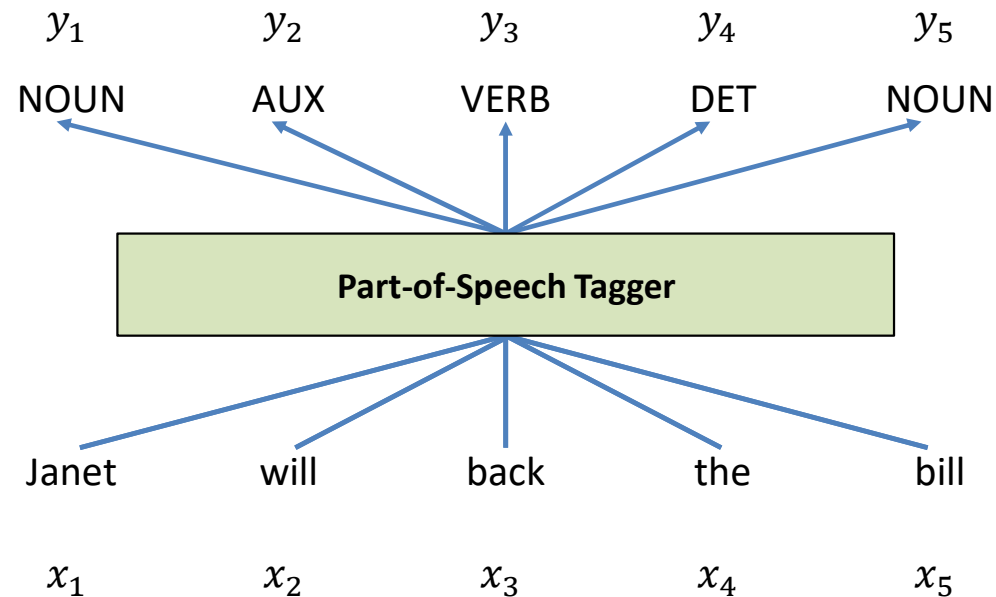
- Tagging

- Part-of-Speech (POS) tagging

- **VERB** : **Book** that flight
 - **NOUN** : Read a **book**.

sequence labeling : from $\{x_1 \dots x_t\}$ to $\{y_1 \dots y_t\}$

words ↖ POS ↖



Introduction

- Tagging

- Named Entity Recognition (NER) tagging

- **PER (Person)** : Janet
- **LOC (Location)** : New York
- **ORG (Organization)** : Samsung
- **TIME (Time)** : Friday

- each word does not have **one tag**

2-word : "Jane Villanueva"
3-word : "United Airlines Holding"
1-word : "Chicago"

Jane Villanueva of United Airlines Holding
discussed the Chicago route.



NER Tagging

Words	Tag
Jane	PER
Villanueva	PER
of	
United	ORG
Airlines	ORG
Holding	ORG
discussed	
the	
Chicago	LOC
route	
.	



BIOES

- BIO tagging for NER

- Tagged by one word

- **B** : token that begins a span
- **I** : tokens inside a span
- **O** : tokens outside of any span

- each word gets **one tag**

[Jane *B-PER*, Villanueva *I-PER*]

[United *B-ORG*, Airlines *I-ORG*, Holding *I-ORG*]

[Chicago *B-LOC*]

Jane Villanueva of United Airlines Holding
discussed the Chicago route.



NER with BIO tagging

Words	Tag	BIO Tag
Jane	PER	B-PER
Villanueva	PER	I-PER
of		O
United	ORG	B-ORG
Airlines	ORG	I-ORG
Holding	ORG	I-ORG
discussed		O
the		O
Chicago	LOC	B-LOC
route		O
.		O

BIOES

- BIOES tagging (BIO variants)

- Tagged by one word

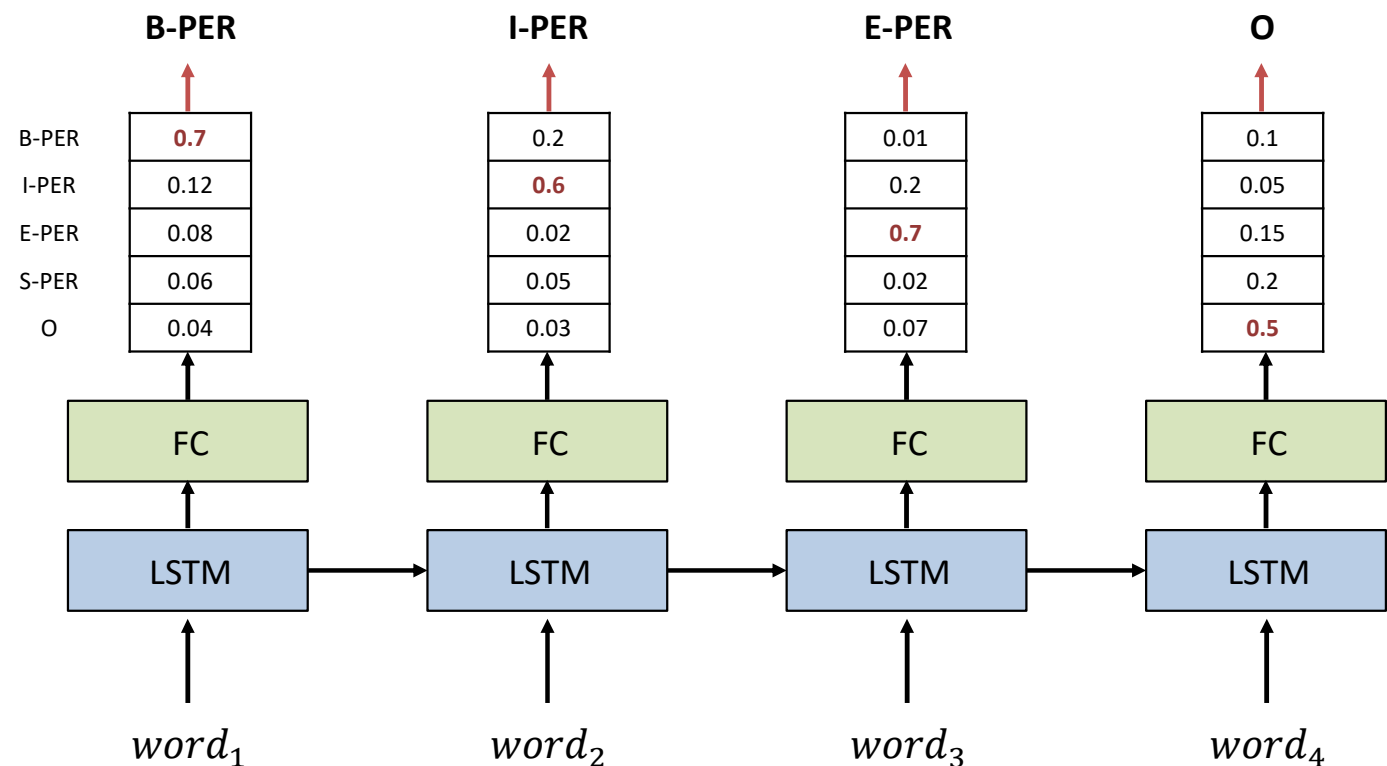
- **B** : token that begins a span
- **I** : tokens inside a span
- **O** : tokens outside of any span
- **E** : token that ends a span
- **S** : token that gets single span

Words	Tag	BIO Tag	BIOES Tag
Jane	PER	B-PER	B-PER
Villanueva	PER	I-PER	E-PER
of		O	O
United	ORG	B-ORG	B-ORG
Airlines	ORG	I-ORG	I-ORG
Holding	ORG	I-ORG	E-ORG
discussed		O	O
the		O	O
Chicago	LOC	B-LOC	S-LOC
route		O	O
.		O	O

Conditional Random Field

- Tagging task (BIOES) with LSTM

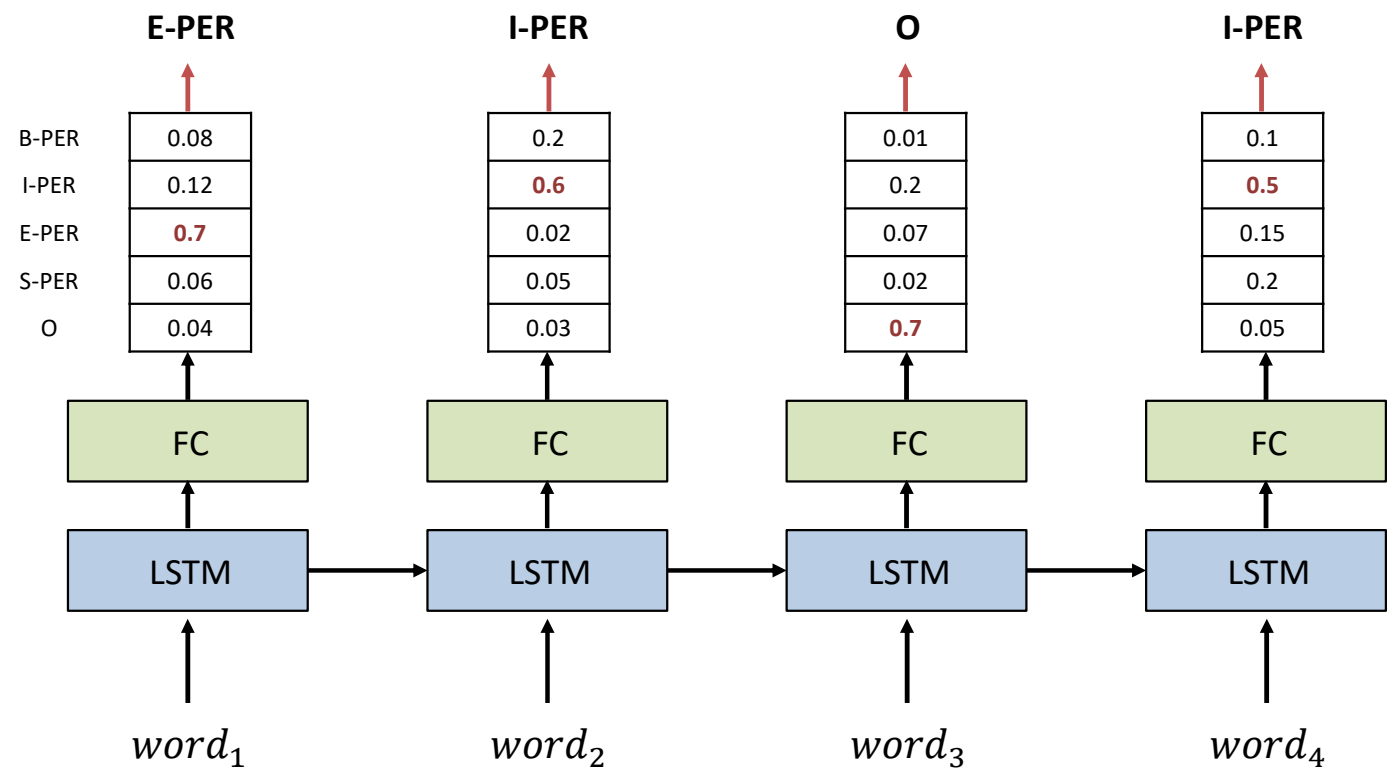
- word 1, 2, 3 represent PERSON
- word 4 is outside span



Conditional Random Field

- Tagging task (BIOES) with LSTM (Problem)

- Cannot start with E
- Cannot start with I

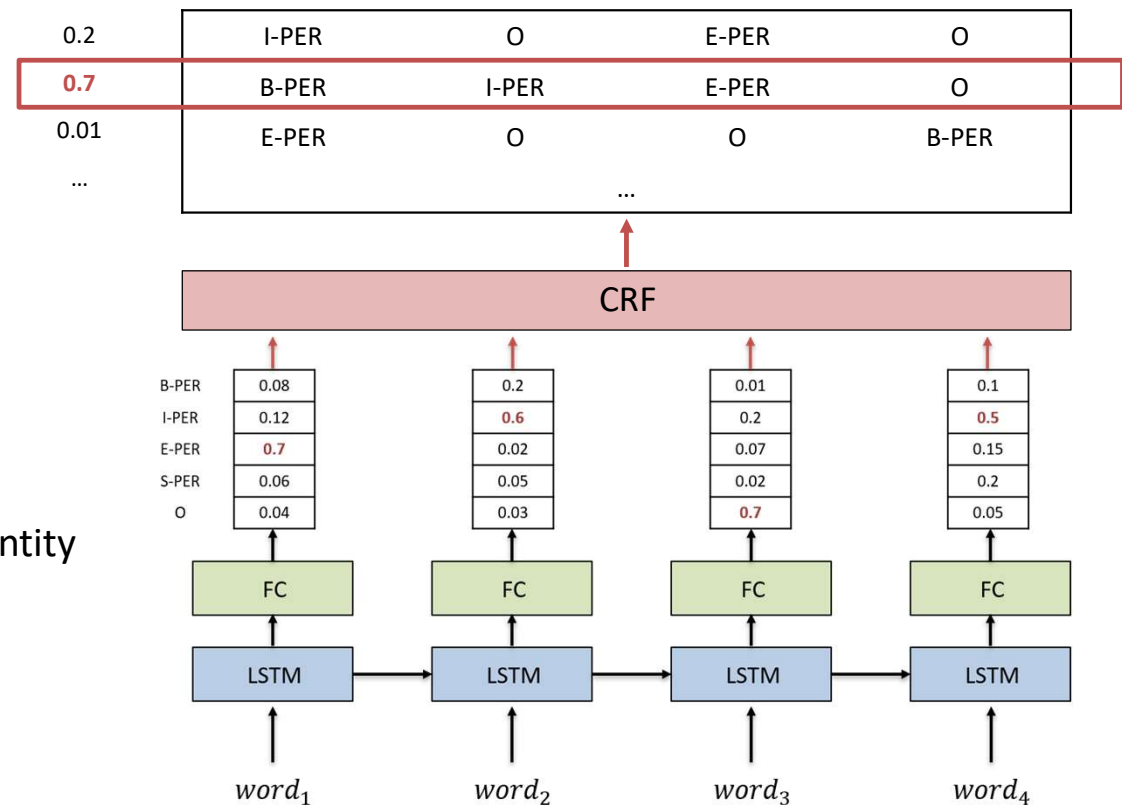


Conditional Random Field

- Tagging task (BIOES) with LSTM + CRF

- CRF learn **BIOES features**
- CRF learn other features of tagging

1. Not start with I, E
2. E comes with end of 2 words entity
3. O-E, O-I patterns not exist



Thank you