国际人工智能会议
AAAI 2021 论文北京预讲会

Entity Structure Within and Throughout:

Modeling Mention Dependencies for Document-Level Relation Extraction

Benfeng Xu[1], Quan Wang[2], Yajuan Lyu[2], Yong Zhu[2], Zhendong Mao[1]

[1]School of Information Science and Technology, University of Science and Technology of China

[2]Baidu Inc.

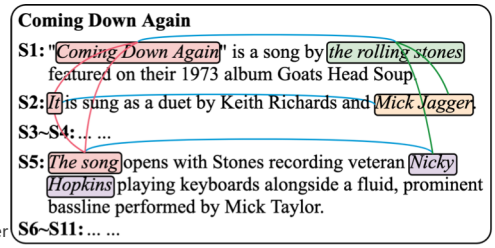
Introduction

- Entities, as the essential elements in relation extraction tasks, exhibit certain structure. We formulate such structure as distinctive dependencies between mention pairs, and propose SSAN to incorporate them within the standard self-attention mechanism and throughout the overall encoding stage.
- Specifically, we design two alternative transformation modules inside each self-attention building block to produce attentive biases so as to adaptively regularize its attention flow.
- We achieve new state-of-the-art results on three popular document-level relation extraction datasets and the visualization shows how entity structure guides the model for better relation extraction.

Motivating Example

Various dependencies indicate rich interactions among entity mentions, and thereby provide informative priors for relation extraction, e.g.:

- Blue link: reside in the same sentence and depend on local context
- Red link: coreference to each other
- Green link: no direct connection, but might be associated via multiple hops



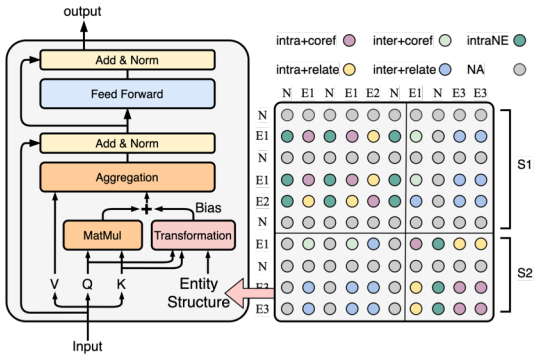
S:Coming Down Again	R:Performer	O:Rolling Stones
S:Mick Jagger	R:Member of	O:Rolling Stones
S:Nicky Hopkins	R:Member of	O:Rolling Stones

Approach

How to formulate the structure of entities?

- **Co-occurrence structure** distinguishes intra-sentential interactions that depend on local context from inter-sentential ones that require cross sentence reasoning.
- **Coreference structure** distinguishes coreferential mention pairs from related (under the schema) mention pairs

		Coreference	
		True	False
Co-occurence	True	*intra+coref*	*intra+relate*
	False	*inter+coref*	*inter+relate*



How to model entity structure?

- Vanilla self-attention

$$e^l_{ij} = \frac{(x^l_i W^Q_l)(x^l_j W^K_l)^T}{\sqrt{d}} = \frac{q^l_i k^{l^T}_i}{\sqrt{d}}$$

- Biaffine Transformation

$$\tilde{e}^l_{ij} = \frac{q^l_i k^{l^T}_i + q^l_i A_{l,s_{i,j}} k^{l^T}_i + b_{l,s_{i,j}}}{\sqrt{d}}$$

- Decomposed Linear Transformation

$$\tilde{e}^l_{ij} = \frac{q^l_i k^{l^T}_i + q^l_i K^T_{l,s_{i,j}} + Q_{l,s_{i,j}} k^{l^T}_i + b_{l,s_{i,j}}}{\sqrt{d}}$$

$s_{i,j}$: structural prior between token i and token j, $s_{i,j} \in$ {intra+coref, inter+coref, intra+relate, inter+relate, intranet, NA}

$A_{l,s_{i,j}}, K^T_{l,s_{i,j}}, Q_{l,s_{i,j}}, b_{l,s_{i,j}}$: learnable modules

Results

DocRED

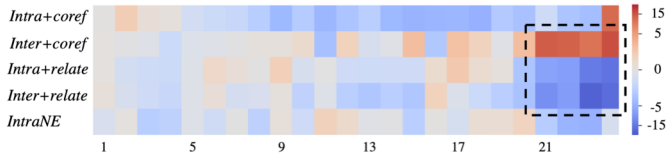
Model	Dev Ign F1 / F1	Test Ign F1 / F1
ContextAware (2019)	48.94 / 51.09	48.40 / 50.70
EoG*(2019)	45.94 / 52.15	49.48 / 51.82
BERT Two-Phase (2019a)	- / 54.42	- / 53.92
GloVe+LSR (2020)	48.82 / 55.17	52.15 / 54.18
HINBERT (2020)	54.29 / 56.31	53.70 / 55.60
CorefBERT Base (2020)	55.32 / 57.51	54.54 / 56.96
CorefBERT Large (2020)	56.73 / 58.88	56.48 / 58.70
BERT+LSR (2020)	52.43 / 59.00	56.97 / 59.05
CorefRoBERTa (2020)	57.84 / 59.93	57.68 / 59.91
BERT Base Baseline	56.29 / 58.60	55.08 / 57.54
$SSAN_{Decomp}$	56.68 / 58.95	**56.06 / 58.41**
$SSAN_{Biaffine}$	**57.03 / 59.19**	55.84 / 58.16
BERT Large Baseline	58.11 / 60.18	57.91 / 60.03
$SSAN_{Decomp}$	58.42 / 60.36	57.97 / 60.01
$SSAN_{Biaffine}$	**59.12 / 61.09**	**58.76 / 60.81**
RoBERTa Base Baseline	57.47 / 59.52	57.27 / 59.48
$SSAN_{Decomp}$	58.29 / 60.22	**57.72** / 59.75
$SSAN_{Biaffine}$	**58.83 / 60.89**	57.71 / **59.94**
RoBERTa Large Baseline	58.45 / 60.58	58.43 / 60.54
$SSAN_{Decomp}$	59.54 / 61.50	59.11 / 61.24
$SSAN_{Biaffine}$	**60.25 / 62.08**	**59.47 / 61.42**
+ Adaptation	**63.76 / 65.69**	**63.78 / 65.92**

CDR & GDA

Model	Dev F1	Test F1	Intra- / Inter-Test F1
(Gu et al. 2017)	-	61.3	57.2 / 11.7
BRAN(2018)	-	62.1	- / -
CNN+CNNchar(2018)	-	62.3	- / -
GCNN(2019)	57.2	58.6	- / -
EoG (2019)	63.6	63.6	68.2 / 50.9
LSR (2020)	-	61.2	66.2 / 50.3
LSR w/o MDP (2020)	-	64.8	68.9 / 53.1
BERT (2020)	-	60.5	- / -
SciBERT (2020)	-	64.0	- / -
BERT Base Baseline	61.7	61.4	69.3 / 44.9
$SSAN_{Decomp}$	63.0	61.2	68.6 / **45.1**
$SSAN_{Biaffine}$	**64.7**	**62.7**	**70.4** / 44.7
BERT Large Baseline	65.3	63.6	70.8 / 49.0
$SSAN_{Decomp}$	64.9	64.5	71.2 / 50.2
$SSAN_{Biaffine}$	**65.8**	**65.3**	**71.4 / 52.0**
SciBERT Baseline	68.2	65.8	71.9 / 53.3
$SSAN_{Decomp}$	67.9	67.0	72.6 / 55.8
$SSAN_{Biaffine}$	**68.4**	**68.7**	**74.5 / 56.2**

Model	Dev F1	Test F1	Intra- / Inter-Test F1
EoG (2019)	78.7	81.5	85.2 / 49.3
LSR (2020)	-	79.6	83.1 / 49.6
LSR w/o MDP (2020)	-	82.2	85.4 / 51.1
BERT Base Baseline	79.8	81.2	84.7 / 60.3
$SSAN_{Decomp}$	81.5	**83.4**	**86.7 / 62.3**
$SSAN_{Biaffine}$	**81.6**	82.1	86.1 / 56.8
BERT Large Baseline	80.4	81.6	84.9 / 61.5
$SSAN_{Decomp}$	82.0	83.8	86.6 / **65.0**
$SSAN_{Biaffine}$	**82.2**	**83.9**	**86.9** / 63.9
SciBERT Baseline	81.4	83.6	**87.2** / 61.8
$SSAN_{Decomp}$	82.5	83.2	87.0 / 60.0
$SSAN_{Biaffine}$	**82.8**	**83.7**	86.6 / **65.3**

Ablation



Visualization of produced structural biases w.r.t. each dependency in each layer

Dependency	Ign F1	F1
$SSAN_{Biaffine}$ (RoBERTa Large)	60.25	62.08
− *intra+coref*	59.59	61.57
− *intra+relate*	59.92	61.91
− *inter+coref*	59.87	61.74
− *inter+relate*	59.92	61.84
− *intraNE*	59.96	61.97
− all	58.45	60.58

Ablation on entity structure formulation

Bias Term	Ign F1	F1
RoBERTa Large baseline (w/o bias)	58.45	60.58
$+b_{s_{i,j}}$	58.62	60.59
$+Q_{s_{i,j}} k^T_j$	58.79	60.65
$+q_i K^T_{s_{i,j}}$	59.26	61.31
$+q_i K^T_{s_{i,j}} + Q_{s_{i,j}} k^T_j + b_{s_{i,j}}$	59.54	61.50
$+q_i A_{s_{i,j}} k^T_j$	59.83	61.75
$+q_i A_{s_{i,j}} k^T_j + b_{s_{i,j}}$	60.25	62.08

Ablation on Transformation modules

主办方：中国中文信息学会青年工作委员会

承办方：智源社区