OneRel: Joint Entity and Relation Extraction with One Module in One Step

Yu-Ming Shang[1], Heyan Huang[1,2], Xian-Ling Mao[1*]

[1]School of Computer Science & Technology, Beijing Institute of Technology, Beijing, China
[2]Beijing Engineering Research Center of High Volume Language Information Processing
and Cloud Computing Applications, Beijing, China
{ymshang, hhy63, maoxl}@bit.edu.cn

AAAI 2022
Code: https://github.com/ssnvxia/OneRel

Reported by Nengqiang Xiang

Introduction

- Existing joint methods suffer from the problems of cascading errors and redundant information.

- To address these issues, in this paper, we propose a novel joint entity and relation extraction model, named OneRel, which casts joint extraction as a fine-grained triple classification problem.

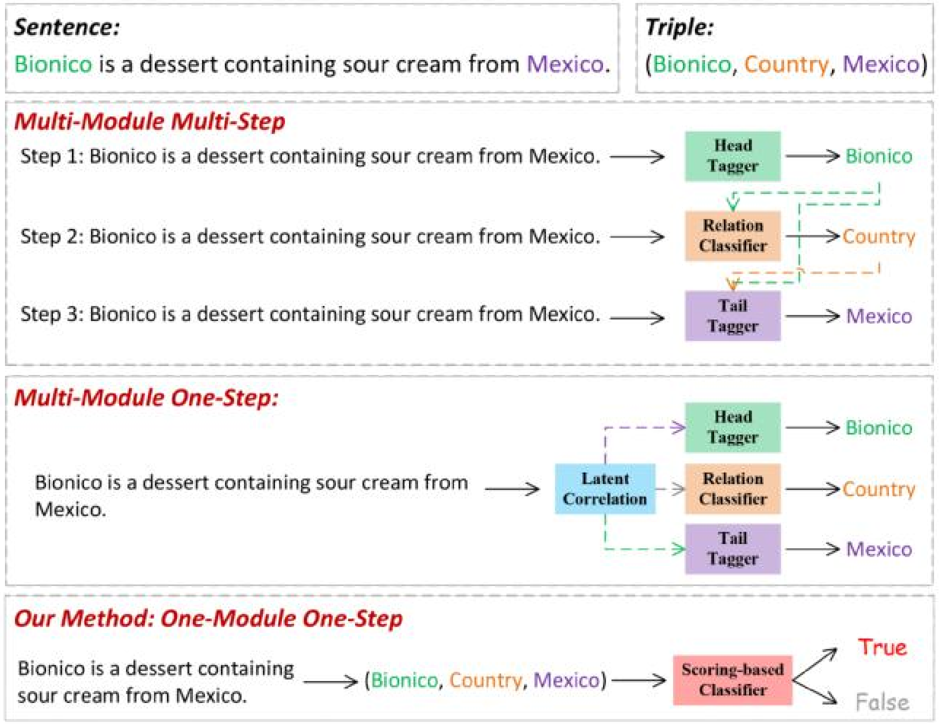
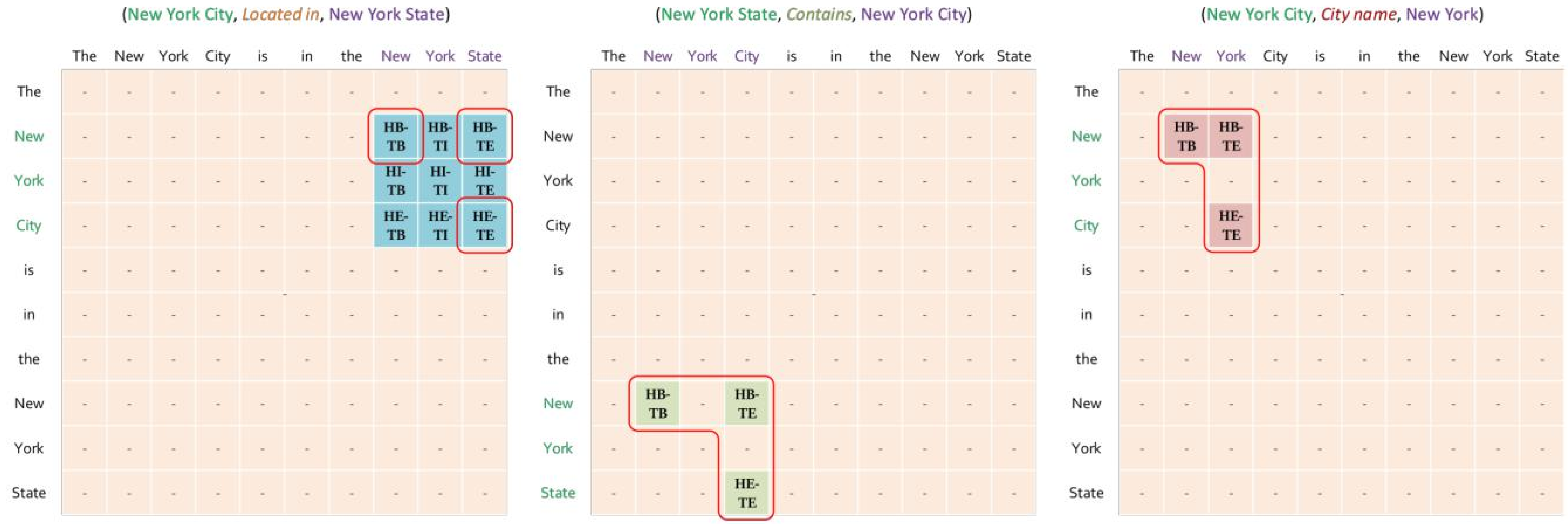


Figure 1: The extraction processes of existing approaches and our method. The dotted arrow indicates the dependencies between triple elements. Note that there are various extraction paradigms in *multi-module multi-step* approaches, e.g., $(h, t) \rightarrow r$, $h \rightarrow r \rightarrow t$, $r \rightarrow (h, t)$. We just use $h \rightarrow r \rightarrow t$ to illustrate the shortcomings of this kind of methods.

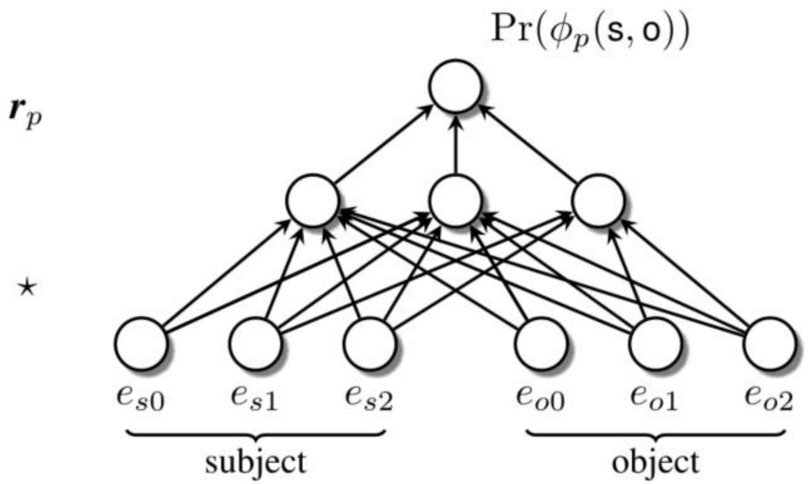
Method



(a) $M_{r=Located\ in}$ (b) $M_{r=Contains}$ (c) $M_{r=City\ name}$

Figure 2: Examples of the Rel-Spec Horns Tagging. For the convenience to explanation, we illustrate the sub-matrix with a given relation, e.g., `Located in`. So, the matrix M degenerates into two dimensions, where the rows represent head entities and the columns represent tail entities.

Method



(b) HOLE

$$f_r(h,t) = \boldsymbol{r}^T(\boldsymbol{h} \star \boldsymbol{t}), \tag{2}$$

$$\boldsymbol{h} \star \boldsymbol{t} = \phi(\boldsymbol{W}[\boldsymbol{h};\boldsymbol{t}]^T + \boldsymbol{b}), \tag{3}$$

$$\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_L\} = BERT(\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_L\}), \tag{1}$$

$$\boldsymbol{v}_{(w_i,r_k,w_j)_{k=1}^K} = \boldsymbol{R}^T \phi(\mathtt{drop}(\boldsymbol{W}[\mathbf{e}_i;\boldsymbol{e}_j]^T + \boldsymbol{b})), \tag{4}$$

$$P(\mathrm{y}_{(w_i,r_k,w_j)}|\mathcal{S}) = Softmax(\boldsymbol{v}_{(w_i,r_k,w_j)}) \tag{5}$$

$$\mathcal{L}_{triple} = -\frac{1}{L \times K \times L} \times \sum_{i=1}^{L}\sum_{k=1}^{K}\sum_{j=1}^{L} \log P(\mathrm{y}_{(w_i,r_k,w_j)} = g_{(w_i,r_k,w_j)}|\mathcal{S}), \tag{6}$$

Experiments

Category	Dataset				Details of Test Set									
	Train	Valid	Test	Relations	Normal	SEO	EPO	HTO	N=1	N=2	N=3	N=4	N>5	Triples
NYT*	56,195	4,999	5,000	24	3,266	1,297	978	45	3,244	1,045	312	291	108	8,110
WebNLG*	5,019	500	703	171	245	457	26	84	266	171	131	90	45	1,591
NYT	56,195	5,000	5,000	24	3,222	1,273	969	117	3,240	1,047	314	290	109	8,120
WebNLG	5,019	500	703	216	239	448	6	85	256	175	138	93	41	1,607

Table 1: Statistics of datasets. N is the number of triples in a sentence.

Experiments

Model	Partial Match						Exact Match					
	NYT*			WebNLG*			NYT			WebNLG		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
GraphRel (Fu, Li, and Ma 2019)	63.9	60.0	61.9	44.7	41.1	42.9	-	-	-	-	-	-
RSAN (Yuan et al. 2020)	-	-	-	-	-	-	85.7	83.6	84.6	80.5	83.8	82.1
MHSA (Liu et al. 2020)	88.1	78.5	83.0	89.5	86.0	87.7	-	-	-	-	-	-
CasRel (Wei et al. 2020)	89.7	89.5	89.6	93.4	90.1	91.8	-	-	-	-	-	-
TPLinker (Wang et al. 2020)	91.3	92.5	91.9	91.8	92.0	91.9	91.4	92.6	92.0	88.9	84.5	86.7
SPN (Sui et al. 2020)	93.3	91.7	92.5	93.1	93.6	93.4	92.5	92.2	92.3	-	-	-
CGT (Ye et al. 2021)	**94.7**	84.2	89.1	92.9	75.6	83.4	-	-	-	-	-	-
CasDE (Ma, Ren, and Zhang 2021)	90.2	90.9	90.5	90.3	91.5	90.9	89.9	91.4	90.6	88.0	88.9	88.4
RIFRE (Zhao et al. 2021a)	93.6	90.5	92.0	93.3	92.0	92.6	-	-	-	-	-	-
PRGC (Zheng et al. 2021)	93.3	91.9	92.6	94.0	92.1	93.0	**93.5**	91.9	92.7	89.9	87.2	88.5
OneRel^{-}	91.3	90.5	90.9	93.8	91.4	92.6	91.1	90.4	90.8	90.5	88.2	89.4
OneRel	92.8	**92.9**	**92.8**	**94.1**	**94.4**	**94.3**	93.2	**92.6**	**92.9**	**91.8**	**90.3**	**91.0**

Table 2: Precision(%), Recall (%) and F1-score (%) of our proposed OneRel and baselines.

Experiments

Model	NYT*									WebNLG*								
	Normal	EPO	SEO	HTO	N=1	N=2	N=3	N=4	N≥5	Normal	EPO	SEO	HTO	N=1	N=2	N=3	N=4	N≥5
CasRel	87.3	92.0	91.4	77.0§	88.2	90.3	91.9	94.2	83.7	89.4	94.7	92.2	90.4§	89.3	90.8	94.2	92.4	90.9
TPLinker	90.1	94.0	93.4	90.1§	90.0	92.8	93.1	96.1	90.0	87.9	95.3	92.5	86.0§	88.0	90.1	94.6	93.3	91.6
SPN	90.8	94.1	94.0	-	90.9	93.4	**94.2**	95.5	90.6	-	-	-	-	89.5	91.3	**96.4**	94.7	93.8
PRGC	**91.0**	94.5	94.0	81.8	**91.1**	93.0	93.5	95.5	93.0	90.4	**95.9**	93.6	94.6	89.9	91.6	95.0	94.8	92.8
OneRel	90.6	**95.1**	**94.8**	**90.8**	90.5	**93.4**	93.9	**96.5**	**94.2**	**91.9**	95.4	**94.7**	**94.9**	**91.4**	**93.0**	95.9	**95.7**	**94.5**

Table 3: F1-score (%) on sentences with different overlapping patterns and different triple numbers. § marks the results reported by (Zheng et al. 2021).

Experiments

Model	Element	NYT*			WebNLG*		
		Prec.	Rec.	F1	Prec.	Rec.	F1
CasRel	(h,t)	89.2	90.1	89.7	95.3	91.7	93.5
	r	96.0	93.8	94.9	96.6	91.5	94.0
	(h,r,t)	89.7	89.5	89.6	93.4	90.1	91.8
SPN	(h,t)	93.2	92.7	92.9	95.0	95.4	95.2
	r	96.3	95.7	96.0	95.2	95.7	95.4
	(h,r,t)	93.3	91.7	92.5	93.1	93.6	93.4
PRGC	(h,t)	**94.0**	92.3	93.1	96.0	93.4	94.7
	r	95.3	96.3	95.8	92.8	96.2	94.5
	(h,r,t)	**93.3**	91.9	92.6	94.0	92.1	93.0
OneRel	(h,t)	93.3	**93.4**	**93.3**	**96.2**	**96.5**	**96.3**
	r	**96.7**	**96.9**	**96.8**	**96.7**	**97.0**	**96.8**
	(h,r,t)	92.8	**92.9**	**92.8**	**94.1**	**94.4**	**94.3**

Table 4: Results on triple elements. (h,t) denotes the entity pair and r means the relation.

Experiments

Dataset	Model	Training Time	Inference Time	F1
NYT*	TPLinker	1592	46.2	91.9
	OneRel	**1195**	**41.5**	**92.9**
WebNLG*	TPLinker	599	**40.1**	91.9
	OneRel	**88**	45	**94.3**

Table 5: Comparison of the model efficiency. Training Time (s) means the time required to train one epoch, Inference Time (ms) is the time to predict triples of one sentence.

Experiments

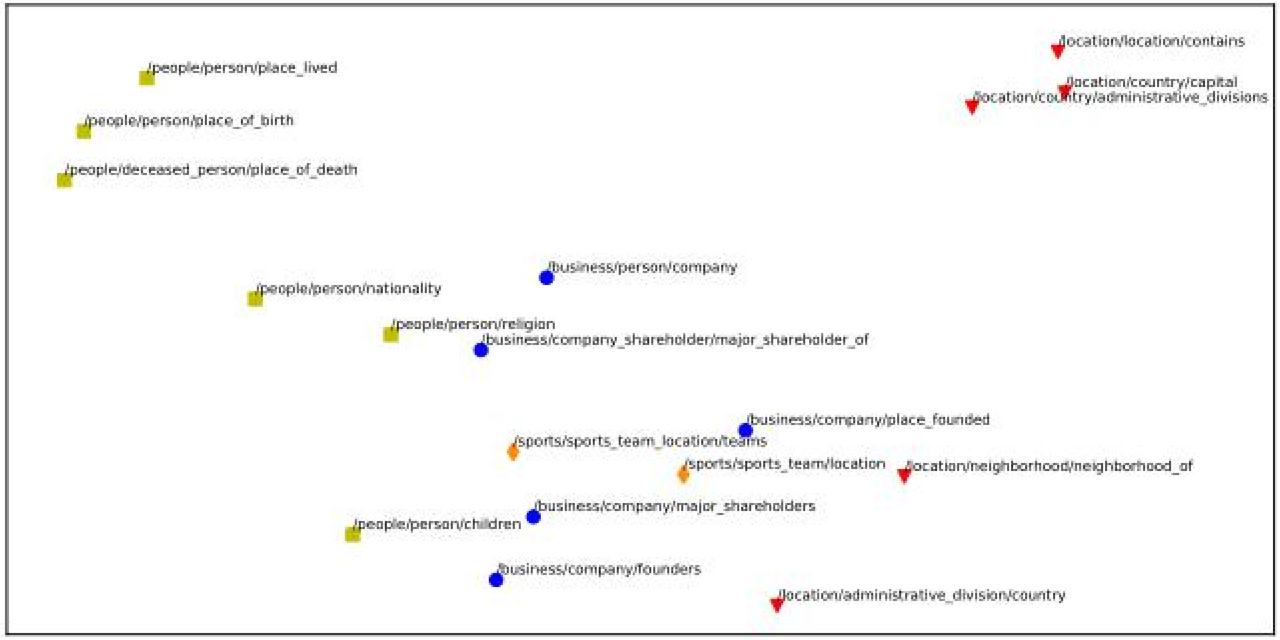


Figure 3: (Best viewed in color and zoom in.) Visualization of relations on NYT dataset.

Thanks