



Why Attention? Analyze BiLSTM Deficiency and Its Remedies in the Case of NER

Peng-Hsuan Li

jacobvsdaniel@iis.sinica.edu.tw

Tsu-Jui Fu

tsu-juifu@ucsb.edu

Wei-Yun Ma

ma@iis.sinica.edu.tw

XOR Limitation for BiLSTM

1. Key and Peele (*work-of-art*)
2. You and I (*work-of-art*)
3. Key and I
4. You and Peele

$$\begin{aligned}s_t &= H_t W_p + b \\ &= \vec{H}_t \vec{W}_p + \overleftarrow{H}_t \overleftarrow{W}_p + b \\ &= \vec{s}_t + \overleftarrow{s}_t + b\end{aligned}$$

i: inside-work-of-art
j: outside

$$\begin{aligned}\vec{s}_{2i}^1 + \overleftarrow{s}_{2i}^1 + b_i &> \vec{s}_{2j}^1 + \overleftarrow{s}_{2j}^1 + b_j \\ \vec{s}_{2i}^2 + \overleftarrow{s}_{2i}^2 + b_i &> \vec{s}_{2j}^2 + \overleftarrow{s}_{2j}^2 + b_j \\ \vec{s}_{2i}^3 + \overleftarrow{s}_{2i}^3 + b_i &< \vec{s}_{2j}^3 + \overleftarrow{s}_{2j}^3 + b_j \\ \vec{s}_{2i}^4 + \overleftarrow{s}_{2i}^4 + b_i &< \vec{s}_{2j}^4 + \overleftarrow{s}_{2j}^4 + b_j\end{aligned}$$

$$\begin{aligned}\vec{s}_{2i}^1 + \overleftarrow{s}_{2i}^1 + b_i &> \vec{s}_{2j}^1 + \overleftarrow{s}_{2j}^1 + b_j \\ \vec{s}_{2i}^2 + \overleftarrow{s}_{2i}^2 + b_i &> \vec{s}_{2j}^2 + \overleftarrow{s}_{2j}^2 + b_j \\ \vec{s}_{2i}^1 + \overleftarrow{s}_{2i}^2 + b_i &< \vec{s}_{2j}^1 + \overleftarrow{s}_{2j}^2 + b_j \\ \vec{s}_{2i}^2 + \overleftarrow{s}_{2i}^1 + b_i &< \vec{s}_{2j}^2 + \overleftarrow{s}_{2j}^1 + b_j\end{aligned}$$

XOR Limitation for CRF-BiLSTM

1. a_o m_s c_o
2. b_o m_s d_o
3. a_o m_o d_o
4. b_o m_o c_o

Viterbi with softmax (p) and transition (q)

$$\begin{aligned}lp(-s|am) + lp(s|mc) + lq(oso) &> lp(-o|am) + lp(o|mc) + lq(ooo) \\ lp(-s|bm) + lp(s|md) + lq(oso) &> lp(-o|bm) + lp(o|md) + lq(ooo) \\ lp(-s|am) + lp(s|md) + lq(oso) &< lp(-o|am) + lp(o|md) + lq(ooo) \\ lp(-s|bm) + lp(s|mc) + lq(oso) &< lp(-o|bm) + lp(o|mc) + lq(ooo)\end{aligned}$$

Viterbi, suppose no transition

$$\begin{aligned}p(oso|amc) &> p(ooo|amc) \\ p(oso|bmd) &> p(ooo|bmd) \\ p(oso|amd) &< p(ooo|amd) \\ p(oso|bmc) &< p(ooo|bmc)\end{aligned}$$

$$\begin{aligned}lp(-s|amc) &> lp(-o|amc) \\ lp(-s|bmd) &> lp(-o|bmd) \\ lp(-s|amd) &< lp(-o|amd) \\ lp(-s|bmc) &< lp(-o|bmc)\end{aligned}$$

$$\begin{aligned}lp(-s|am) + lp(s|mc) &> lp(-o|am) + lp(o|mc) \\ lp(-s|bm) + lp(s|md) &> lp(-o|bm) + lp(o|md) \\ lp(-s|am) + lp(s|md) &< lp(-o|am) + lp(o|md) \\ lp(-s|bm) + lp(s|mc) &< lp(-o|bm) + lp(o|mc)\end{aligned}$$

BiLSTM

$$\vec{H} = \overrightarrow{LSTM}_2(\overrightarrow{LSTM}_1(X))$$

$$\overleftarrow{H} = \overleftarrow{LSTM}_4(\overleftarrow{LSTM}_3(X))$$

$$H = \vec{H} \parallel \overleftarrow{H},$$

$$s_t = H_t W_p + b$$

-> Cross

$$\vec{H}^1 = \overrightarrow{LSTM}_1(X)$$

$$\overleftarrow{H}^3 = \overleftarrow{LSTM}_3(X)$$

$$\vec{H}^2 = \overrightarrow{LSTM}_2(\vec{H}^1 \parallel \overleftarrow{H}^3)$$

$$\overleftarrow{H}^4 = \overleftarrow{LSTM}_4(\vec{H}^1 \parallel \overleftarrow{H}^3)$$

$$H = \vec{H}^2 \parallel \overleftarrow{H}^4$$

-> Att

$$\alpha^i = \sigma \left(\frac{HW^{qi}(HW^{ki})^T}{\sqrt{d_c}} \right)$$

$$C^i = \alpha^i HW^{vi},$$

$$s_t^c = (H_t \parallel C_t^1 \parallel C_t^2 \parallel \dots \parallel C_t^m) W_c + b$$

Results

	OntoNotes 5.0			WNUT 2017		
	Prec.	Rec.	F1	Prec.	Rec.	F1
BiLSTM-CNN	86.04	86.53	86.28±0.26	-	-	-
CRF-IDCNN	-	-	86.84±0.19	-	-	-
CRF-BiLSTM(-BiLSTM*)	-	-	86.99±0.22	-	-	38.24
Baseline-BiLSTM-CNN	88.37	87.14	87.75±0.14	53.24	32.93	40.68±1.78
Cross-BiLSTM-CNN	88.37	88.17	88.27±0.17	58.28	33.92	42.85±0.99
Att-BiLSTM-CNN	88.71	88.11	88.40±0.18	55.82	34.08	42.26±0.82

-> By Length

	OntoNotes 5.0				WNUT 2017			
	1	2	3	3+	1	2	3	3+
Cross	+0.3%	+0.6%	+1.8%	+1.3%	+1.7%	+2.9%	+8.7%	+5.4%
Att	+0.1%	+1.1%	+2.3%	+1.8%	+1.5%	+2.0%	+2.6%	+0.9%

-> By Component

	$H C^{all}$	Att-BiLSTM-CNN							Baseline-...
		H	C^{all}	C^1	C^2	C^3	C^4	C^5	H
O	99.05	-1.68	0.75	0.95	-1.67	-45.57	-0.81	-35.46	-0.03
S	93.74	2.69	-91.02	-90.56	-90.88	-25.61	-86.25	-84.32	0.13
B	90.99	1.21	-52.26	-90.78	-88.08	-90.88	-12.21	-87.45	-0.63
I	90.09	-28.18	-3.80	-87.93	-60.56	-50.19	-57.19	-79.63	-0.41
E	93.23	2.00	-71.50	-93.12	-36.45	-39.19	-91.90	-90.83	-0.38

Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli

Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli
Clinton	will	host	a	meeting	at	the	White	house	Saturday	with	Israeli

President **Clinton** will host a meeting at **the White house** **Saturday** with **Israeli** and **Palestinian** negotiators ...