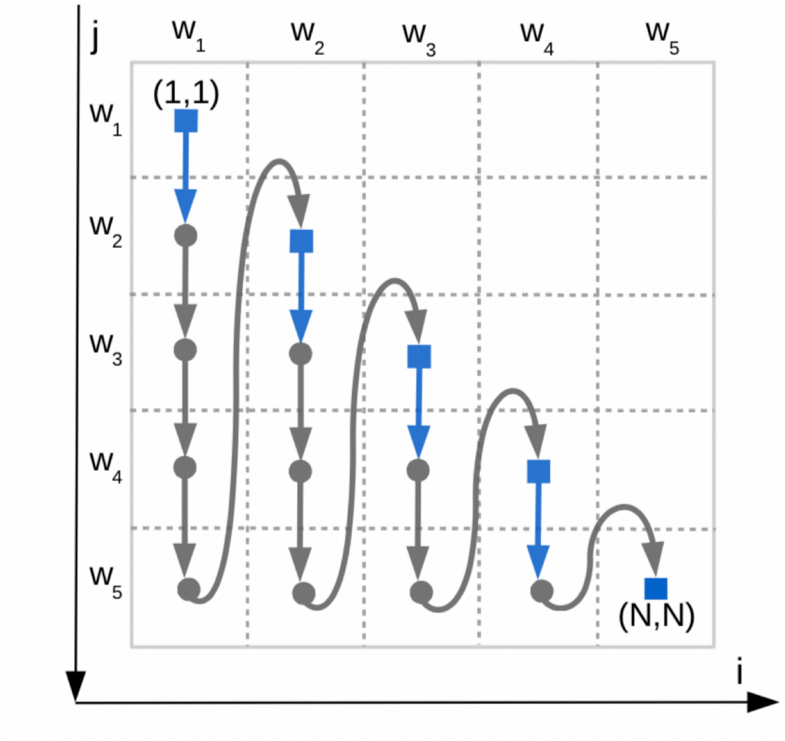
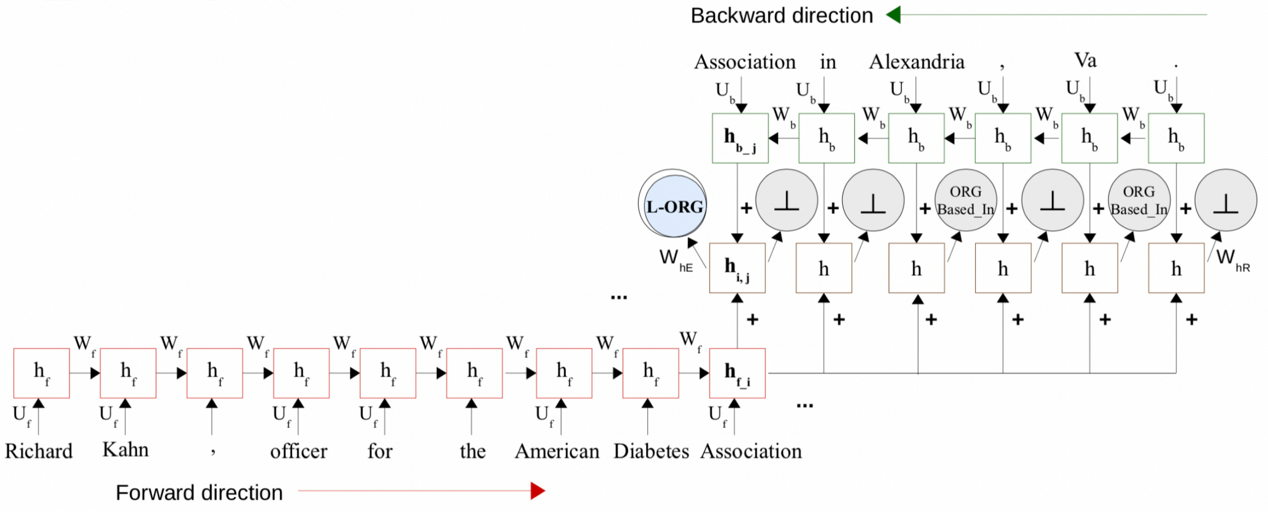
A Novel Hierarchical Binary Tagging Framework for Joint Extraction of Entities and Relations

Zhepei Wei[1], Jianlin Su[2], Yue Wang[3], Yuan Tian[1], Yi Chang[1]
[1]School of Artificial Intelligence, Jilin University
[2]Shenzhen Zhuiyi Technology Co., Ltd.
[3]School of Information and Library Science, University of North Carolina at Chapel Hill
weizp19@mails.jlu.edu.cn, bojonesu@wezhuiyi.com, wangyue@email.unc.edu, yuantian@jlu.edu.cn, yichang@jlu.edu.cn

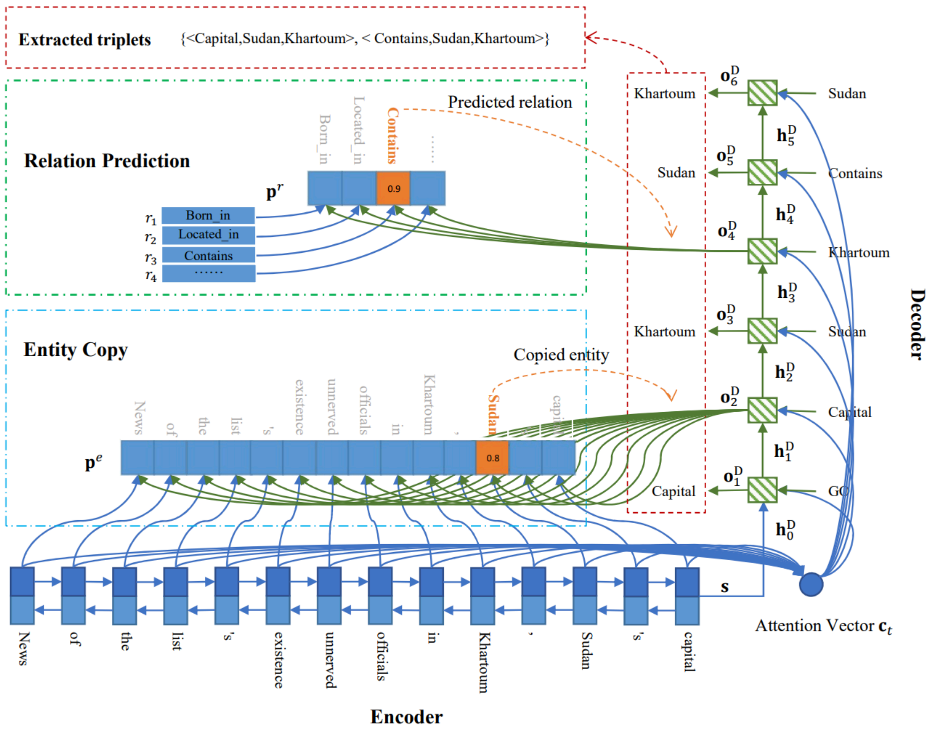
Relation Extraction Models

1. Table Filling

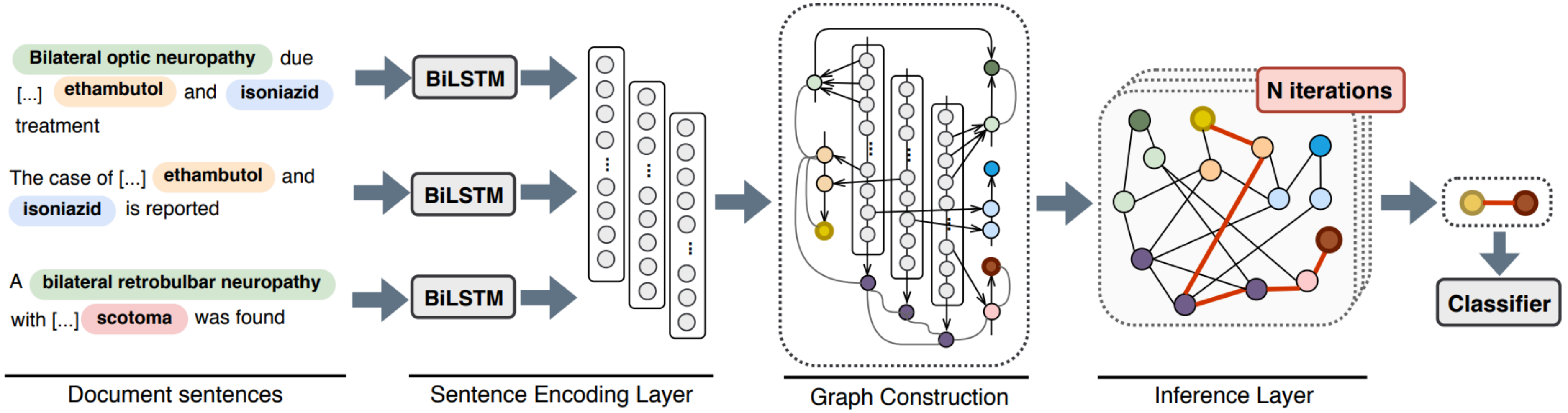




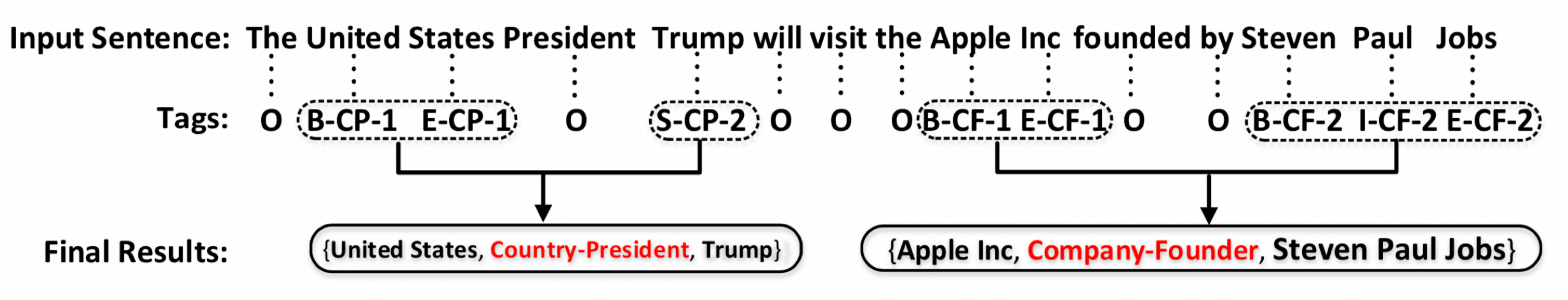
2. Seq2Seq



3. GNN



4. Tagging



一个实体/词语最多只能被指派到一个relation中

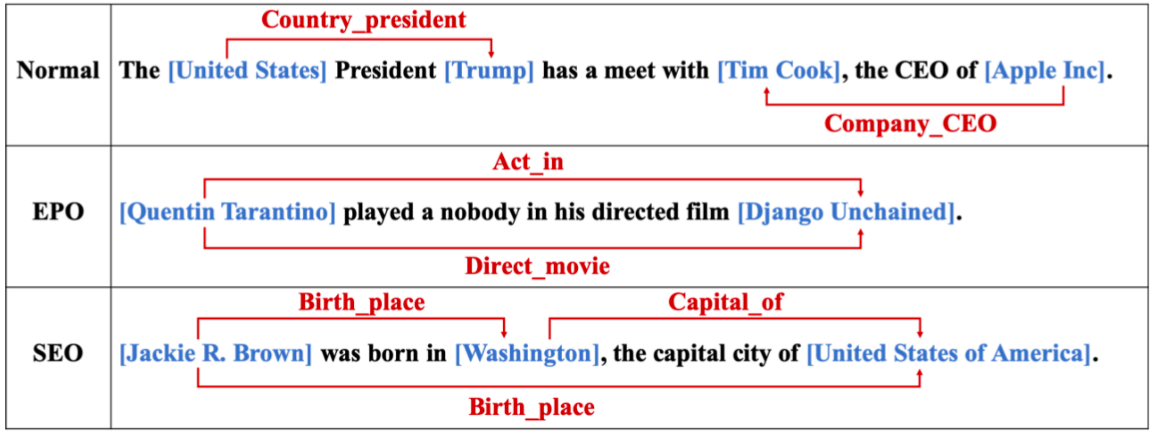


Figure 1: Examples of *Normal*, *EntityPairOverlap (EPO)* and *SingleEntityOverlap (SEO)* overlapping patterns.

提出能够解决实体属于多个relation的基于标注的模型：HBT（Hierarchical Binary Tagging）

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, Li, and Ma 2019)	62.9	57.3	60.0	42.3	39.2	40.7
GraphRel$_{2p}$ (Fu, Li, and Ma 2019)	63.9	60.0	61.9	44.7	41.1	42.9
HBT$_{random}$	84.7	72.3	78.0	67.9	40.4	50.6
HBT	**89.7**	**85.4**	**87.5**	**89.5**	**88.0**	**88.8**

Table 2: Results of different methods on NYT and WebNLG datasets.

Two-step process:

1. Identify all possible subjects in a sentence

2. Apply relation-specific taggers to identify all possible relations and the corresponding objects

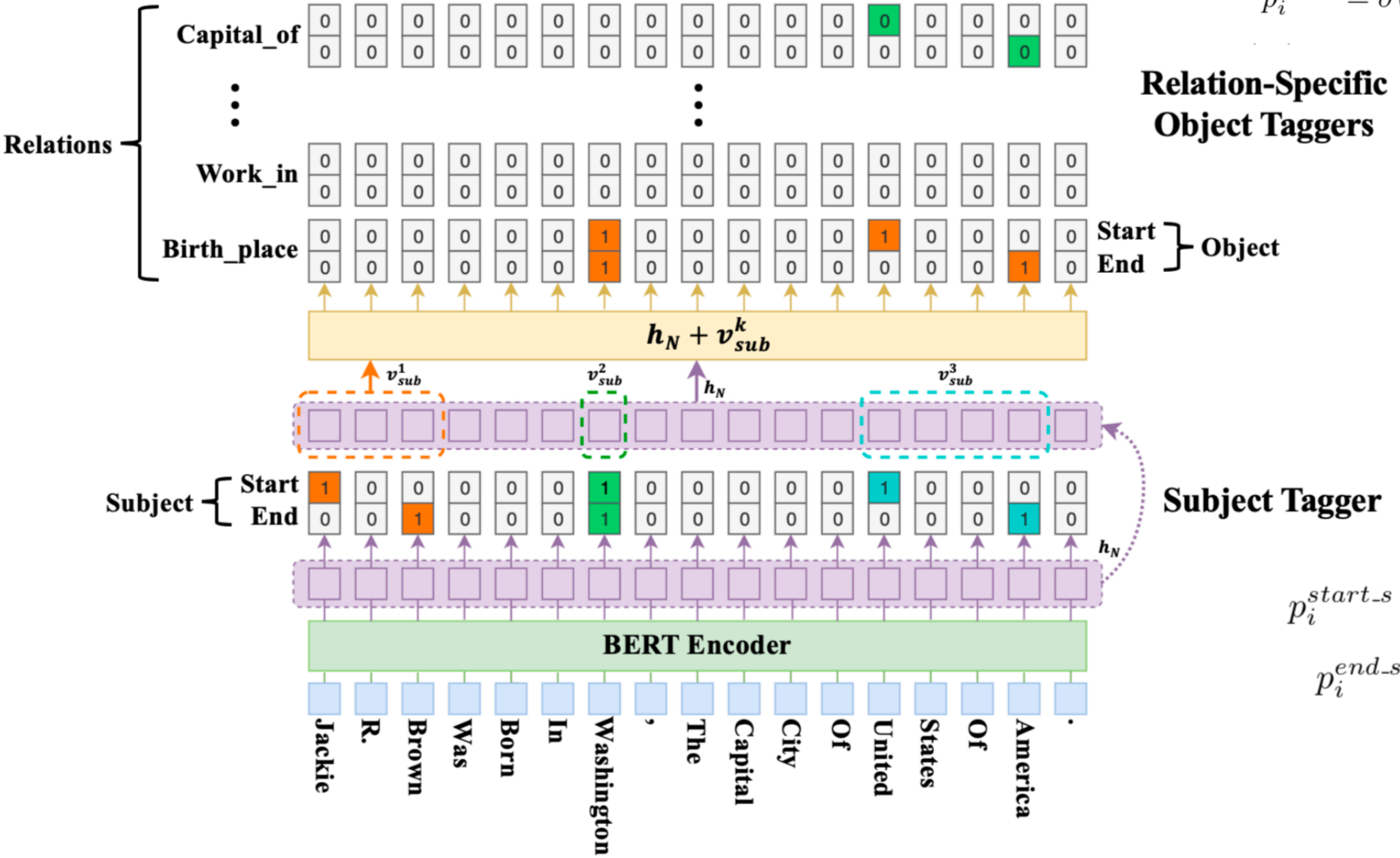
$$\prod_{j=1}^{|D|}\left[\prod_{(s,r,o)\in T_j} p((s,r,o)|x_j)\right] \tag{1}$$

$$=\prod_{j=1}^{|D|}\left[\prod_{s\in T_j} p(s|x_j)\prod_{(r,o)\in T_j|s} p((r,o)|s,x_j)\right] \tag{2}$$

$$=\prod_{j=1}^{|D|}\left[\prod_{s\in T_j} p(s|x_j)\prod_{r\in T_j|s} p_r(o|s,x_j)\prod_{r\in R\backslash T_j|s} p_r(o_\varnothing|s,x_j)\right] \tag{3}$$

$$p_i^{start\_o} = \sigma(\mathbf{W}^r_{start}(\mathbf{x}_i + \mathbf{v}^k_{sub}) + \mathbf{b}^r_{start})$$

$$p_i^{end\_o} = \sigma(\mathbf{W}^r_{end}(\mathbf{x}_i + \mathbf{v}^k_{sub}) + \mathbf{b}^r_{end})$$



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

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

监督信号/Loss计算

$$p_\theta(s|\mathbf{x}) = \prod_{t \in \{start\_s, end\_s\}} \prod_{i=1}^{L} \left(p_i^t\right)^{\mathbf{1}\{y_i^t=1\}} \left(1 - p_i^t\right)^{\mathbf{1}\{y_i^t=0\}} \tag{8}$$

$$p_{\phi_r}(o|s,\mathbf{x}) = \prod_{t \in \{start\_o, end\_o\}} \prod_{i=1}^{L} \left(p_i^t\right)^{\mathbf{1}\{y_i^t=1\}} \left(1 - p_i^t\right)^{\mathbf{1}\{y_i^t=0\}} . \tag{11}$$

实验部分

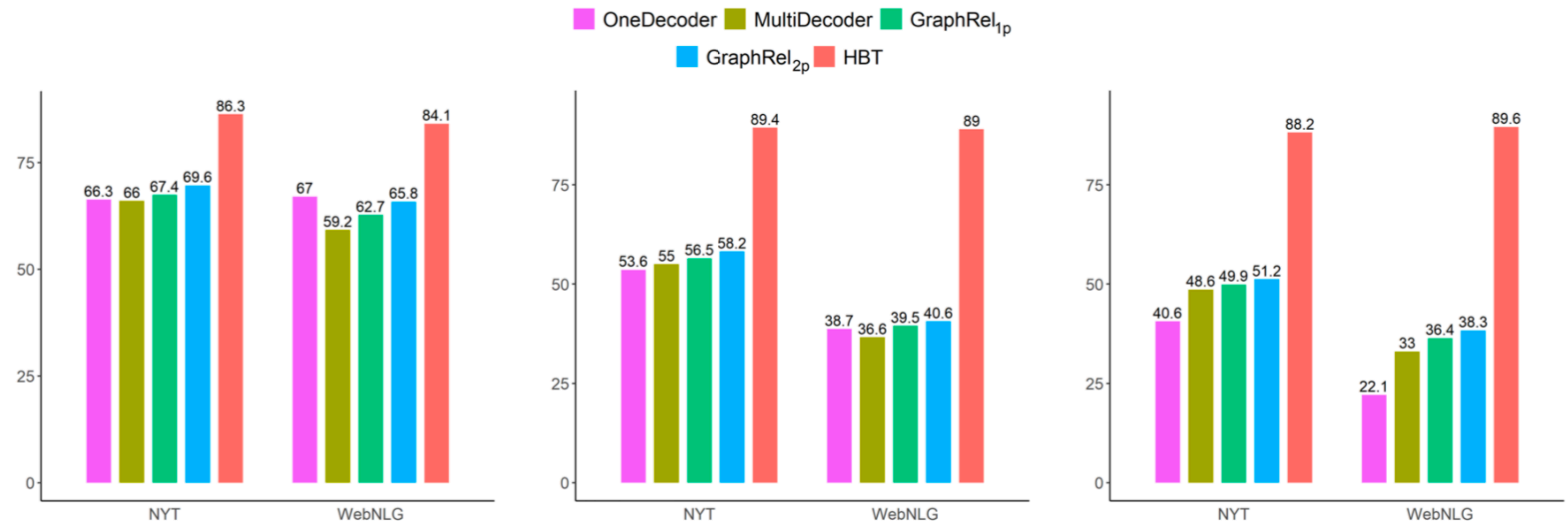
数据集：NYT+WebNLG

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

Table 1: Statistics of datasets. Note that a sentence can belong to both *EPO* class and *SEO* class.

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, Li, and Ma 2019)	62.9	57.3	60.0	42.3	39.2	40.7
GraphRel$_{2p}$ (Fu, Li, and Ma 2019)	63.9	60.0	61.9	44.7	41.1	42.9
HBT$_{random}$	84.7	72.3	78.0	67.9	40.4	50.6
HBT	**89.7**	**85.4**	**87.5**	**89.5**	**88.0**	**88.8**

Table 2: Results of different methods on NYT and WebNLG datasets.



(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.

1. 对基于标注方法目前的缺点进行改进，对每个relation分开判断
2. 参数量随relation种类数增加而增加，对于relation种类比较大的不友好
3. 这都能work？效果还这么好？
4. BERT+Tagging