Joint Extraction of Entity Mentions and Relations without Dependency Trees

胡伟龙

武汉大学计算机学院

2017 年 12 月 14 日

Table of Content

1.Introduction

1. 现存的句子级别的实体和关系抽取方法有两种：
   - pipeline 模型：看做两个单独的任务
   - 联合模型：同时识别实体和关系，表现更好。
2. Minwa and Bansal(2016) 提出利用基于序列的 LSTM 进行实体抽取和单独的基于依赖树的 LSTM 层进行关系分类。
3. 本文提出的实体与关系联合抽取模型：
   - 基于 RNNs 的模型, 一个多层双向 LSTM
   - 在每一个时间步，对先前的解码序列使用一个类似 attention 的模型
   - 使用额外的一层 LSTM 对输出序列进行从右往左编码

- RNNs 已经应用于很多序列建模和预测任务，但这些模型都假设输出层元素间是条件独立的
- 对于实体和关系的联合抽取，有基于特征的结构预测模型(Li and Ji,2014;)，联合推理整数线性规划模型(Yang and Cardie,2013)，card-pyramid parsing(Kate and Mooney,2010)和概率图模型(Singh et al.,2013)
- Miwa and Bansal(2016)提出端到端基于序列 LSTM 和树结构模型，通过序列层抽取实体并通过最短依赖路径网络识别实体间关系
- Pointer networks(Vinyals et al., 2015)是 attention 的变体，将 token 级别的权重作为指针指向输入元素。比如 Zhai et al.（2017）利用 Pn 做 neural chunking,Cheng and Lapata(2016) 做 summarization

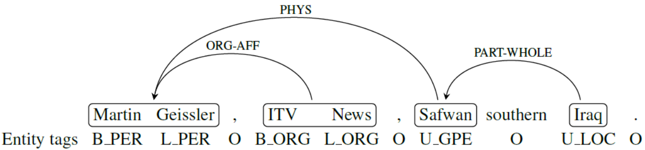
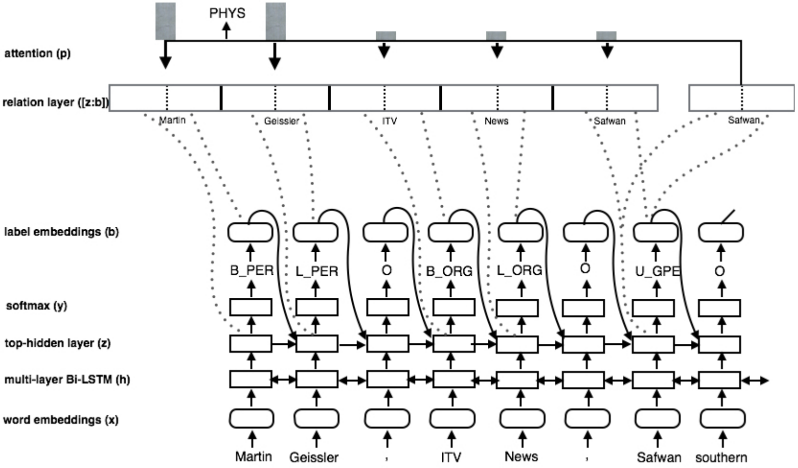


Figure 1: Gold standard annotation for an example sentence from ACE05 dataset.

- 关系标签从实体级别转换到了 token 级别
- 分别建模 “ITV” 和 “News” 与 “Martin Geissler” 的关系

3.1 Multi-layer Bi-directional Recurrent Network

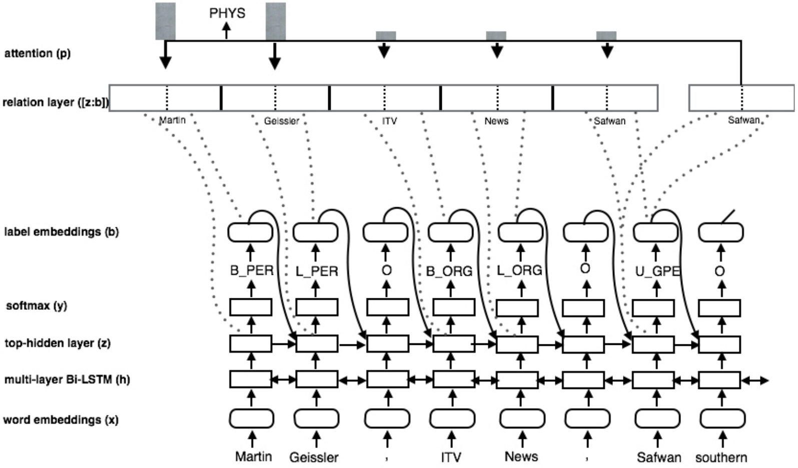


$$\overrightarrow{h}_t = LSTM(x_t, \overrightarrow{h}_{t-1})$$
$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t-1})$$

最顶层 L 的输出计算如下:

$$z_t' = \overrightarrow{V}\overrightarrow{h}_t^{(L)} + \overleftarrow{V}\overleftarrow{h}_t^{(L)} + c$$

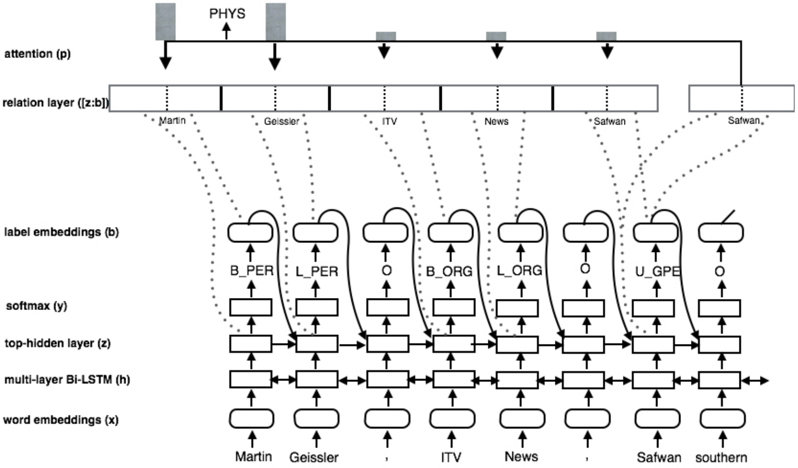
3.2 Entity detection



$$y_t = softmax(Uz_t' + b)$$

$$z_t = LSTM([z_t'; b_{t-1}^k], h_{t-1})$$

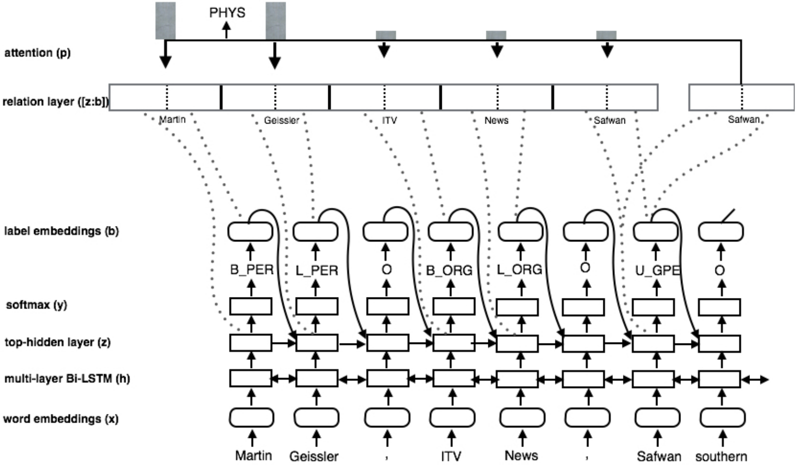
$$y_t = softmax(Uz_t + b')$$



$u_t^i = v^T tanh(W_1 z + W_2 d_i)$

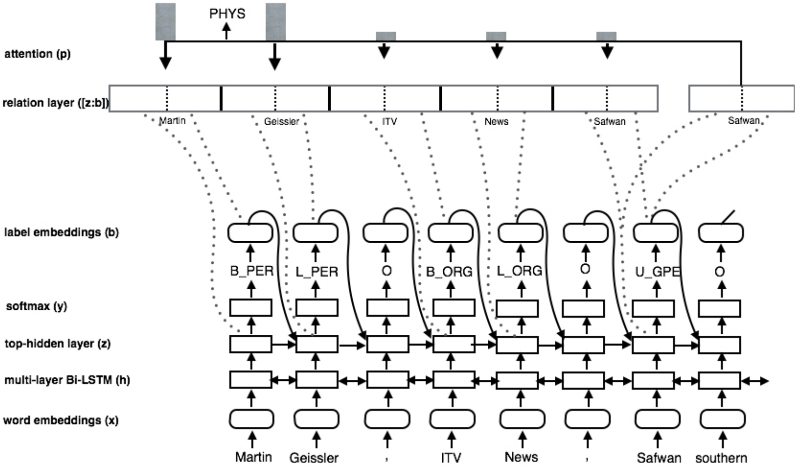
$p_t^i = softmax(u_t^i)$ 其中 d_i 是 decoder 序列中第 i^{th} 个 token, v 是隐含表示转换为 attention score 的权重矩阵

3.4 Relation detection



$$u_{\leq t}^{t} = v^{T} tanh(W_1[z_{\leq t}; b_{\leq t}] + W_2[z_t; d_t])$$

$p_{\leq t}^{t} = softmax(u_{\leq t}^{t})$ $p_{\leq t}^{t}$ 对应了之前所有 token 与当前 token 相关的概率



在上图的的基础上加了另一层 Bi-LSTM 从右向左编码。两种编码共享多层双向 LSTM，除了顶层的隐含层。因此，模型有两个输出层，单独输出 entity tag 和 relation tag。在推断过程中，本文运用启发式从两个方向结合两个输出。

给定正确的 entity E 和 relation R，序列 S，网络学习的目标是最大化如下对数概率：

$$logp(E,R|S,\theta)$$
$$= \frac{1}{|S|}\sum_{i\in|S|} logp(e_i,r_i|e_{<i},r_{<i},S,\theta)$$
$$= \frac{1}{|S|}\sum_{i\in|S|} [logp(e_i|e_{<i},r_{<i}) + logp(r_i|e_{\leq i},r_{<i})]$$

在推断时：$(\hat{E},\hat{R}) = \arg\max_{E,R} p(E,R)$

Multiple Relations

$$\sum_{|j:r'_{i,j}>0|} r'_{i,j} logp(r_{i,j}|e_{\leq i},r_{<i},S,\theta)$$

r'_i 是关系的真实分布，r_i 是模型 softmax 输出值

数据集：ACE04，ACE05 (Automatic Content Extraction)

	Entity			Relation			Entity+Relation		
Method	P	R	F1	P	R	F1	P	R	F1
Li and Ji (2014)	.852	.769	.808	.689	.419	.521	.654	.398	.495
SPTree	.829	.839	.834	–	–	–	.572	.540	.556
SPTree[1]	.823	.839	.831	.605	.553	.578	.578	.529	.553
Our Model	.840	.813	.826	.579	.540	.559	.555	.518	.536

Table 1: Performance on ACE05 test dataset. The dashed ("–") performance numbers were missing in the original paper (Miwa and Bansal, 2016).

	Entity			Relation			Entity+Relation		
Encoding	P	R	F1	P	R	F1	P	R	F1
Left-to-Right	.821	.812	.817	.622	.449	.522	.601	.434	.504
+Multiple Relations	.835	.811	.823	.560	.492	.524	.539	.473	.504
+Bi-directional (Our Model)	.840	.813	.826	.579	.540	.559	.555	.518	.536

Table 2: Performance of different encoding methods on ACE05 dataset.

	Entity			Relation			Entity+Relation		
Method	P	R	F1	P	R	F1	P	R	F1
Li and Ji (2014)	.835	.762	.797	.647	.385	.483	.608	.361	.453
SPTree	.808	.829	.818	–	–	–	.487	.481	.484
Our Model	.812	.781	.796	.502	.488	.493	.464	.453	.457

Table 3: Performance on ACE04 test dataset. The dashed ("–") performance numbers were missing in the original paper (Miwa and Bansal, 2016).

6.Error Analysis

Relation Type	Method	R	P	F1
ART	SPTree Our model	.363 .431	.552 .611	.438 **.505**
PART-WHOLE	SPTree Our model	.560 .520	.538 .538	**.548** .528
PER-SOC	SPTree Our model	.671 .657	.671 .648	.671 .652
PHYS	SPTree Our model	.489 .388	.513 .426	**.500** .406
GEN-AFF	SPTree Our model	.414 .484	.640 .516	.502 .500
ORG-AFF	SPTree Our model	.692 .706	.704 .700	.697 .703

		Relation		
Distance	Method	R	P	F1
≤ 7	SPTree Our model	.589 .591	.628 .605	.608 .598
> 7	SPTree Our model	.275 .153	.375 .259	**.267** .192

Table 5: Performance based on the distance between entity arguments in relations for ACE05 test dataset.

S1 :	the [men]PER:ART-1 held on the sinking [vessel]VEH:ART-1 until the [passenger]PER:ART-2 [ship]VEH:ART-2 was able...
SPTree :	the [men]PER held on the sinking [vessel]VEH until the [passenger]PER [ship]VEH was able to reach them.
Our Model :	the [men]PER:ART-1 held on the sinking [vessel]VEH:ART-1 until the [passenger]PER:ART-2 [ship]VEH:ART-2 was able...
S2 :	[her]PER research was conducted [here]FAC at a [location]FAC:PHYS1 well-known to [u.n.]ORG:ORG-AFF1 [arms]WEA [inspectors]PER:ORG-AFF1. 300 miles west of [baghdad]GPE:PHYS1.
SPTree :	[her]PER research was conducted [here]GPE at a [location]LOC:PHYS1 well-known to u.n. [arms]WEA [[inspectors] PER:PHYS1,PHY2. 300 miles west of [baghdad]GPE:PHYS2.
Our Model :	[her]PER research was conducted [here]FAC:PHYS1 at a [location]GPE well-known to [u.n.]ORG:ORG-AFF1 [arms]WEA [inspectors]PER:ORG-AFF1. 300 miles west of [baghdad]GPE:PHYS1.
S3 :	... [Abigail Fletcher]PER:PHYS1 , a [marcher]FAC:GEN-AFF2 from [Florida]FAC:GEN-AFF2, said outside the [president]PER:ART3 's [[residence] FAC:ART3, PHYS1.
SPTree :	... [Abigail Fletcher]PER:PHYS1 , a [marcher]FAC:GEN-AFF2 from [Florida]FAC:GEN-AFF2, said outside the [president]PER:ART3 's [[residence]]FAC:ART3, PHYS1.
Our Model :	... [Abigail Fletcher]PER , a [marcher]FAC:GEN-AFF2 from [Florida]FAC:GEN-AFF2, said outside the [president]PER 's residence.

Table 6: Examples from the dataset with label annotations from SPTree and our model for comparison


THANKS