

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

Zhepei Wei¹, Jianlin Su², Yue Wang³, Yuan Tian¹, Yi Chang¹

¹School of Artificial Intelligence, Jilin University

²Shenzhen Zhuiyi Technology Co., Ltd.

³School 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

这是CopyMTL之后比较新的又一个工作，刷了sota，而且幅度还不小。

主要针对解决的问题和CopyMTL针对copyRE的不同，这篇主要是要解决entity重叠的问题。

传统如果直接使用tagger方法，一个词/entity只能属于一个relation

但是实际上，可以有多个relation的情况。

所以这篇就是来解决这个问题的。

当然了copyRE是第一个显式解决这个问题的，但是解决得还不够好。

其实想法也是比较简单的。

首先是输入句子经过一个bert encoder进行编码。

然后decoder是一个层次结构。

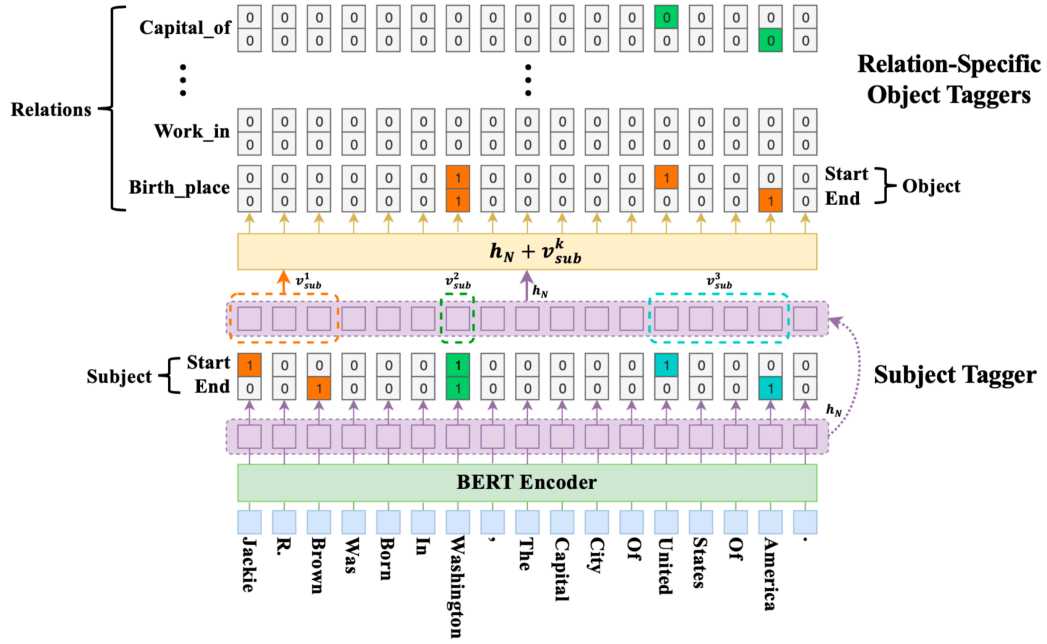
第一层是一个subject tagger，用来标注可能的头实体。

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

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

大于某个threshold（实验是0.5）就是1，否则0。

然后就能够标注出头实体。



之后标注出了k个可能的头实体之后，对每个头实体进行遍历。

每个relation都有自己一套参数，能够标注给定头实体情况下，这个关系可能的尾实体。

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

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

然后就可以进行监督学习啦

$$J(\Theta) = \sum_{j=1}^{|D|} \left[\sum_{s \in T_j} \log p_{\theta}(s | \mathbf{x}_j) + \sum_{r \in T_j | s} \log p_{\phi_r}(o | s, \mathbf{x}_j) \right. \\ \left. + \sum_{r \in R \setminus T_j | s} \log p_{\phi_r}(o_{\emptyset} | s, \mathbf{x}_j) \right]. \quad (12)$$

数据集同样是NYT和WebNLG，结果变态好

Method	NYT			WebNLG		
	<i>Prec.</i>	<i>Rec.</i>	<i>F1</i>	<i>Prec.</i>	<i>Rec.</i>	<i>F1</i>
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.

而且说其他方法NYT不错但是WebNLG不好，原因是WebNLG实体重叠现象更多，所以更难。