Nested Named Entity Recognition Revisited

Arzoo Katiyar and **Claire Cardie**
Department of Computer Science
Cornell University

JOSE COVES

INTRODUCTION

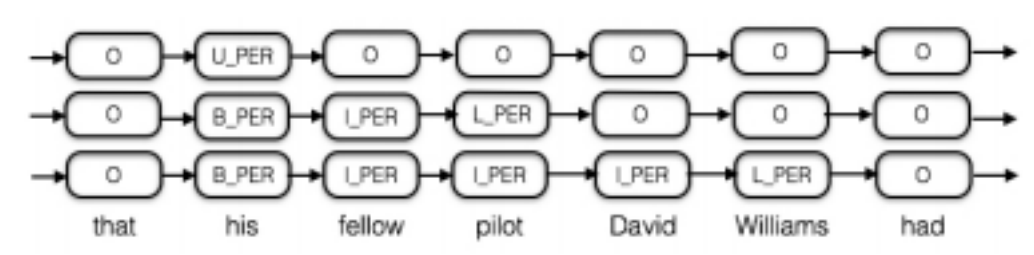
- RNN for nested NER and mention detection
- 17% and 30% in some corpora (GENIA, ACE)
- Previously, Conditional Random Fields
- BILOU encoding
- Hypergprah
  - Edges, arcs
  - Hyperarcs (connects sets of nodes)

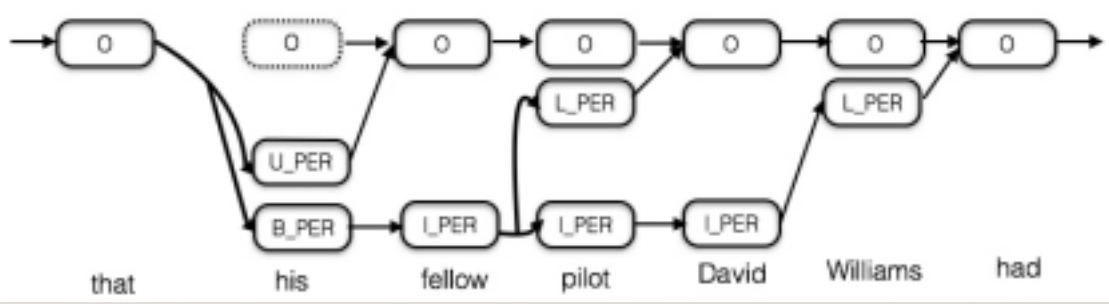
(S1) Employing the [EBV - transformed [human B cell line]$_{\text{CELL\_LINE}}$]$_{\text{CELL\_LINE}}$ SKW6.4, we demonstrate . . .

(S2) . . . [the burial site of [Sheikh Abbad]$_{\text{PERSON}}$]$_{\text{LOCATION}}$ is located . . .

HYPERGRAPH

- 'O' at every time step; possible new entity
- Edges: Simple labeling task
- Hyperarcs: Multi-label learning problem
- Assign probabilities to edges
- Argmax for edges
- Hyperarcs above threshold
- Adj matrix using dfs





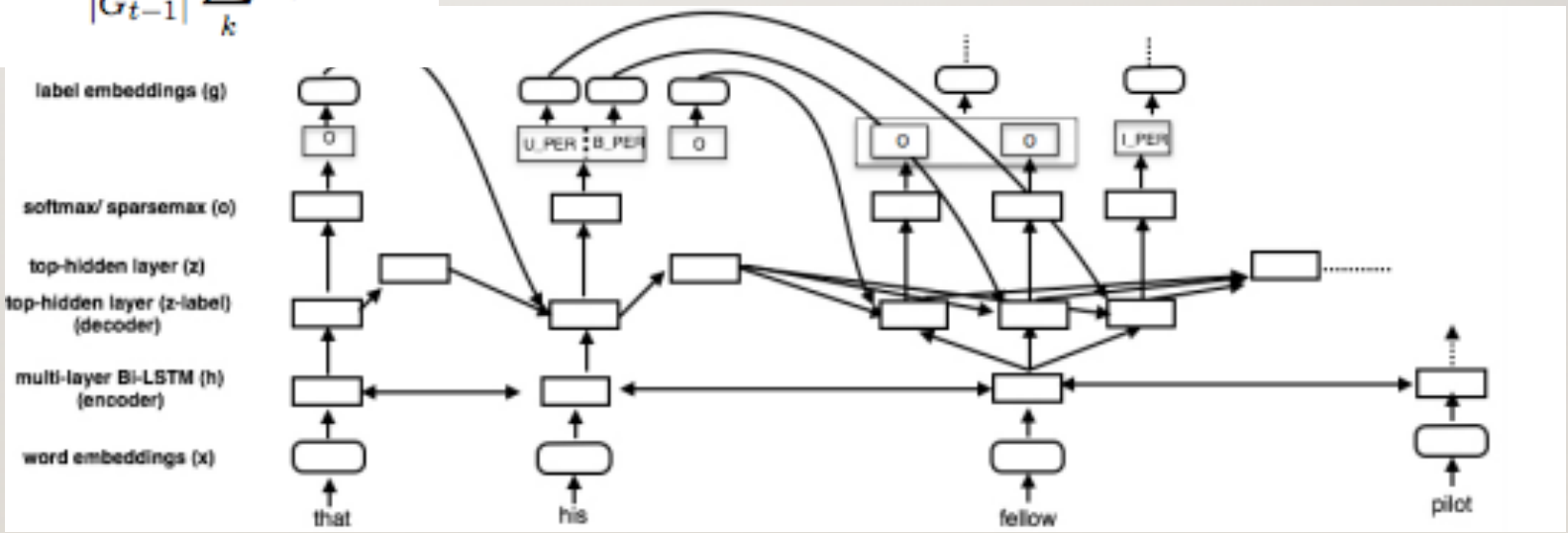
MODEL

- Multi-layer bi-LSTM (concat forward and backward output)
- Top hidden layer (pred)
- Next step

$$\overrightarrow{\boldsymbol{h}}_t^{(l)} = \mathrm{LSTM}(\boldsymbol{x}_t, \overrightarrow{\boldsymbol{h}}_{t-1})$$
$$\overleftarrow{\boldsymbol{h}}_t^{(l)} = \mathrm{LSTM}(\boldsymbol{x}_t, \overleftarrow{\boldsymbol{h}}_{t+1})$$
$$\boldsymbol{z}_t^{(l)} = \overrightarrow{\boldsymbol{V}}\overrightarrow{\boldsymbol{h}}_t^{(l)} + \overleftarrow{\boldsymbol{V}}\overleftarrow{\boldsymbol{h}}_t^{(l)} + \boldsymbol{b}^l$$

$$\boldsymbol{h}_t^{(L),k} = \mathrm{LSTM}(\boldsymbol{z}_t^{(L-1)}, \boldsymbol{h}_{t-1}^{(L)}, \boldsymbol{g}_{t-1}^{k})$$

$$\boldsymbol{h}_t^{(L)} = \frac{1}{|G_{t-1}|}\sum_k \boldsymbol{h}_t^{(L),k}$$



ENTITY EXTRACTION + TRAINING

- Probability distribution over candidate labels
- Softmax , Sparsemax (zeros)
- 3 hidden layers, unit dim = 100, T = 0.3

$$\boldsymbol{o}_t^k = \boldsymbol{U}\boldsymbol{h}_t^{(L),k} + \boldsymbol{b}$$
$$\hat{\boldsymbol{e}}_t^k = \mathrm{softmax}(\boldsymbol{o}_t^k)$$
$$p(y_t = c | y_{t-1} = g_{t-1}^k) = (\boldsymbol{e}_t^k)_c$$

$$\ell_{t(\mathrm{softmax})}^k = -\sum_c \left(\boldsymbol{e}_t^k\right)_c \log\left(\hat{\boldsymbol{e}}_t^k\right)_c$$

$$\ell_{t(\mathrm{sparsemax})}^k = -2\boldsymbol{e}_t^{k\top}\boldsymbol{o}_t^k + \sum_{c:(\hat{\boldsymbol{e}}_t^k)_c \neq 0} \left((\boldsymbol{o}_t^k)_c^2 - \tau^2\right)$$

$$\mathcal{L} = \sum_t \sum_{k \in G_{t-1}} \ell_t^k.$$

RESULTS

	ACE2004			ACE2005		
Method	P	R	F1	P	R	F1
MH-F (Lu and Roth, 2015)	74.4	50.0	59.8	63.4	53.8	58.3
Our model(softmax)	68.2	60.5	64.2	67.5	62.3	64.8
Our model(sparsemax)	**72.3**	**66.8**	**69.7**	**70.6**	**69.8**	**70.2**

Method	P	R	F1
Finkel and Manning (2009)	75.4	65.9	70.3
MH-F (Lu and Roth, 2015)	72.5	65.2	68.7
Muis and Lu (2017)	75.4	66.8	70.8
LSTM-flat	75.5	63.5	68.9
LSTM-output_layer	78.4	67.9	72.8
Our model (softmax)	76.7	**71.1**	**73.8**
Our model (sparsemax)	**79.8**	68.2	73.6