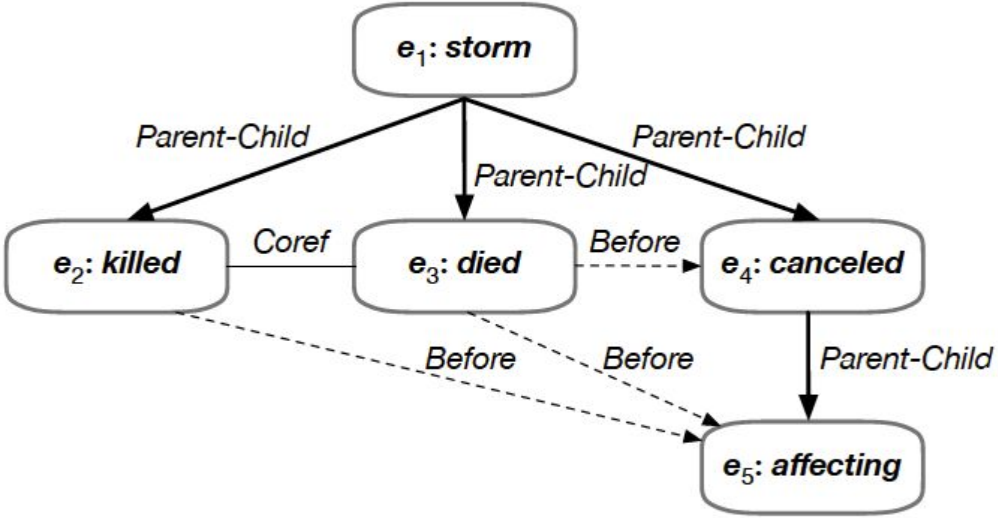
Joint Constrained Learning for Event-Event Relation Extraction

Haoyu Wang, Muhao Chen, Hongming Zhang, Dan Roth

EMNLP 2020

Use case

On Tuesday, there was a typhoon-strength (e_1:***storm***) in Japan. One man got (e_2:***killed***) and thousands of people were left stranded. Police said an 81-year-old man (e_3:***died***) in central Toyama when the wind blew over a shed, trapping him underneath. Later this afternoon, with the agency warning of possible tornadoes, Japan Airlines (e_4:***canceled***) 230 domestic flights, (e_5:***affecting***) 31,600 passengers.



Events are expressed at different granularities and have complex structures

Motivation

This paper aims to induce the complex of event relations to improve the understanding of text:

- Granularity
- Temporal order

This is done by regularizing:

- Annotation consistency
- Symmetric consistency
- Conjunction consistency

Notation

Document: $D = [t_1, \cdots, e_1, \cdots, e_2, \cdots, t_n]$

Events: $\mathcal{E}_D = \{e_1, e_2, \cdots, e_k\}$

Event relations: $\mathcal{R} = \mathcal{R}_T \cup \mathcal{R}_H$.

Temporal relations $\mathcal{R}_T$ BEFORE, AFTER, EQUAL, and VAGUE

Granularity relations $\mathcal{R}_H$ PARENT-CHILD, CHILD-PARENT, COREF and NOREL

Architecture

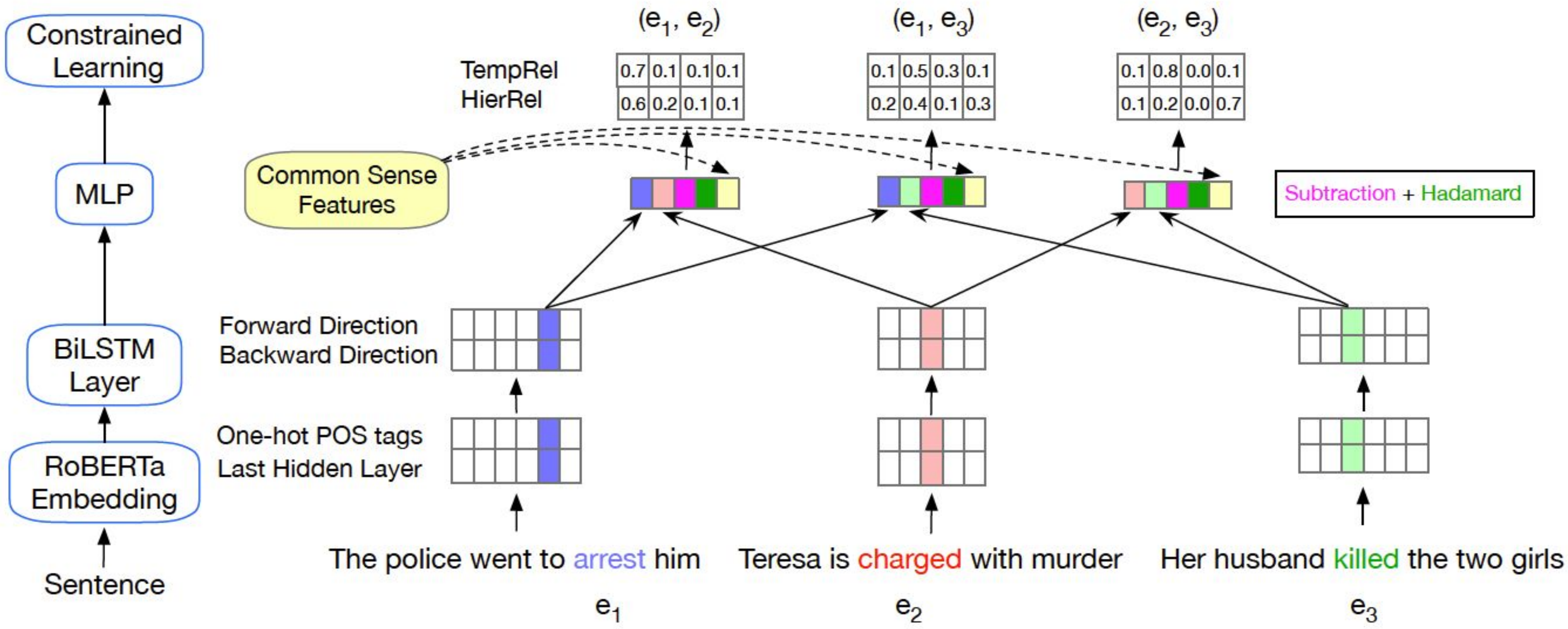


Figure 2: Model architecture. The model incorporates contextual features and commonsense knowledge to represent event pairs (§3.2). The joint learning enforces logical consistency on TempRel and subevent relations (§3.3).

Event pair representation

- Contextualized representation
  - Concatenation,
  - Element-wise product
  - Subtraction
- Commonsense knowledges
  - Extract from ConceptNet
    - 30K pairs and their relations
    - 30K corrupted pairs
  - TemProb
    - Provides temporal order of events
  - MLP models are trained on this and then fixed to generate **common sense features**

Annotation consistency

If labels are provided, the model should predict what is given in the annotation

$$L_A = \sum_{e_1, e_2 \in \mathcal{E}_D} -w_r \log r_{(e_1, e_2)},$$

Symmetric consistency

Given any pair of events, the converse relation holds

$$\bigwedge_{e_1, e_2 \in \mathcal{E}_D,\ \alpha \in \mathcal{R}_S} \alpha(e_1, e_2) \leftrightarrow \bar{\alpha}(e_2, e_1),$$

$$L_S = \sum_{e_1, e_2 \in \mathcal{E}, \alpha \in \mathcal{R}_S} |\log \alpha_{(e_1, e_2)} - \log \bar{\alpha}_{(e_2, e_1)}|.$$

Conjunction Consistency

Given $(e_1, e_2), (e_2, e_3)$ and (e_1, e_3) there are some rules mandate the relations between these pairs

$$\bigwedge_{\substack{e_1, e_2, e_3 \in \mathcal{E}_D \\ \alpha, \beta \in \mathcal{R},\ \gamma \in \mathrm{De}(\alpha, \beta)}} \alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \gamma(e_1, e_3).$$

$$\bigwedge_{\substack{e_1, e_2, e_3 \in \mathcal{E}_D \\ \alpha, \beta \in \mathcal{R},\ \delta \notin \mathrm{De}(\alpha, \beta)}} \alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \neg\delta(e_1, e_3).$$

Conjunction Consistency

α \ β	PC	CP	CR	NR	BF	AF	EQ	VG
PC	PC, ¬AF	–	PC, ¬AF	¬CP, ¬CR	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	–
CP	–	CP, ¬BF	CP, ¬BF	¬PC, ¬CR	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	–
CR	PC, ¬AF	CP, ¬BF	CR, EQ	NR	BF , ¬CP, ¬CR	AF, ¬PC, ¬CR	EQ	VG
NR	¬CP, ¬CR	¬PC, ¬CR	NR	–	–	–	–	–
BF	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	¬AF, ¬EQ
AF	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	–	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	¬BF , ¬EQ
EQ	¬AF	¬BF	EQ	–	BF , ¬CP, ¬CR	AF, ¬PC, ¬CR	EQ	VG, ¬CR
VG	–	–	VG, ¬CR	–	¬AF, ¬EQ	¬BF , ¬EQ	VG	–

Table 1: The induction table for conjunctive constraints on temporal and subevent relations. Given the relations $\alpha(e_1, e_2)$ in the left-most column and $\beta(e_2, e_3)$ in the top row, each entry in the table includes all the relations and negations that can be deduced from their conjunction for e_1 and e_3, i.e. $\mathrm{De}(\alpha, \beta)$. The abbreviations PC, CP, CR, NR, BF, AF, EQ and VG denote PARENT-CHILD, CHILD-PARENT, COREF, NOREL, BEFORE, AFTER, EQUAL and VAGUE, respectively. Vertical relations are in black, and TempRel are in blue. "–" denotes no constraints.

$$L_C = \sum_{\substack{e_1,e_2,e_3 \in \mathcal{E}_D,\\ \alpha,\beta \in \mathcal{R}, \gamma \in \mathrm{De}(\alpha,\beta)}} |L_{t_1}| + \sum_{\substack{e_1,e_2,e_3 \in \mathcal{E}_D,\\ \alpha,\beta \in \mathcal{R}, \delta \notin \mathrm{De}(\alpha,\beta)}} |L_{t_2}|$$

$$L_{t_1} = \log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log \gamma_{(e_1,e_3)}$$
$$L_{t_2} = \log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log(1 - \delta_{(e_1,e_3)})$$

Results

Model	P	R	F_1
CogCompTime (Ning et al., 2018c)	0.616	0.725	0.666
Perceptron (Ning et al., 2018b)	0.660	0.723	0.690
BiLSTM+MAP (Han et al., 2019b)	-	-	0.755
LSTM+CSE+ILP (Ning et al., 2019)	0.713	0.821	0.763
Joint Constrained Learning (ours)	**0.734**	**0.850**	**0.788**

Table 2: TempRel extraction results on MATRES. Precision and recall are not reported by (Han et al., 2019b).

	F_1 score		
Model	PC	CP	Avg.
StructLR (Glavaš et al., 2014)	0.522	**0.634**	0.577
TacoLM (Zhou et al., 2020a)	0.485	0.494	0.489
Joint Constrained Learning (ours)	**0.625**	0.564	**0.595**

Table 4: Subevent relation extraction results on HiEve. PC, CP and Avg. respectively denote Parent-Child, Child-Parent and their micro-average.