

源码：<https://github.com/xmb-cipher/fofe-ner>
背景：由2015年的一种变长编码方法FOFE（Fixed-size Ordinally Forgetting Encoding）提出来的一个解决NER的方法。过程是对句子的切分词段以及其上下文进行FOFE编码表示后再通过简单的多层神经网络输出

模型：

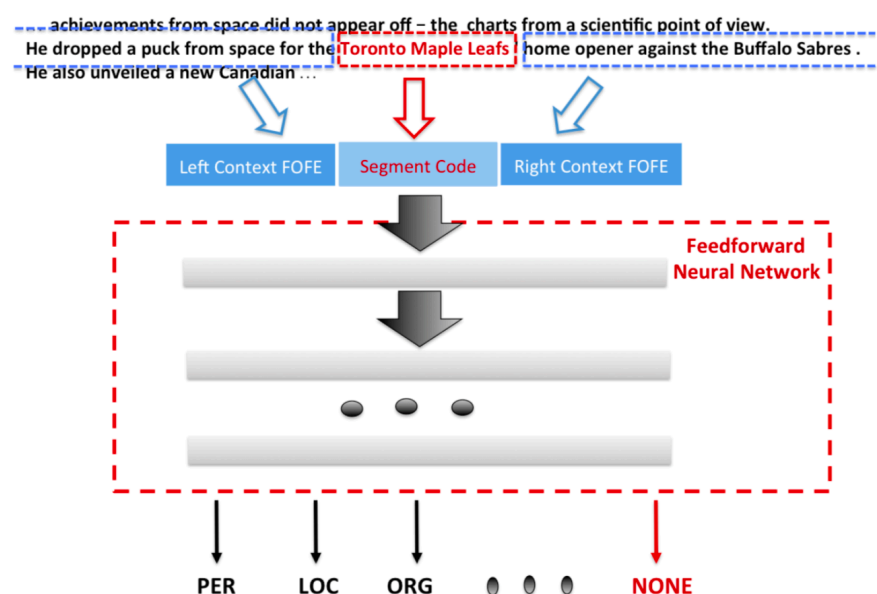


Figure 1: Illustration of the local detection approach for NER using FOFE codes as input and a feedforward neural network as model. The window currently examines the fragment of *Toronto Maple Leafs*. The window will scan and scrutinize all fragments up to K words.

1.FOFE编码方法

对于所有长度<n的序列进行编码

t -th word in S , where $1 \leq t \leq T$. The FOFE of each partial sequence z_t from the first word to the t -th word is recursively defined as:

$$z_t = \begin{cases} \mathbf{0}, & \text{if } t = 0 \\ \alpha \cdot z_{t-1} + \mathbf{e}_t, & \text{otherwise} \end{cases} \tag{8}$$

where the constant α is called forgetting factor, and it is chosen picked between 0 and 1 exclusively. Obviously, the size of z_t is $|V|$, and it is irrelevant to the length of original sequence, T .

α 为从左到右编码的遗忘因子， \mathbf{e}_t 为one-hot向量， z_t 为该部分编码，维度为整个词典大小
由原论文证明该编码是唯一的且不会损失信息

例子 ABCBC编码为 $[\alpha_4, \alpha + \alpha_3, 1 + \alpha_2]$.

过程 A $[1, 0, 0]$ B $[0, 1, 0]$ C $[0, 0, 1]$

从左到右编码: $[1, 0, 0]$ $[\alpha, 1, 0]$ $[\alpha^2, \alpha, 1]$ $[\alpha^3, \alpha^2 + 1, \alpha]$ $[\alpha^4, \alpha^3 + \alpha, 1 + \alpha^2]$

- 2.类似CBOW用2个矩阵映射到低维向量
- 3.低维向量维度确定，通过神经网络softmax输出分类

训练过程对模型的补充：

- 1.加入字符级特征：把已知实体的逐个字符进行双向FOFE编码，用矩阵映射到低维，加入神经网络(应该是只有训练时才用到)
- 2.softmax输出分类为正确的标签和NONE（忽略部分匹配）
- 3.若不想出现实体嵌套或重叠，可直接选择出现分数更高或长度更长的，也可保留嵌套特征，这在序列标注模型是难以做到的(优点1)
- 4.保留大小写...下采样平衡实体和非实体数量...

对模型的解释和其他优点

- 1.类似人类寻找实体的过程，关注NER整体和周围上下文特征，实体内部的关联相对较少，实体和实体间序列上的依赖也相对较少
- 2.无需特征工程，未完全标注的训练集也可以(序列标注效果会很差)
- 3.训练使用3层神经网络简单快速

效果：接近state of art

	word	char	gaz	cap	pos	F1
(Collobert et al., 2011)	✓	✗	✓	✓	✗	89.59
(Huang et al., 2015)	✓	✓	✓	✓	✓	90.10
(Rondeau and Su, 2016)	✓	✗	✓	✓	✓	89.28
(Chiu and Nichols, 2016)	✓	✓	✓	✗	✗	91.62
this work	✓	✓	✗	✗	✗	90.71

Table 2: Performance (F_1 score) comparison among various neural models reported on the CoNLL dataset, and the different features used in these methods.

gaz为地名索引、cap为大小写特征、pos序列标注特征

个人总结：

- 1.模型极度依赖编码方式，从结果说明虽然训练加入了字符特征，主要依赖上下文窗口的方式还是能取得不错的效果，找个时间看源码确认一下
- 2.窗口方式灵活性强，序列标注的障碍在于依赖语料每个字符的输出影响到了句子层面的需求和任务如实体嵌套 以及out-of-vocabulary的问题
- 3.命名实体的周围窗口这里是软识别，保留了概率值但标签一律为NONE，可否像CRF一样针对实体周围加入硬标签？
- 4.实际训练时 α 是否会因为高次方而影响结果，加一个补偿因子？如ABCBBC中出现了4次方，考虑到实体长度有限这点不敢确定。
- 5.实体、实体左右语境等若干输入特征可否有权重分配(比如说实体、实体+实体右部权重比只有实体右部权重更大？)
- 6.主要由窗口判断的话可否引入句法分析层面的特征？