A Simple but Effective Bidirectional Framework for Relational Triple Extraction

Feiliang Ren*†
Northeastern University
Shenyang, China

Longhui Zhang†
Northeastern University
Shenyang, China

Xiaofeng Zhao
Northeastern University
Shenyang, China

Shujuan Yin
Northeastern University
Shenyang, China

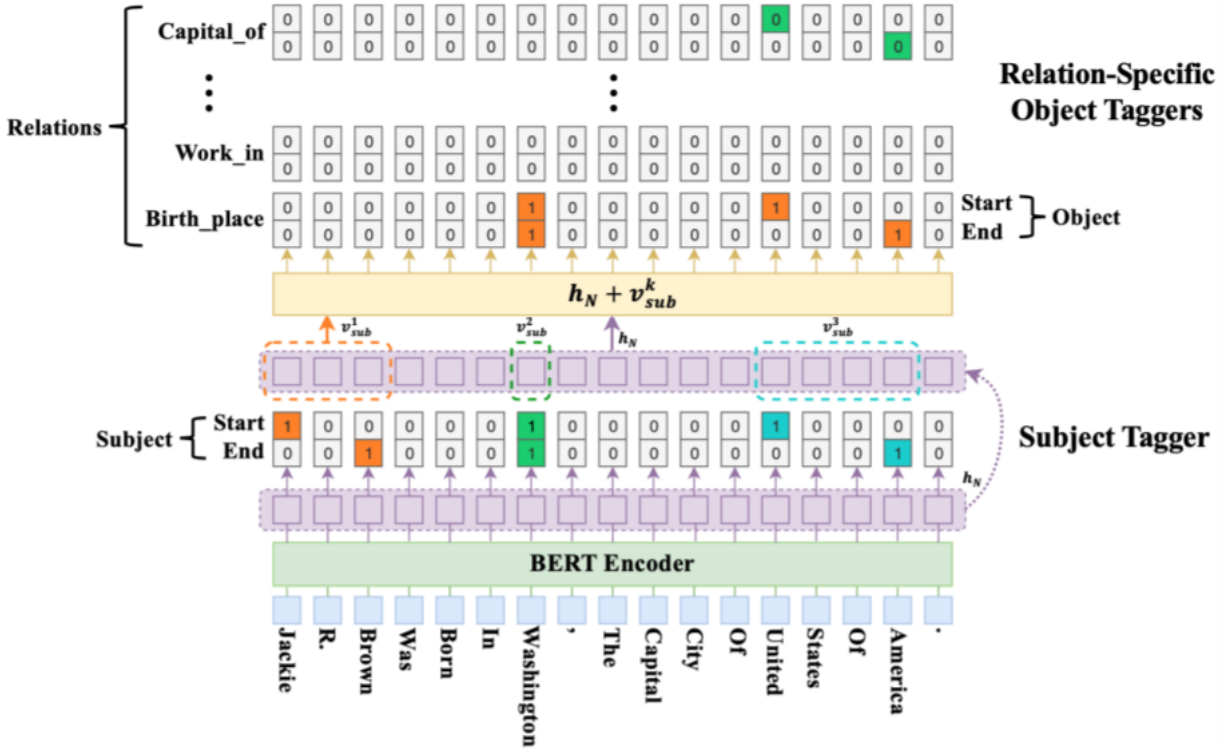
Shilei Liu
Northeastern University
Shenyang, China

Bochao Li
Northeastern University
Shenyang, China

(WSDM 2022)

Introduction

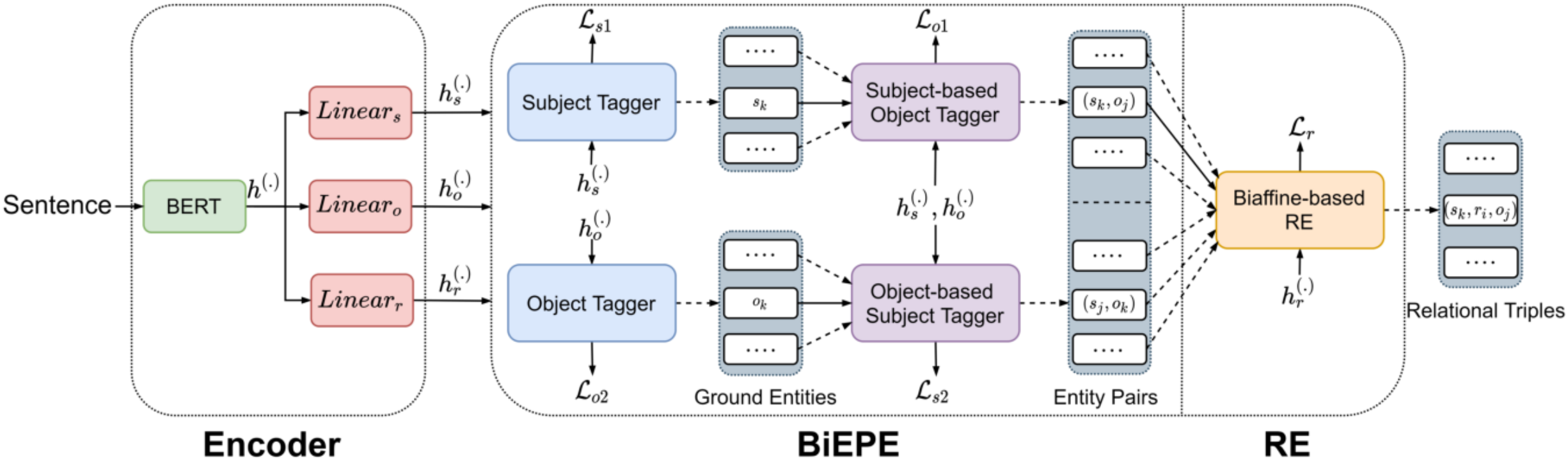
$$f_r(s) \rightarrow o$$



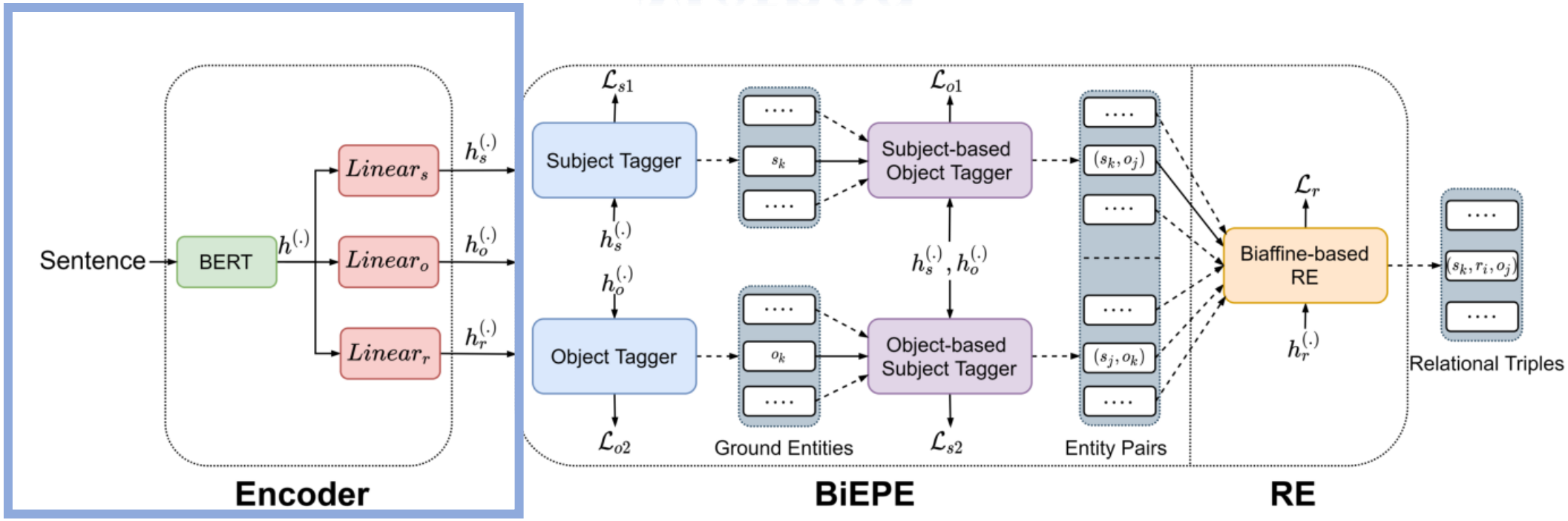
- bidirectional extraction framework
  - $subject \rightarrow object \rightarrow relation\ (s2o)$
  - $object \rightarrow subject \rightarrow relation\ (o2s)$
  - biaffine model
- the share-aware learning mechanism

CasRel: A Novel Cascade Binary Tagging Framework for Relational Triple Extraction, ACL 2020

Method


Sentence
BERT
$h^{(.)}$
$Linear_s$
$Linear_o$
$Linear_r$
$h_s^{(.)}$
$h_o^{(.)}$
$h_r^{(.)}$
Subject Tagger
Object Tagger
$\mathcal{L}_{s1}$
$\mathcal{L}_{o2}$
s_k
o_k
Ground Entities
Subject-based Object Tagger
Object-based Subject Tagger
$\mathcal{L}_{o1}$
$\mathcal{L}_{s2}$
$h_s^{(.)}, h_o^{(.)}$
(s_k, o_j)
(s_j, o_k)
Entity Pairs
Biaffine-based RE
$\mathcal{L}_r$
(s_k, r_i, o_j)
Relational Triples
Encoder
BiEPE
RE

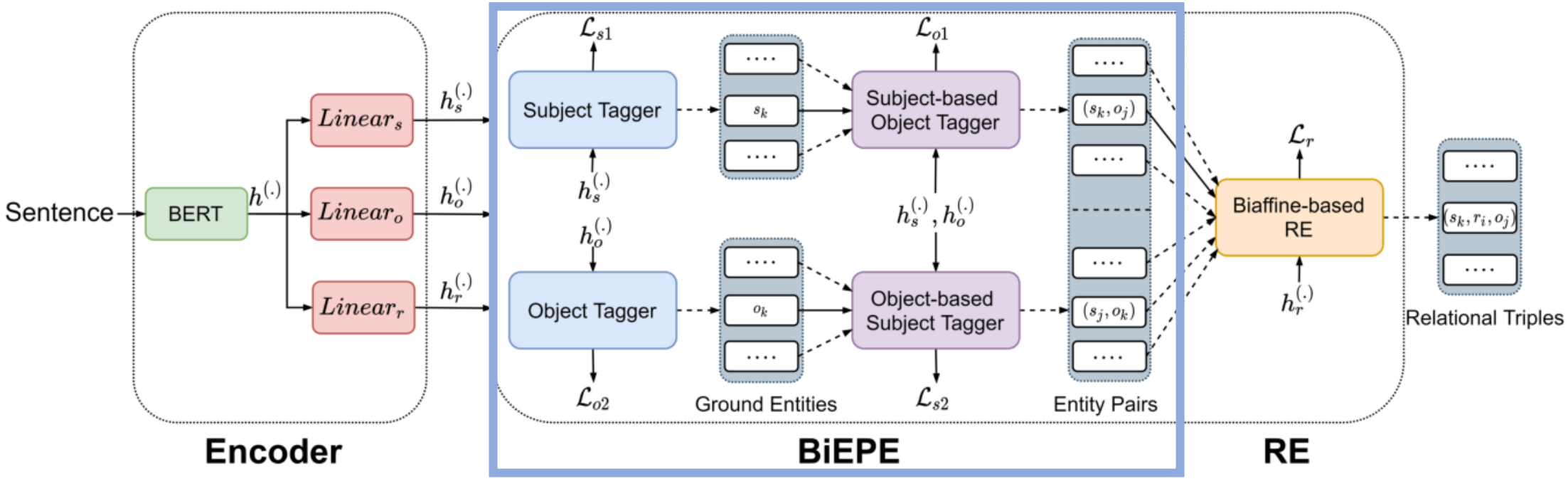
Method



$$
\begin{aligned}
\mathrm{h}_s^i &= \mathrm{W}_s \mathrm{h}^i + \mathrm{b}_s \\
\mathrm{h}_o^i &= \mathrm{W}_o \mathrm{h}^i + \mathrm{b}_o \\
\mathrm{h}_r^i &= \mathrm{W}_r \mathrm{h}^i + \mathrm{b}_r
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\mathrm{h}_s^i &= \mathrm{h}_s^i + \mathrm{h}_o^{cls} \\
\mathrm{h}_o^i &= \mathrm{h}_o^i + \mathrm{h}_s^{cls}
\end{aligned} \tag{2}
$$

Method



$$p_s^{i,start} = \sigma\left(\mathrm{W}_s^{start}\mathrm{h}_s^i + \mathrm{b}_s^{start}\right)$$
$$p_s^{i,end} = \sigma\left(\mathrm{W}_s^{end}\mathrm{h}_s^i + \mathrm{b}_s^{end}\right)$$

(3)

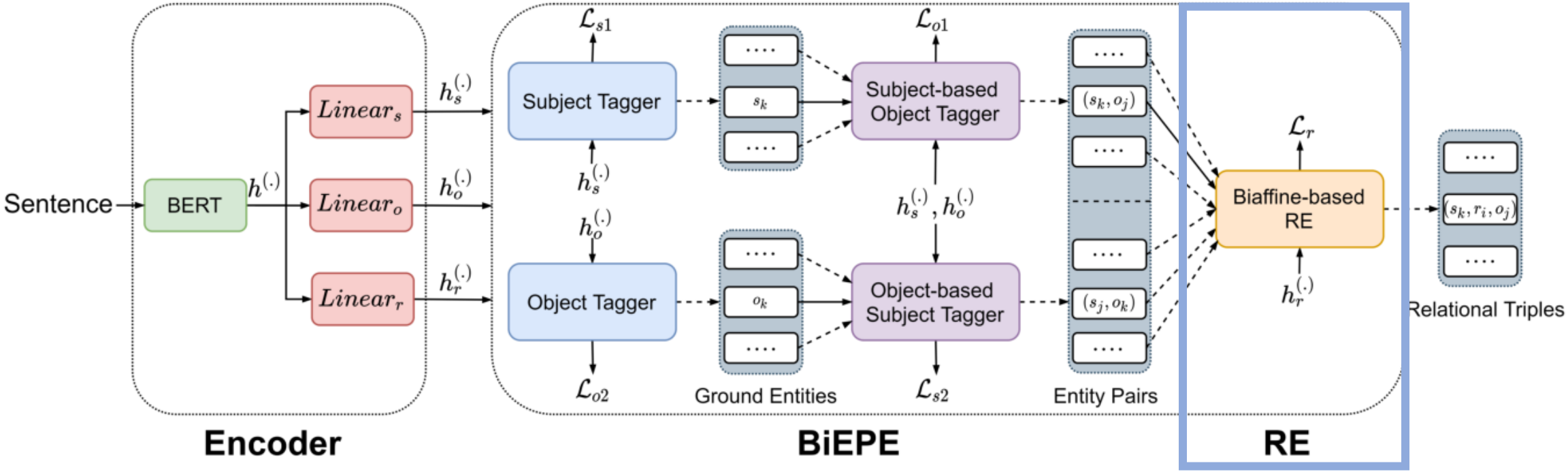
$$\mathrm{v}_s^{s\_k} = \mathrm{maxpool}\left(\mathrm{h}_s^{s\_k\_start}, \ldots, \mathrm{h}_s^{s\_k\_end}\right)$$
$$p_o^{i,start} = \sigma\left(\mathrm{W}_o^{start}\left(\mathrm{h}_o^i \circ \mathrm{v}_s^{s\_k}\right) + \mathrm{b}_o^{start}\right)$$
$$p_o^{i,end} = \sigma\left(\mathrm{W}_o^{end}\left(\mathrm{h}_o^i \circ \mathrm{v}_s^{s\_k}\right) + \mathrm{b}_o^{end}\right)$$

(4)

$$\mathrm{ce}(p,t) = -\left[t\log p + (1-t)\log(1-p)\right]$$
$$\mathcal{L}_{s1} = \frac{1}{2\times l}\sum_{m\in\{\mathrm{start,end}\}}\sum_{i=1}^{l}\mathrm{ce}\left(p_s^{i,m}, t_s^{i,m}\right)$$
$$\mathcal{L}_{o1} = \frac{1}{2\times l}\sum_{m\in\{\mathrm{start,end}\}}\sum_{i=1}^{l}\mathrm{ce}\left(p_o^{i,m}, t_o^{i,m}\right)$$

(5)

Method

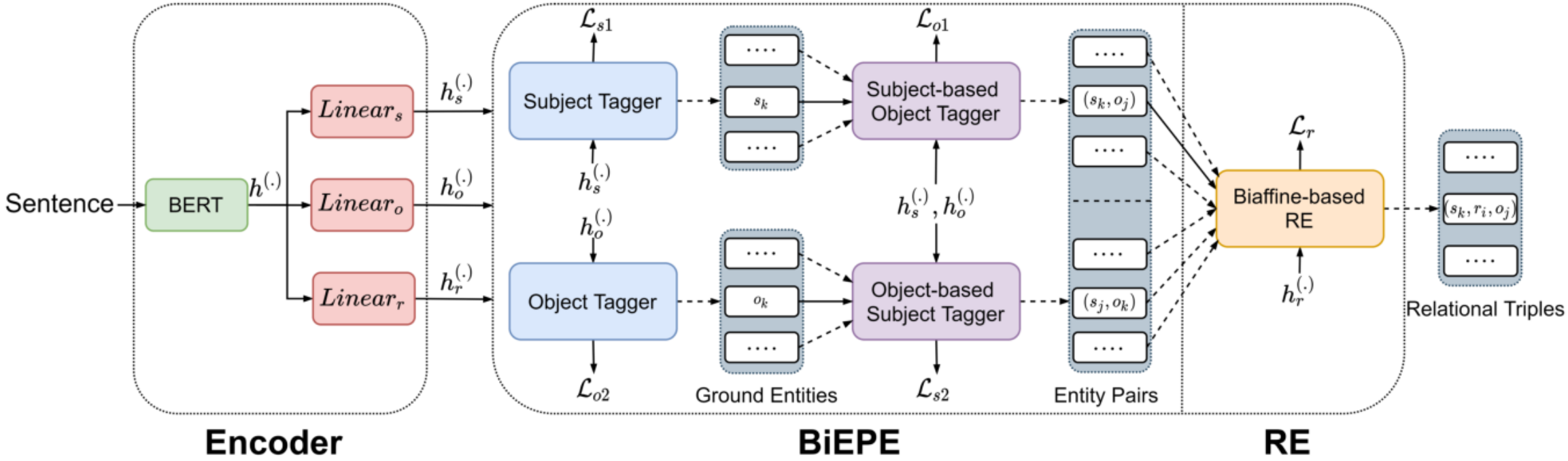


$$
\begin{aligned}
\mathrm{v}_r^{s\_k} &= \mathrm{maxpool}\left(\mathrm{h}_r^{s\_k\_start}, \ldots, \mathrm{h}_r^{s\_k\_end}\right) \\
\mathrm{v}_r^{o\_j} &= \mathrm{maxpool}\left(\mathrm{h}_r^{o\_j\_start}, \ldots, \mathrm{h}_r^{o\_j\_end}\right) \\
p_r^i &= \sigma\left(\begin{bmatrix} \mathrm{v}_r^{s\_k} \\ 1 \end{bmatrix}^{\top} \mathrm{W}_r^i \begin{bmatrix} \mathrm{v}_r^{o\_j} \\ 1 \end{bmatrix}\right)
\end{aligned} \tag{6}
$$

$$
\mathcal{L}_r = \frac{1}{|R|} \sum_{i=1}^{|R|} \mathrm{ce}\left(p_r^i, t_r^i\right) \tag{7}
$$

Method

the share-aware learning mechanism



$$\mathcal{L} = \mathcal{L}_{s1} + \mathcal{L}_{o1} + \mathcal{L}_{s2} + \mathcal{L}_{o2} + \mathcal{L}_r \qquad (8)$$

$$\xi_i = \begin{cases} \xi, & k_i = 1 \\ \frac{(1+\delta)}{f(k_i)} * \xi, & k_i > 1 \end{cases} \qquad (9)$$

Experiments

Category	NYT		WebNLG		NYT10		NYT11	
	Train	Test	Train	Test	Train	Test	Train	Test
Normal	37013	3266	1596	246	59396	2963	53395	368
EPO	9782	978	227	26	5376	715	2100	0
SEO	14735	1297	3406	457	8772	742	7365	1
ALL	56195	5000	5019	703	70339	4006	62648	369

Table 1: Statistics of datasets. *EPO* and *SEO* refer to the *entity pair overlapping* and *single entity overlapping* respectively [31]. Note a sentence can belong to both *EPO* and *SEO*.

Experiments

Model	Partial Match						Exact Match					
	NYT*			WebNLG*			NYT			WebNLG		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
ETL-Span [26]	84.9	72.3	78.1	84.0	91.5	87.6	85.5	71.7	78.0	84.3	82.0	83.1
WDec [16]	–	–	–	–	–	–	88.1	76.1	81.7	–	–	–
RSAN [27]	–	–	–	–	–	–	85.7	83.6	84.6	80.5	83.8	82.1
RIN [19]	87.2	87.3	87.3	87.6	87.0	87.3	83.9	85.5	84.7	77.3	76.8	77.0
CasRel$_{LSTM}$ [24]	84.2	83.0	83.6	86.9	80.6	83.7	–	–	–	–	–	–
PMEI$_{LSTM}$ [20]	88.7	86.8	87.8	88.7	87.6	88.1	84.5	84.0	84.2	78.8	77.7	78.2
TPLinker$_{LSTM}$ [23]	83.8	83.4	83.6	90.8	90.3	90.5	86.0	82.0	84.0	**91.9**	81.6	86.4
R-BPtrNet$_{LSTM}$‡ [3]	90.9	91.3	91.1	90.7	**94.6**	92.6	–	–	–	–	–	–
CGT$_{UniLM}$ [25]	**94.7**	84.2	89.1	92.9	75.6	83.4	–	–	–	–	–	–
CasRel$_{BERT}$ [24]	89.7	89.5	89.6	93.4	90.1	91.8	89.8*	88.2*	89.0*	88.3*	84.6*	86.4*
PMEI$_{BERT}$ [20]	90.5	89.8	90.1	91.0	92.9	92.0	88.4	88.9	88.7	80.8	82.8	81.8
TPLinker$_{BERT}$ [23]	91.3	92.5	91.9	91.8	92.0	91.9	91.4	92.6	92.0	88.9	84.5	86.7
StereoRel$_{BERT}$ [22]	92.0	92.3	92.2	91.6	92.6	92.1	92.0	92.3	92.2	–	–	–
PRGC$_{BERT}$ [33]	93.3	91.9	92.6	94.0	92.1	93.0	**93.5**	91.9	92.7	89.9	87.2	88.5
R-BPtrNet$_{BERT}$‡ [3]	92.7	92.5	92.6	**93.7**	92.8	93.3	–	–	–	–	–	–
BiRTE$_{LSTM}$	86.5	89.0	87.7	90.5	91.6	91.0	86.4	87.1	86.7	85.2	87.3	86.2
BiRTE$_{BERT}$	92.2	**93.8**	**93.0**	93.2	94.0	**93.6**	91.9	**93.7**	**92.8**	89.0	**89.5**	**89.3**

Model	Partial Match						Exact Match					
	NYT10			NYT11			NYT10			NYT11		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
PMEI$_{LSTM}$ [20]	79.1	67.2	72.6	56.0	58.6	57.2	75.4	65.8	70.2	**55.3**	57.8	56.5
CasRel$_{BERT}$ [24]	77.7	68.8	73.0	50.1	58.4	53.9	76.8*	68.0*	72.1*	49.1*	56.4*	52.5*
StereoRel$_{BERT}$ [22]	80.0	67.4	73.2	53.8	55.4	54.6	–	–	–	–	–	–
PMEI$_{BERT}$ [20]	79.1	70.4	74.5	55.8	59.7	57.7	77.3	69.7	73.3	54.9	58.9	56.8
TPLinker$_{BERT}$ [23]	78.9*	71.1*	74.8*	55.9*	60.2*	58.0*	78.5*	68.8*	73.4*	54.8*	59.3*	57.0*
BiRTE$_{LSTM}$	79.0	68.8	73.5	55.1	60.4	57.6	76.1	67.4	71.5	54.1	60.5	57.1
BiRTE$_{BERT}$	**80.6**	**71.8**	**76.0**	**56.4**	**62.0**	**59.1**	**80.1**	**71.4**	**75.5**	55.0	**61.2**	**57.9**

Table 2: Main experiments. Note *CGT* uses *UniLM* [5]. ‡: R-BPtrNet uses extra entity type features while all other models not.

Experiments

Model	NYT*								WebNLG*							
	Normal	SEO	EPO	T = 1	T = 2	T = 3	T = 4	T ≥ 5	Normal	SEO	EPO	T = 1	T = 2	T = 3	T = 4	T ≥ 5
CasRel$_{BERT}$ [24]	87.3	91.4	92.0	88.2	90.3	91.9	94.2	83.7	89.4	92.2	94.7	89.3	90.8	94.2	92.4	90.9
TPLinker$_{BERT}$ [23]	90.1	93.4	94.0	90.0	92.8	93.1	96.1	90.0	87.9	92.5	95.3	88.0	90.1	94.6	93.3	91.6
PRGC$_{BERT}$ [33]	91.0	94.0	94.5	91.1	93.0	93.5	95.5	**93.0**	90.4	93.6	95.9	89.9	91.6	95.0	94.8	92.8
R-BPtrNet$_{BERT}$ [3]	90.4	94.4	**95.2**	89.5	93.1	93.5	**96.7**	91.3	89.5	93.9	**96.1**	88.5	91.4	**96.2**	**94.9**	**94.2**
BiRTE$_{BERT}$	**91.4**	**94.7**	94.2	**91.5**	**93.7**	**93.9**	95.8	92.1	**90.1**	**95.9**	94.3	**90.2**	**92.9**	95.7	94.6	92.0

Table 3: F1 scores on sentences with different overlapping pattern and different triplet number. Results of *CasRel* are copied from *TPLinker* directly. "T" is the number of triples contained 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
BiRTE$_{BERT}$	**92.2**	**93.8**	**93.0**	**93.2**	**94.0**	**93.6**	**91.9**	93.7	**92.8**	**89.0**	**89.5**	**89.3**
BiRTE$_{s2o}$	91.5	91.3	91.4	92.0	90.4	91.2	91.5	91.0	91.2	88.3	87.0	87.6
BiRTE$_{o2s}$	91.4	91.0	91.2	91.8	90.5	91.1	91.5	90.8	91.1	88.5	87.5	88.0
BiRTE$_{FinePipeline}$	90.4	91.2	90.8	91.0	91.6	91.3	89.7	90.1	89.9	84.0	85.6	84.8
BiRTE$_{CoarsePipeline}$	90.9	92.3	91.6	91.9	92.1	92.0	90.5	91.0	90.7	85.7	87.3	86.5
BiRTE$_{OneLr}$	91.0	92.4	91.7	92.5	93.6	93.0	91.2	91.8	91.5	88.1	89.0	88.5
BiRTE$_{uif}$	91.6	92.9	92.2	92.7	93.8	93.2	91.3	92.5	91.9	88.8	88.6	88.7
BiRTE$_{tru}$	92.1	93.4	92.7	93.2	93.8	93.5	91.5	93.2	92.3	88.9	89.3	89.1
BiRTE$_{BIO}$	92.1	93.7	92.9	93.0	93.9	93.4	**91.9**	93.8	**92.8**	88.8	**89.5**	89.1
BiRTE$_{2step}$	89.5	92.3	90.9	89.9	91.9	90.9	89.0	91.5	90.2	84.7	87.6	86.1
BiRTE$_{Li}$	91.0	93.6	92.3	91.6	92.9	92.2	90.5	**93.9**	92.2	87.2	89.3	88.2

Table 4: Results of detailed evaluations.

Experiments

Models	Direction	NYT*	WebNLG*	NYT	WebNLG
BiRTE	s2o	95.0	95.3	94.2	91.0
	o2s	94.8	95.6	93.9	91.1
BiRTE_{s2o}	s2o	93.6	92.6	93.1	89.3
BiRTE_{o2s}	o2s	93.2	92.8	92.8	89.5

Table 5: F1 results of the ground entity extraction.

Models	NYT*	WebNLG*	NYT	WebNLG
ETL-Span	54.3	56.1	56.8	60.2
CasRel	49.7	48.5	55.7	51.8
BiRTE_{s2o}	55.2	39.6	56.0	42.8
BiRTE_{o2s}	53.5	51.2	54.8	52.2
BiRTE	9.7	5.4	11.0	9.3

Table 6: Proportions (%) of triples that are not extracted due to the *ground entity extraction failure* issue.

Experiments

Model	Partial Match						Exact Match					
	NYT*			WebNLG*			NYT			WebNLG		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
ETL-Span$_{BiDir}$	84.6	73.5(↑)	78.7(↑)	83.3	92.0(↑)	87.4	85.2	73.0(↑)	78.6(↑)	83.5	83.1(↑)	83.3(↑)
CasRel$_{BiDir}$	89.0	91.1(↑)	90.0(↑)	92.6	91.2(↑)	91.9(↑)	89.0	90.1(↑)	89.5(↑)	87.1	85.1(↑)	86.1
ETL-Span$_{SaLr}$	85.3(↑)	73.0(↑)	78.7(↑)	84.3(↑)	91.7(↑)	87.8(↑)	86.2(↑)	72.3(↑)	78.6(↑)	83.0	84.6(↑)	83.8(↑)
CasRel$_{SaLr}$	90.1(↑)	89.9(↑)	90.0(↑)	93.5(↑)	90.5(↑)	92.0(↑)	90.1(↑)	89.1(↑)	89.6(↑)	87.9	87.1(↑)	87.5(↑)

Table 7: Adaptability evaluations. "↑" denotes the performance is increased.

Conclusions

Two main contributions:

1. the ground entity extraction failure

bidirectional extraction framework

$subject \rightarrow object \rightarrow relation\ (s2o)$

$object \rightarrow subject \rightarrow relation\ (o2s)$

biaffine model

2. the convergence rate inconsistency issue existed in the share structures

the share-aware learning mechanism

Thank you!