A Novel Cascade Binary Tagging Framework for Relational Triple Extraction

Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, Yi Chang

Accepted by ACL2020

Overview

- Most previous work addresses relation extraction (RE) by learning a mapping from a pair of entities (i.e., subject and object) to a relation:

$$f(s, o) \rightarrow r$$

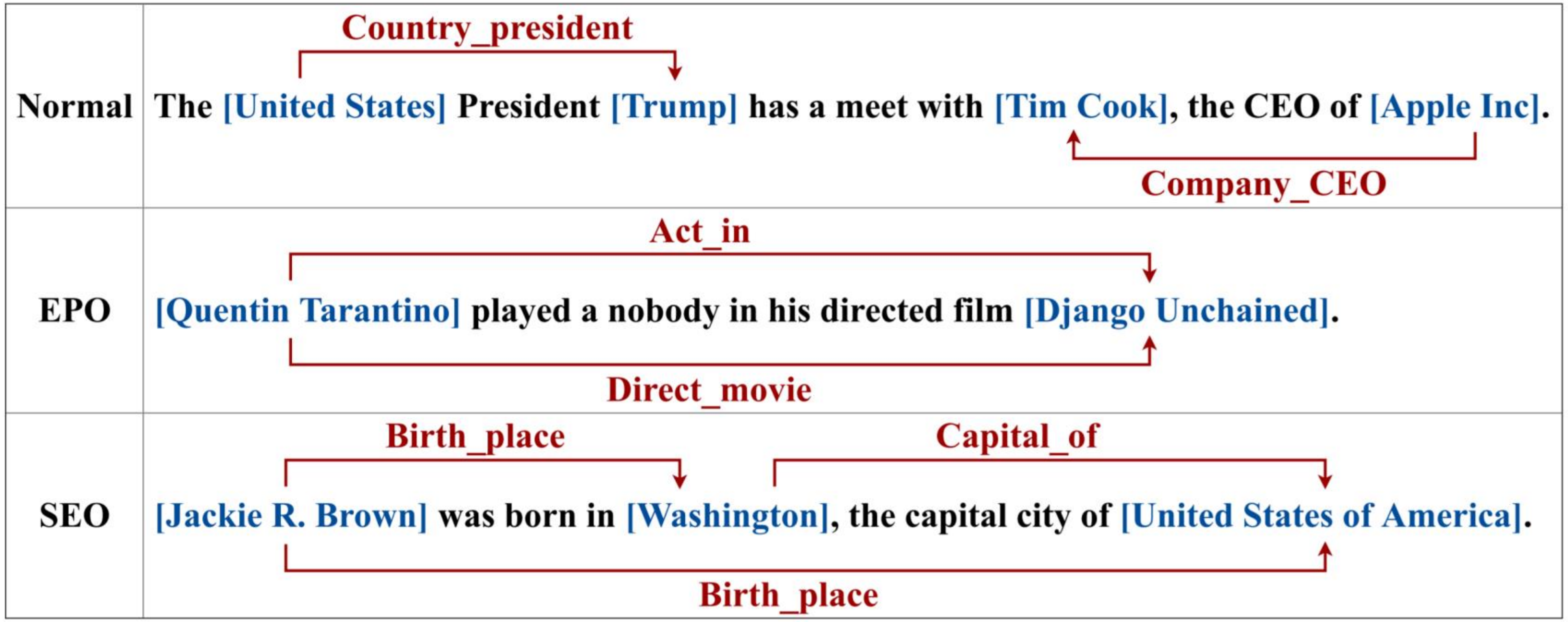
- This work treats RE as a learning problem of functions mapping subjects to objects:

$$f_r(s) \rightarrow o$$

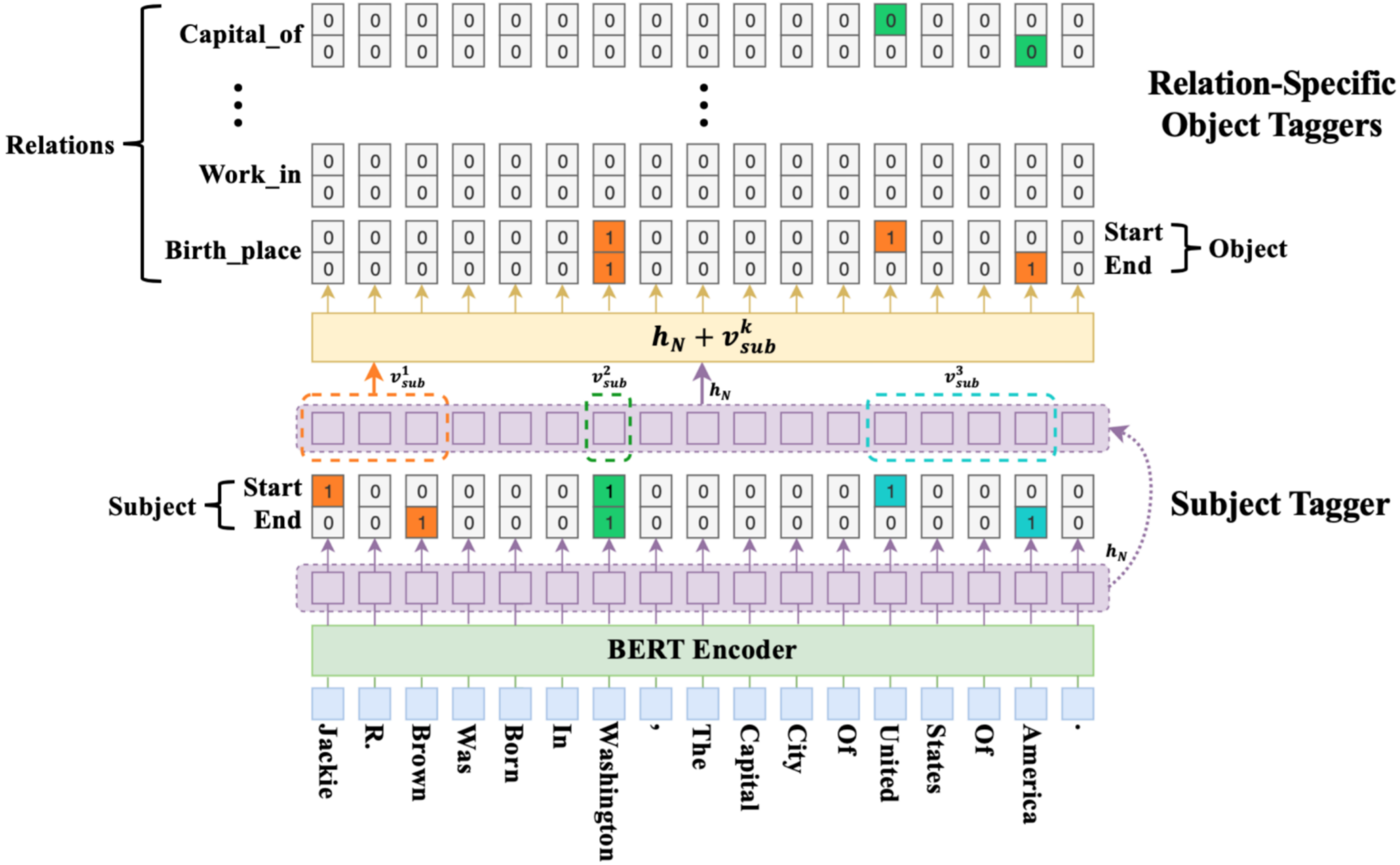
=> Naturally solve the problem of overlapping relations in RE.

Overlapping problem in RE

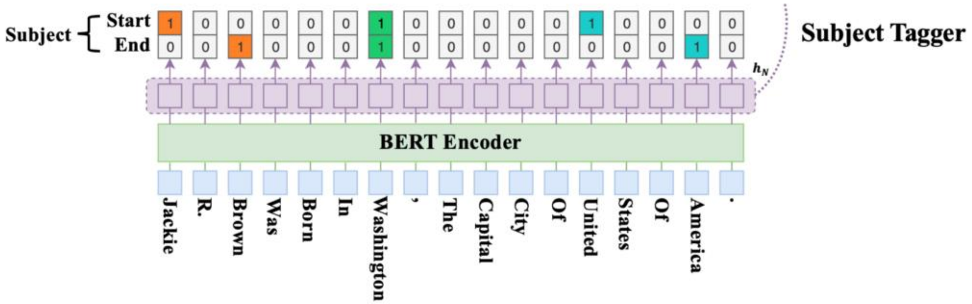
- There are two types of overlapping in RE:
  - Entity Pair Overlap (EPO).
  - Single Entity Overlap (SEO).



Model: Overview



Model: Subject Tagger



- Given a sentence j, it uses BERT as the encoder to obtain $\mathbf{x}_j$
- Employs two classifiers for identifying "start"s and "end"s of subjects.

$$p_i^{start\_s} = \sigma(\mathbf{W}_{start}\mathbf{x}_i + \mathbf{b}_{start})$$
$$p_i^{end\_s} = \sigma(\mathbf{W}_{end}\mathbf{x}_i + \mathbf{b}_{end})$$

- Training objective: maximizes the log likelihood of the groundtruth subject spans:

$$\sum_{s \in T_j} \log p_\theta(s|\mathbf{x}_j)$$

where: $p_\theta(s|\mathbf{x})$

$$= \prod_{t \in \{start\_s, end\_s\}} \prod_{i=1}^{L} \left(p_i^t\right)^{\mathbf{I}\{y_i^t = 1\}} \left(1 - p_i^t\right)^{\mathbf{I}\{y_i^t = 0\}}$$

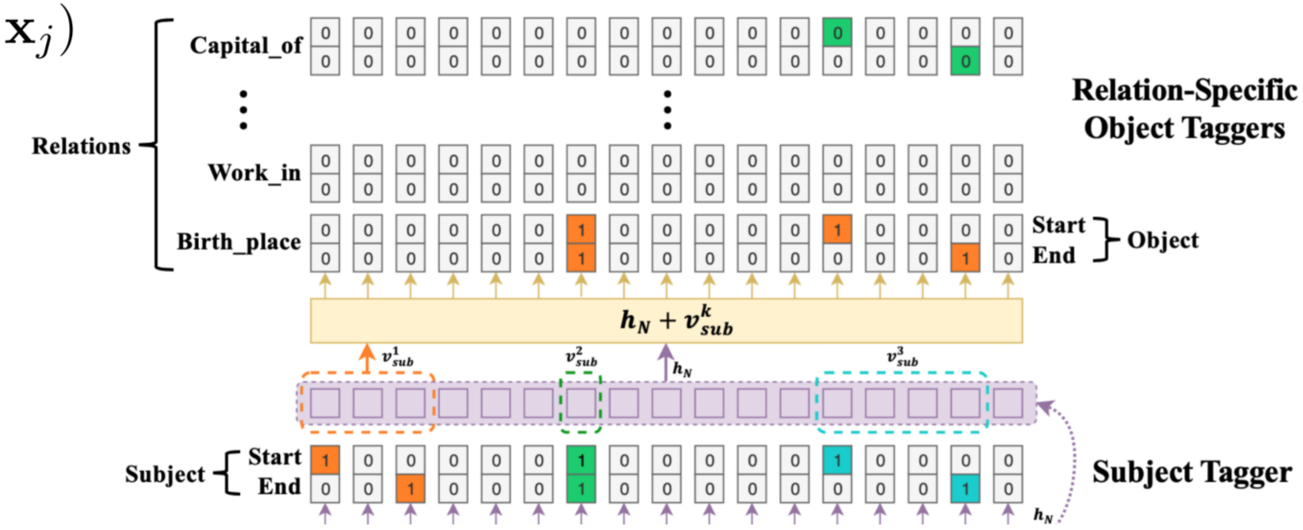
Model: Relation-specific Object Taggers

- For each relation, takes the representation $\mathbf{v}_{sub}^{k}$ of the detected subject into account.

$$p_i^{start\_o} = \sigma(\mathbf{W}_{start}^{r}(\mathbf{x}_i + \mathbf{v}_{sub}^{k}) + \mathbf{b}_{start}^{r})$$
$$p_i^{end\_o} = \sigma(\mathbf{W}_{end}^{r}(\mathbf{x}_i + \mathbf{v}_{sub}^{k}) + \mathbf{b}_{end}^{r})$$

- Training objective: maximizes the following log likelihood:

$$\sum_{r \in T_j|s} \log p_{\phi_r}(o|s, \mathbf{x}_j) + \sum_{r \in R \backslash T_j|s} \log p_{\phi_r}(o_{\varnothing}|s, \mathbf{x}_j)$$



Results

- Datasets:

  - Highly-overlapping RE datasets: NYT and WebNLG.

Category	NYT		WebNLG	
	Train	Test	Train	Test
Normal	37013	3266	1596	246
EPO	9782	978	227	26
SEO	14735	1297	3406	457
ALL	56195	5000	5019	703

  - Also on ACE04, NYT10-HRL, NYT11-HRL, WikiKBP.

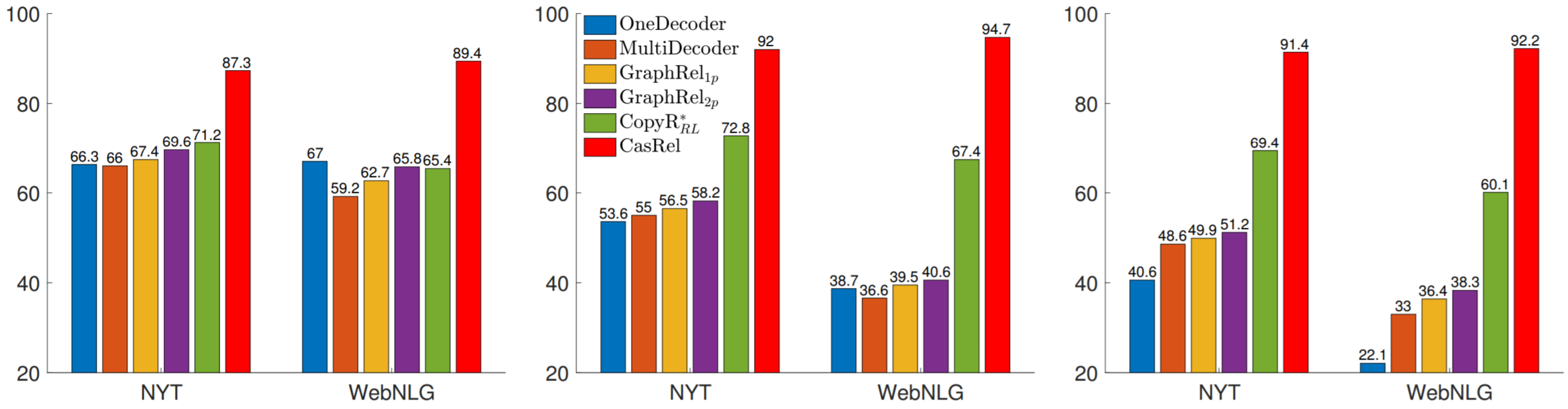
Results

- Achieves 17.5% and 30.2% improvements in F1-score over the best state-of-the-art method (Zeng et al., 2019).

Method	NYT			WebNLG		
	Prec.	*Rec.*	*F1*	*Prec.*	*Rec.*	*F1*
NovelTagging (Zheng et al., 2017)	62.4	31.7	42.0	52.5	19.3	28.3
CopyR$_{OneDecoder}$ (Zeng et al., 2018)	59.4	53.1	56.0	32.2	28.9	30.5
CopyR$_{MultiDecoder}$ (Zeng et al., 2018)	61.0	56.6	58.7	37.7	36.4	37.1
GraphRel$_{1p}$ (Fu et al., 2019)	62.9	57.3	60.0	42.3	39.2	40.7
GraphRel$_{2p}$ (Fu et al., 2019)	63.9	60.0	61.9	44.7	41.1	42.9
CopyR$_{RL}$ (Zeng et al., 2019)	77.9	67.2	72.1	63.3	59.9	61.6
CopyR$^{*}_{RL}$	72.8	69.4	71.1	60.9	61.1	61.0
CasRel$_{random}$	81.5	75.7	78.5	84.7	79.5	82.0
CasRel$_{LSTM}$	84.2	83.0	83.6	86.9	80.6	83.7
CasRel	**89.7**	**89.5**	**89.6**	**93.4**	**90.1**	**91.8**

Results

- Performance comparison in different types of overlapping.



(a) F1 of Normal Class (b) F1 of EntityPairOverlap Class (c) F1 of SingleEntityOverlap Class

Figure 3: F1-score of extracting relational triples from sentences with different overlapping pattern.

Results

- Results on low-overlapping RE datasets:

Method	Partial Match									Exact Match		
	ACE04			NYT10-HRL			NYT11-HRL			Wiki-KBP		
	Prec.	*Rec.*	*F1*	*Prec.*	*Rec.*	*F1*	*Prec.*	*Rec.*	*F1*	*Prec.*	*Rec.*	*F1*
Chan and Roth (2011)	42.9	38.9	40.8	–	–	–	–	–	–	–	–	–
MultiR (Hoffmann et al., 2011)	–	–	–	–	–	–	32.8	30.6	31.7	30.1	53.0	38.0
DS-Joint (Li and Ji, 2014)	**64.7**	38.5	48.3	–	–	–	–	–	–	–	–	–
FCM (Gormley et al., 2015)	–	–	–	–	–	–	43.2	29.4	35.0	–	–	–
SPTree (Miwa and Bansal, 2016)	–	–	–	49.2	55.7	52.2	52.2	54.1	53.1	–	–	–
CoType (Ren et al., 2017)	–	–	–	–	–	–	48.6	38.6	43.0	31.1	**53.7**	38.8
Katiyar and Cardie (2017)	50.2	**48.8**	49.3	–	–	–	–	–	–	–	–	–
NovelTagging (Zheng et al., 2017)	–	–	–	59.3	38.1	46.4	46.9	48.9	47.9	**53.6**	30.3	38.7
ReHession (Liu et al., 2017)	–	–	–	–	–	–	–	–	–	36.7	49.3	42.1
CopyR (Zeng et al., 2018)	–	–	–	56.9	45.2	50.4	34.7	53.4	42.1	–	–	–
HRL (Takanobu et al., 2019)	–	–	–	71.4	58.6	64.4	**53.8**	53.8	53.8	–	–	–
PA-LSTM-CRF (Dai et al., 2019)	–	–	–	–	–	–	–	–	–	51.1	39.3	44.4
CasRel	57.2	47.6	**52.0**	**77.7**	**68.8**	**73.0**	50.1	**58.4**	**53.9**	49.8	42.7	**45.9**