OneEE: A One-Stage Framework for Fast Overlapping and Nested Event Extraction

Hu Cao[1*], Jingye Li[1*], Fangfang Su[1], Fei Li[1], Hao Fei[2], Shengqiong Wu[2], Bobo Li[1], Liang Zhao[3], Donghong Ji[1†]

[1] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, China

[2] School of Computing, National University of Singapore, Singapore

[3] Department of Computing and Mathematics, University of São Paulo, Brazil

{whucaohu, theodorelee, lifei_csnlp, dhji}@whu.edu.cn

COLING-2022

Code: https://github.com/Cao-Hu/OneEE

Leibniz-Institut
für Sozialwissenschaften

Reported by Dongdong Hu

Introduction

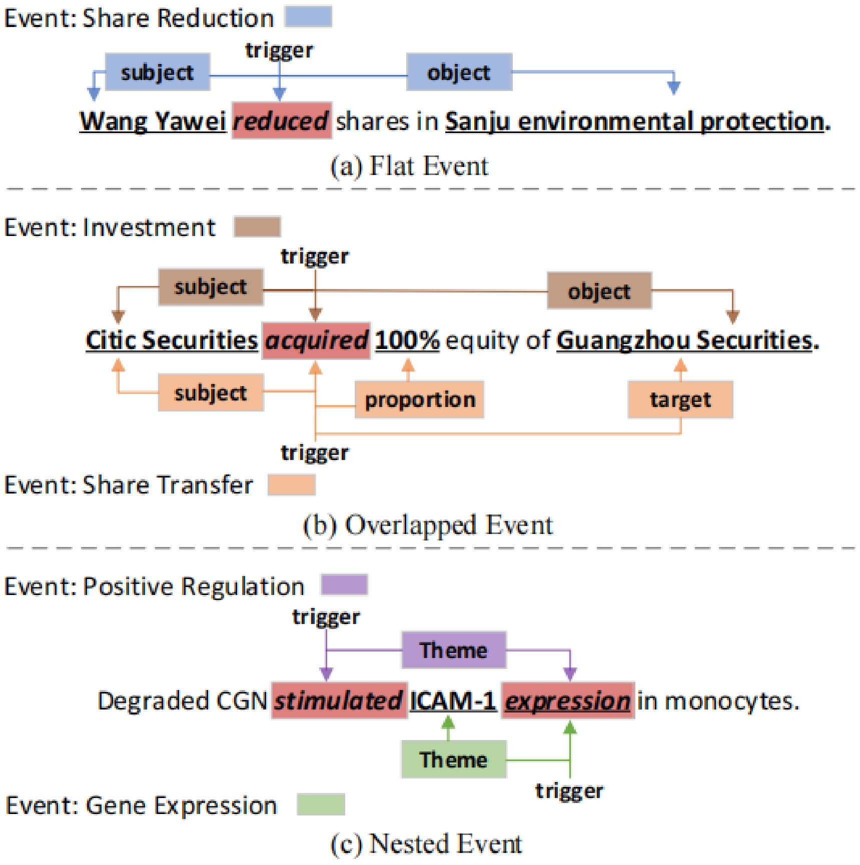


Figure 1: Examples of three kinds of events, including a flat event (a), overlapped events (b), and nested events (c). Different event mentions are denoted in distinct colors. Triggers are marked with red boxes while arguments are underlined.

A few models for overlapped and nested EE includes several successive stages to extract event triggers and arguments, which suffer from error propagation.

Method

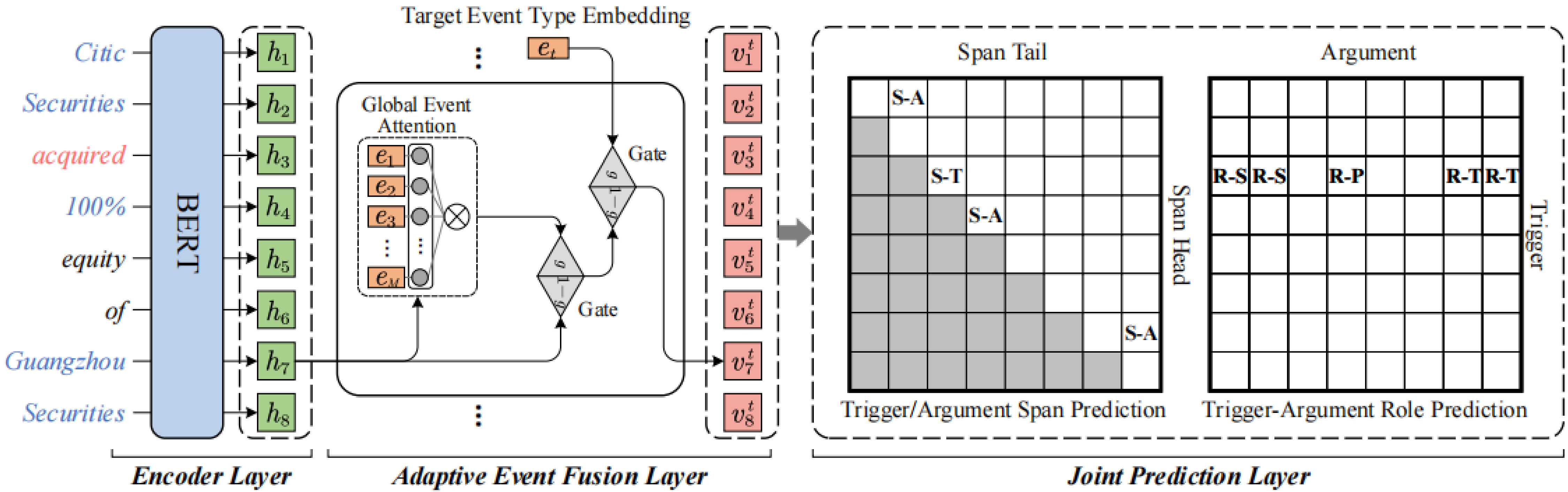
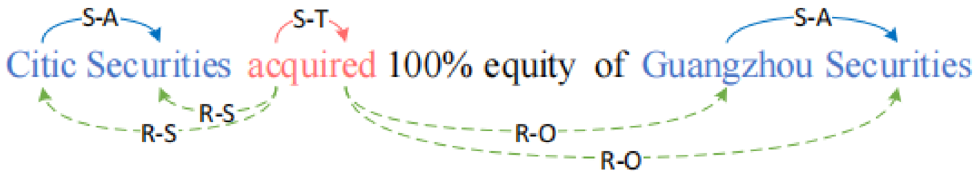
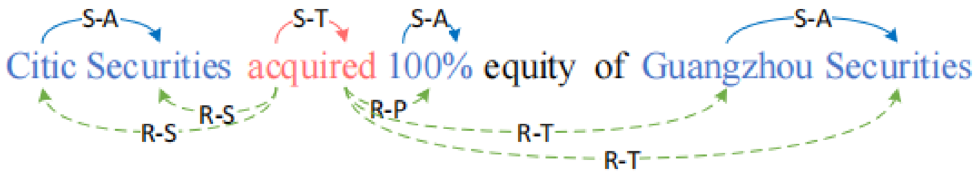


Figure 3: The architecture of our framework. Given a target event type embedding e_t of type t (e.g., transfer share), the goal of our framework is to identify its triggers, arguments, and corresponding roles in the input sentence.

Method



(a) Event: Investment



(b) Event: Transfer Share

Figure 2: Two examples to illustrate our tagging scheme. We formalize the overlapping and nested EE as word-word relation recognition, where S-T and S-A denote the relations between the head and tail boundary words of a trigger or argument, and R-S, R-O, R-T, and R-P denote the relations between the trigger word and the argument words with the roles “subject”, “object”, “target” and “proportion”.

Problem Formulation

$$X = \{x_1, x_2, \ldots, x_N\}$$

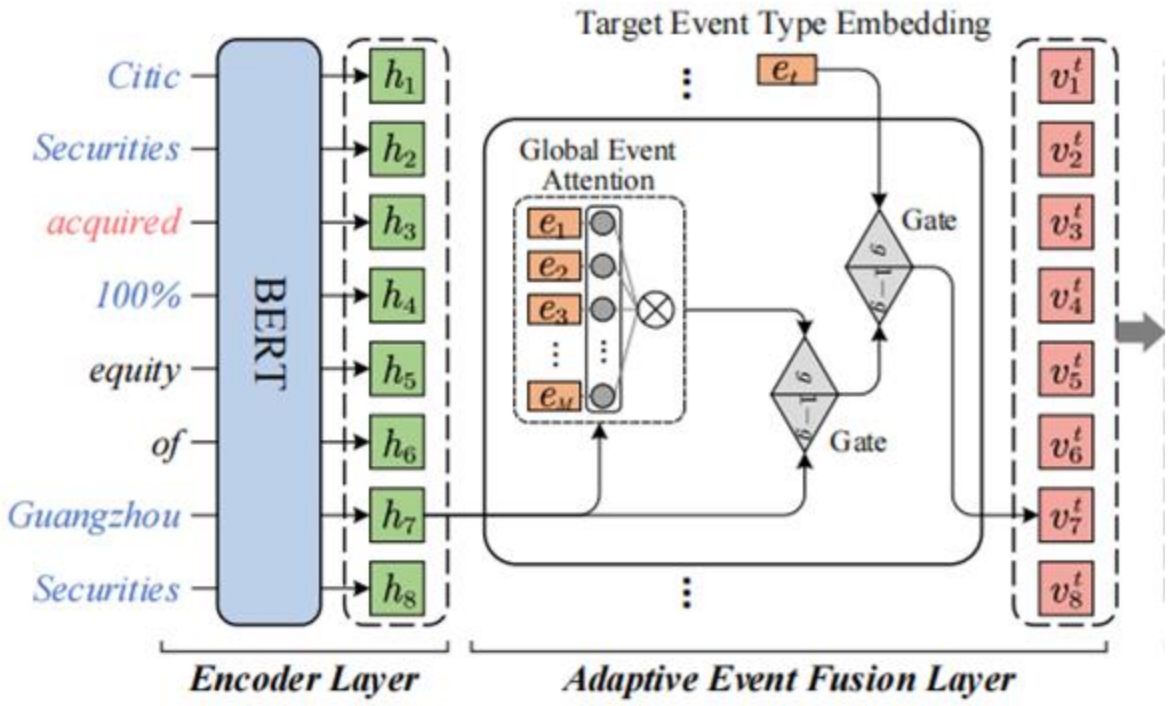
event type $e \in \mathcal{E}$

span relations $\mathcal{S}$

role relations $\mathcal{R}$

- $\mathcal{S}$: the span relation indicates that x_i and x_j are the starting and ending token of the extracted trigger span `S-T` or argument span `S-A`, where $1 \leq i \leq j \leq N$.
- $\mathcal{R}$: the role relation indicates that the argument with x_j acts the certain role `R-*` of the event with the trigger containing x_i, where $1 \leq i, j \leq N$. $\star$ indicates the role type.
- `NONE`, indicating that the word pair does not have any relation defined in this paper.

Method



Encoder Layer

$$X = \{x_1, x_2, \ldots, x_N\}$$

$$\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_N\} \in \mathbb{R}^{N \times d_h}$$

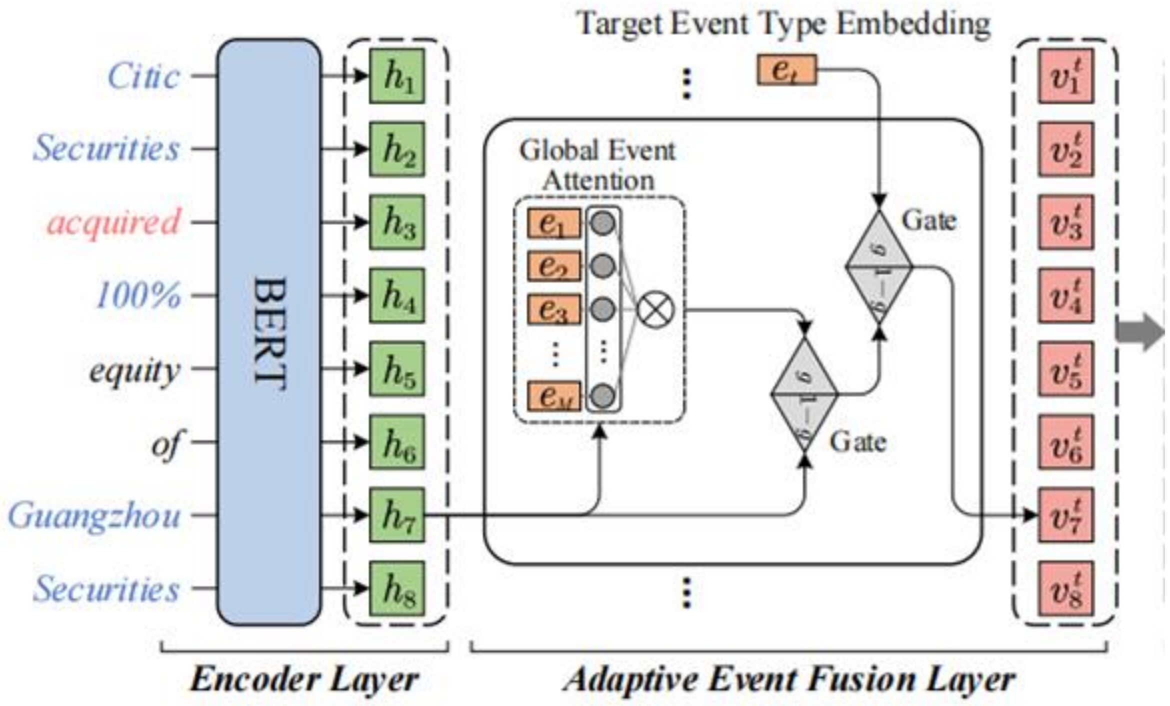
Adaptive Event Fusion Layer

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^\top}{\sqrt{d_h}})\boldsymbol{V} , \quad (1)$$

$$\boldsymbol{E}^g = \text{Attention}(\boldsymbol{W}_q\boldsymbol{H}, \boldsymbol{W}_k\boldsymbol{E}, \boldsymbol{W}_v\boldsymbol{E}) , \quad (4)$$

$$\boldsymbol{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_M\} \in \mathbb{R}^{M \times d_h}$$

Method



$$\boldsymbol{H}^g = \mathrm{Gate}(\boldsymbol{H}, \boldsymbol{E}^g) \;, \quad (5)$$

$$\boldsymbol{V}^t = \mathrm{Gate}(\boldsymbol{H}^g, \boldsymbol{e}_t) \;, \quad (6)$$

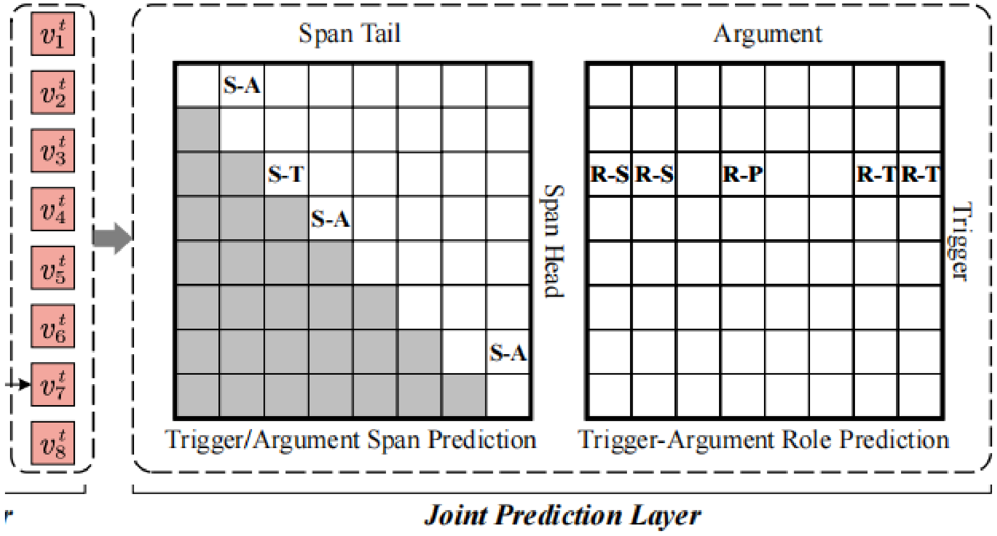
$$\mathrm{Gate}(\boldsymbol{p}, \boldsymbol{q}) = \boldsymbol{g} \odot \boldsymbol{p} + (1 - \boldsymbol{g}) \odot \boldsymbol{q} \;, \quad (2)$$

$$\boldsymbol{g} = \sigma(\boldsymbol{W}_g[\boldsymbol{p}; \boldsymbol{q}] + \boldsymbol{b}_g) \;, \quad (3)$$

where $\boldsymbol{p}$ and $\boldsymbol{q}$ are input vectors.

$$\boldsymbol{V}^t = \{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N\} \in \mathbb{R}^{N \times d_h}$$

Experiments



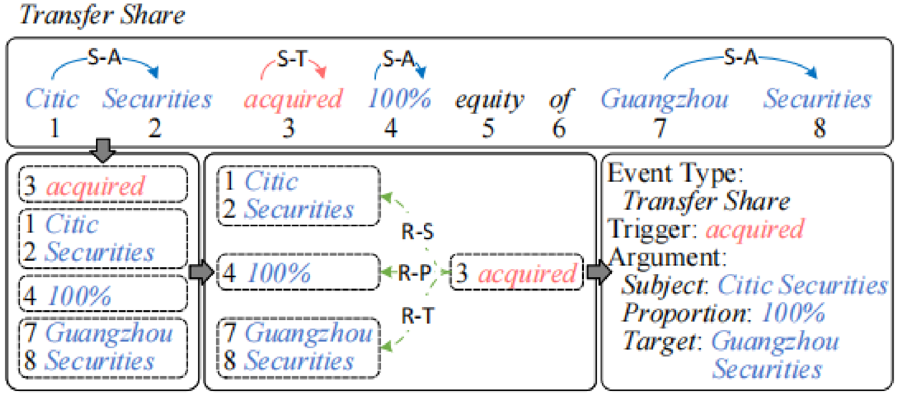


Figure 4: A decoding case of our system with four steps.

Joint Prediction Laver

$$\begin{aligned} \mathrm{Score}(\boldsymbol{p}_i, \boldsymbol{p}_j) &= (\boldsymbol{R}_i \boldsymbol{p}_i)^\top (\boldsymbol{R}_j \boldsymbol{p}_j) \\ &= \boldsymbol{p}_i^\top \boldsymbol{R}_{j-i} \boldsymbol{p}_j \ , \end{aligned} \tag{7}$$

where $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$ are position embeddings of $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$, $\boldsymbol{R}_{j-i} = \boldsymbol{R}_i^\top \boldsymbol{R}_j$.

$$c_{ij}^s = \mathrm{Score}(\boldsymbol{W}_{s1} \boldsymbol{v}_i^t, \boldsymbol{W}_{s2} \boldsymbol{v}_j^t) \ , \tag{8}$$

$$c_{ij}^r = \mathrm{Score}(\boldsymbol{W}_{r1} \boldsymbol{v}_i^t, \boldsymbol{W}_{r2} \boldsymbol{v}_j^t) \ , \tag{9}$$

$$\mathcal{L}^\star = \log(e^\delta + \sum_{(i,j)\in\Omega^\star} e^{-c_{ij}^\star}) + \log(e^\delta + \sum_{(i,j)\notin\Omega^\star} e^{c_{ij}^\star}) \ , \tag{10}$$

where $\Omega^\star$ denotes the pair set of relation $\star$, δ is set to zero.

$$\mathcal{L} = \sum_{t\in\mathcal{E}'} (\sum_{s\in\mathcal{S}} \mathcal{L}^s + \sum_{r\in\mathcal{R}} \mathcal{L}^r) \ , \tag{11}$$

where $\mathcal{S}'$ is a subset sampled from $\mathcal{S}$

Method

		#Ovlp.	#Nest.	#Sent.	#Events
FewFC	train	1,560	-	7,185	10,227
	dev	205	-	899	1,281
	test	210	-	898	1,332
Genia11	train	954	1,628	8,730	6,401
	dev	121	199	1,091	824
	test	125	197	1,092	775
Genia13	train	347	784	4,000	2,743
	dev	44	100	500	352
	test	42	88	500	320

Table 1: Statistics of the datasets. “Ovlp.” and “Nest.” denote the sentences with overlapped and nested events, respectively.

Experiments

		TI(%)			TC(%)			AI(%)			AC(%)		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
Flat EE	BERT-softmax	89.8	79.0	84.0	80.2	61.8	69.8	74.6	62.8	68.2	72.5	60.2	65.8
	BERT-CRF	**90.8**	80.8	85.5	**81.7**	63.6	71.5	75.1	64.3	69.3	72.9	61.8	66.9
	BERT-CRF-joint	89.5	79.8	84.4	80.7	63.0	70.8	**76.1**	63.5	69.2	**74.2**	61.2	67.1
Ovlp. & Nest. EE	PLMEE	83.7	85.8	84.7	75.6	74.5	75.1	74.3	67.3	70.6	72.5	65.5	68.8
	MQAEE	89.1	85.5	87.4	79.7	76.1	77.8	70.3	68.3	69.3	68.2	66.5	67.3
	CasEE	89.4	87.7	88.6	77.9	78.5	78.2	72.8	73.1	72.9	71.3	71.5	71.4
Ours	OneEE	88.7	**88.7**	**88.7**	79.1	**80.3**	**79.7**	75.4	**77.0**	**76.2**	74.0	**72.9**	**73.4**

Table 2: Results for extracting all kinds of events on FewFC, where TI, TC, AI, AC denote trigger identification, trigger classification, argument identification, and argument classification, respectively. We run our model for 5 times with different random seeds and report the median values.

Experiments

	TI(%)	TC(%)	AI(%)	AC(%)
• **Genia11**				
BERT-softmax	67.8	64.4	57.4	56.0
BERT-CRF	68.3	64.8	58.3	56.9
BERT-CRF-joint	67.0	64.1	60.2	58.1
PLMEE	67.3	65.5	60.7	59.4
CasEE	70.0	67.0	62.0	60.4
OneEE	**71.5**	**69.5**	**65.9**	**62.5**
• **Genia13**				
BERT-softmax	77.4	75.9	69.9	67.7
BERT-CRF	78.8	77.4	70.1	68.2
BERT-CRF-joint	77.6	75.7	71.9	68.2
PLMEE	79.3	78.3	72.1	70.7
CasEE	80.5	78.5	73.7	71.9
OneEE	**81.9**	**80.8**	**76.8**	**72.7**

Table 3: F1 scores for extracting all events on Genia11 and Genia13.

	TI(%)	TC(%)	AI(%)	AC(%)
OneEE	88.7	79.7	76.2	73.4
w/o Attention	88.3	79.5	75.9	72.8
w/o Gate	88.4	79.3	75.3	72.6
w/o Fusion Layer	88.0	78.7	75.2	72.2
w/o Position Emb.	88.1	78.7	74.1	71.8

Table 4: Ablation studies using FewFC.

Experiments

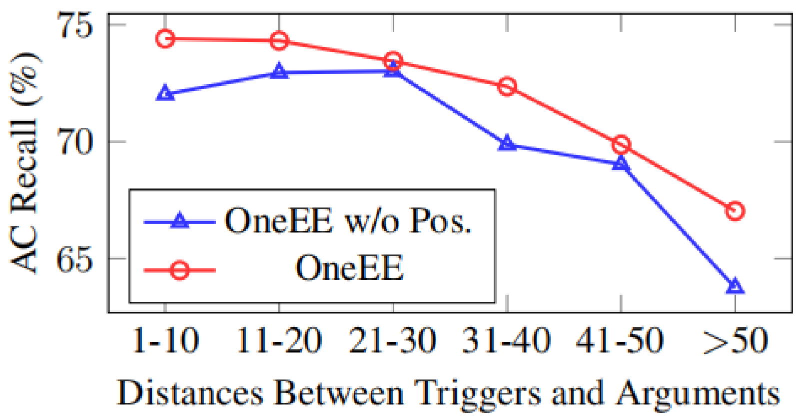


Figure 6: FewFC results of extracting triggers and arguments with different distances. The red line denotes that position embeddings are used while the blue line not.

Model	Stage	#Param.	Speed (sent/s)	Ratio
PLMEE	Two	204.6M	19.8	×1.0
CasEE	Three	120.7M	62.3	×3.1
OneEE	One	114.2M	79.4	×4.0
OneEE†	One	114.2M	186.5	×9.4

Table 5: Parameter number and inference speed comparisons on FewFC. All models are tested with batch size 1, † denotes that the model is tested with batch size 8. The ratio denotes the multiple of the speed increase with regard to PLMEE.

Experiments

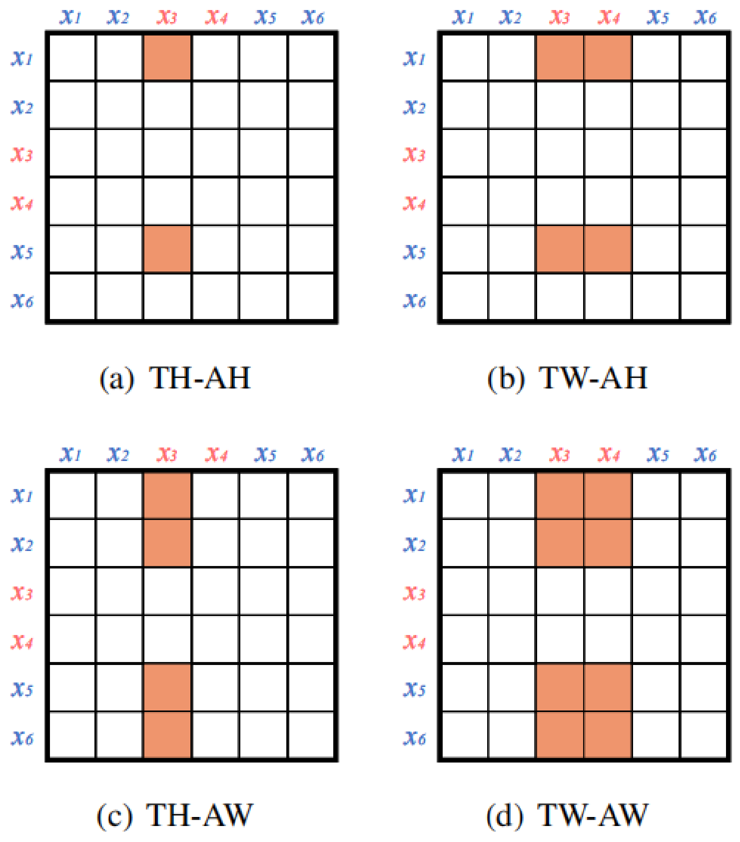


Figure 7: Four kinds of role label strategies. The goal is to predict the relation between trigger head and argument head (a), trigger word and argument head (b), trigger head and argument word (c), and trigger word and argument word (d).

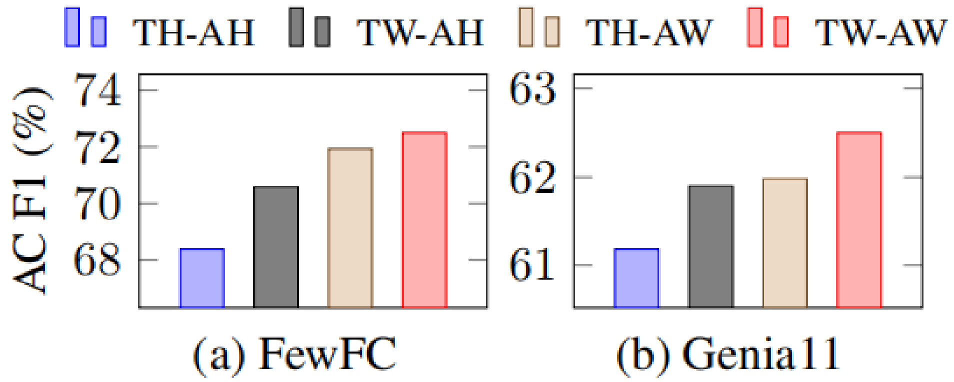


Figure 8: Results of AC with different role label strategies on FewFC (a) and Genia11 (b) datasets.

Experiments

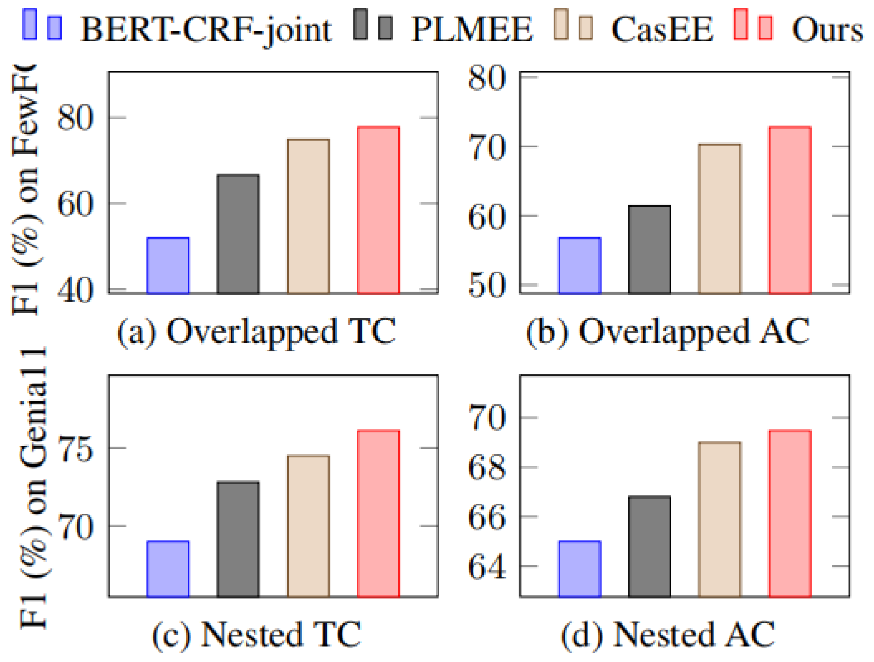


Figure 5: Results for overlapped trigger (a) and argument (b) extraction on FewFC, and nested trigger (c), and argument (d) extraction on Genia11. Note that only the sentences that contain at least one such event are used.

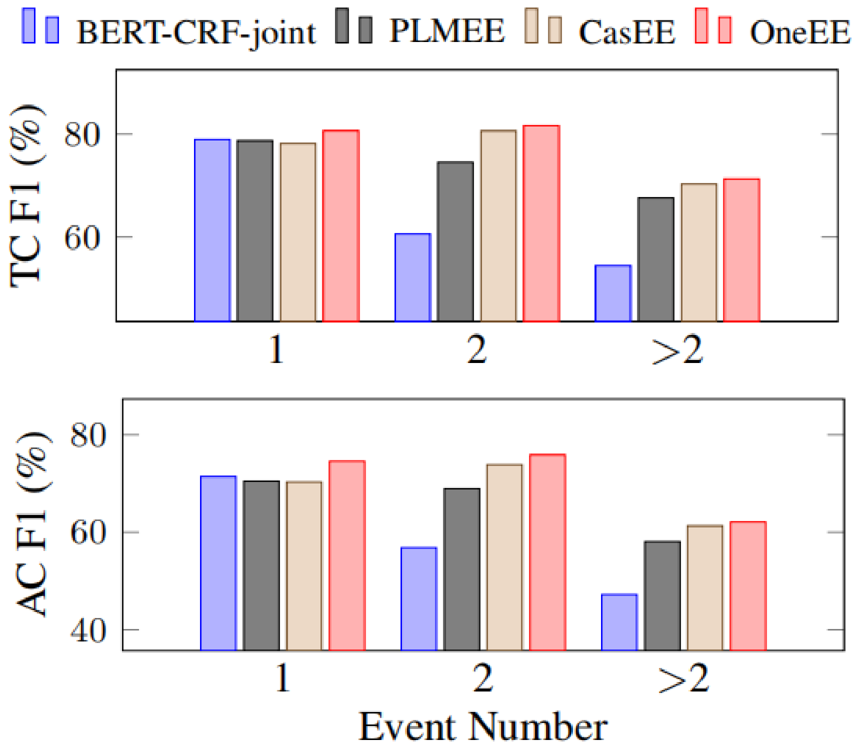


Figure 9: Results of different event numbers on FewFC.

Thanks