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.

[*] This work was done while the first author is an intern at Baidu Inc.

Introduction

Task: Document-level Relation Extraction

- Input:
  - document-level text
  - multiple entities along with their mentions
- Target:
  - identify relations for every entity pair

Kungliga Hovkapellet

[1] ***Kungliga Hovkapellet*** (The ***Royal Court Orchestra***) is a *Swedish* orchestra, originally part of the *Royal Court* in ***Sweden***'s capital *Stockholm*. [2] The orchestra originally consisted of both musicians and singers. [3] It had only male members until *1727*, when *Sophia Schröder* and *Judith Fischer* were employed as vocalists; in the *1850s*, the harpist *Marie Pauline Åhman* became the first female instrumentalist. [4] From *1731*, public concerts were performed at ***Riddarhuset*** in *Stockholm*. [5] Since *1773*, when the ***Royal Swedish Opera*** was founded by *Gustav III* of ***Sweden***, the ***Kungliga Hovkapellet*** has been part of the opera's company.

Subject:	***Kungliga Hovkapellet; Royal Court Orchestra***	
Object:	***Royal Swedish Opera***	
Relation:	`part_of`	**Supporting Evidence:** 5
Subject:	***Riddarhuset***	
Object:	***Sweden***	
Relation:	`country`	**Supporting Evidence:** 1, 4

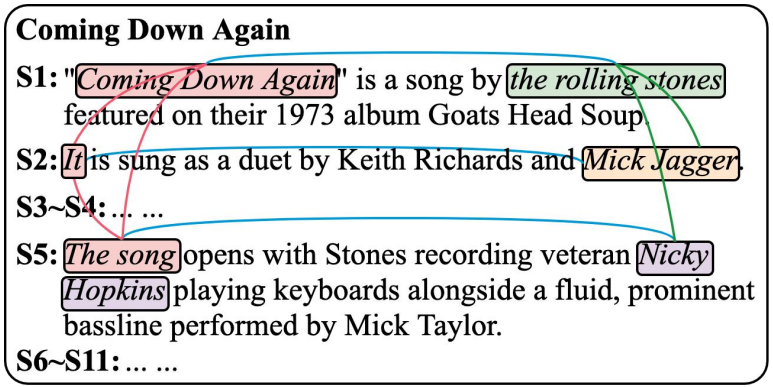
An example from DocRED dataset[1]

[1] Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., ... & Sun, M. (2019). DocRED: A large-scale document-level relation extraction dataset. arXiv preprint arXiv:1906.06127.

A Motivating Example

Various dependencies among mentions indicates rich interactions, and thereby provides informative priors for relation extraction task:

- 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

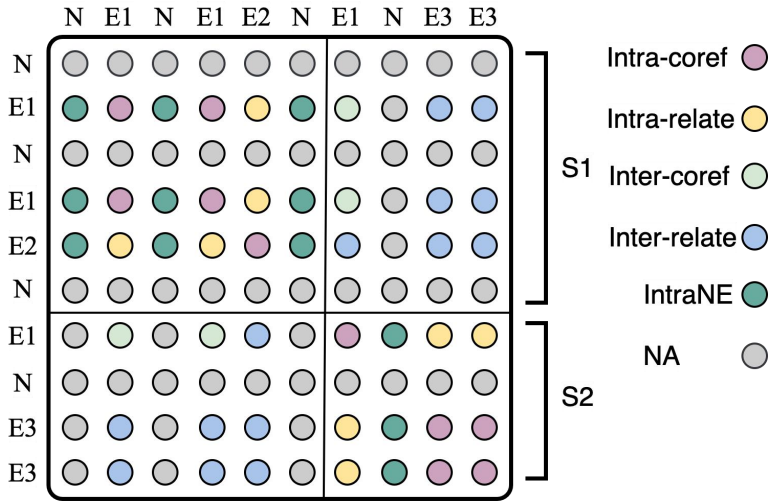


S:*Coming Down Again*	R:Performer	O:*Rolling Stones*
S:*Mick Jagger*	R:Member of	O:*Rolling Stones*
S:*Nicky Hopkins*	R:Member of	O:*Rolling Stones*

Approach: Formulate Entity Structure

- Co-occurrence structure: whether or not two mentions reside in the same sentence.
- Coreference structure: whether or not two mentions refer to the same entity.
- IntraNE: intra-sentential non-entity words (local

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



Approach: Structured Self-Attention Network

- Vanilla self-attention

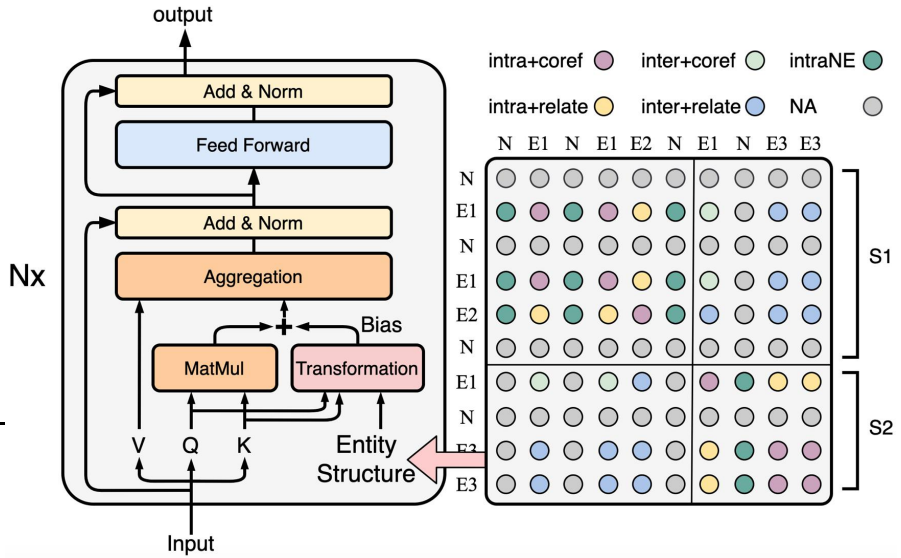
$$e^{l}_{ij} = \frac{(x^{l}_{i} W^{Q}_{l})(x^{l}_{j} W^{K}_{l})^{T}}{\sqrt{d}} = \frac{q^{l}_{i} k^{l\,T}_{i}}{\sqrt{d}}$$

- Biaffine Transformation

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

- Decomposed Linear Transformation

$$e^{l}_{ij} = \frac{q^{l}_{i} k^{l\,T}_{i} + q^{l}_{i} K^{T}_{l,\,s_{i,j}} + Q_{l,\,s_{i,j}} k^{l\,T}_{i} + b_{l,\,s_{i,j}}}{\sqrt{d}}$$



$s_{i,j}$: 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

Experiments

Model	Dev Ign F1 / F1	Test Ign F1 / F1
ContexAware (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	57.63 / 59.73	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**

DocRED

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	- / -
methods using external resources			
(Peng, Wei, and Lu 2016)	-	63.1	- / -
(Li et al. 2016b)	-	67.7	58.9 / -
(Panyam et al. 2018)	-	60.3	65.1 / 45.7
(Zheng et al. 2018)	-	61.5	- / -
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**

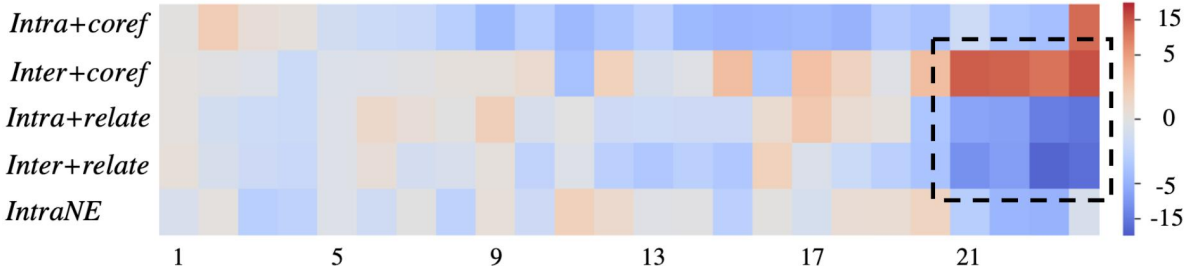
CDR

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**

GDA

Ablation Study

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



- Ablation on entity structure formulation

Dependency	Ign F1	F1
$\text{SSAN}_{\text{Biaffine}}$ (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 Transformation modules

Bias Term	Ign F1	F1
RoBERTa Large baseline (w/o bias)	58.45	60.58
$+b_{s_{ij}}$	58.62	60.59
$+\boldsymbol{Q}_{s_{ij}}\boldsymbol{k}_j^T$	58.79	60.65
$+\boldsymbol{q}_i\boldsymbol{K}_{s_{ij}}^T$	59.26	61.31
$+\boldsymbol{q}_i\boldsymbol{K}_{s_{ij}}^T + \boldsymbol{Q}_{s_{ij}}\boldsymbol{k}_j^T + b_{s_{ij}}$	59.54	61.50
$+\boldsymbol{q}_i\boldsymbol{A}_{s_{ij}}\boldsymbol{k}_j^T$	59.83	61.75
$+\boldsymbol{q}_i\boldsymbol{A}_{s_{ij}}\boldsymbol{k}_j^T + b_{s_{ij}}$	60.25	62.08

国际人工智能会议
AAAI 2021 论文北京预讲会

THANKS

2020.12.19