

关于命名实体识别的一些基本方法

一 摘要

最近通过阅读一些经典顶会论文，对命名实体识别的一些基本方法有了一个大致的了解。本文主要介绍了一些在命名实体识别方向上的一些基本的同时也是非常典型的方法，包括隐马尔科夫模型 (HMM) 和条件随机场 (CRF)，以及基于深度学习的一些神经网络架构如双向长短期记忆循环神经网络 (BiLSTM)，以及用卷积神经网络 (CNN) 来提取字符层面的特征后联合词嵌入送进 BiLSTM 这一架构。

关键词：HMM; CRF; BiLSTM; CNN

二 简短介绍

命名实体识别是序列标注类的任务之一，目的是对于给定的句子预测出句子中每一个单词对应的标签。如图所示，B-ORG 代表的是一个单词，它是一个组织名的开头。B-PER 代表的是一个单

```
the 0
Commission B-ORG
's 0
chief 0
spokesman 0
Nikolaus B-PER
van I-PER
der I-PER
Pas I-PER
told 0
a 0
news 0
briefing 0
. 0
```

词，它是人名的开头，I-PER 则代表人名结尾。o 代表其它类型单词，也就是这个单词既不是组织名，也不是地点名同时也不是人名。

三 模型简介

隐马尔科夫模型

隐马尔科夫模型是可以用于序列标注的一种时序概率图模型，描述的是由一组隐藏的状态序列（在 ner 里面就是各个单词对应的标签）生成一个观测序列（就是各个单词）。序列的每一个位置就是一个时刻。

假设 x_1, x_2, \dots, x_n 代表句子中的 n 个单词， y_1, y_2, \dots, y_n 就代表这 n 个单词对应的标签。要计算的是 $\max(P(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n))$ 的概率。HMM 通过贝叶斯公式

$$P(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_n) P(y_1, y_2, \dots, y_n)}{P(x_1, x_2, \dots, x_n)}$$

将计算 $P(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n)$ 改为计算 $P(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_n) P(y_1, y_2, \dots, y_n)$ 。对于 $P(x_1, x_2, \dots, x_n)$ 是给定的输入，可以忽略。

HMM 有两个基本的假设

1. 齐次马尔科夫假设，即当前时刻的状态仅仅依赖于前一个时刻的状态。也就是说 t 时刻第 i 个单词的标签仅仅依赖于第 $i-1$ 个单词对应的标签。因此 $P(y_1, y_2, \dots, y_n)$ 可以写成

$$P(y_1, y_2, \dots, y_n) = P(y_1) p(y_2 | y_1) p(y_3 | y_2) \cdots P(y_n | y_{n-1})$$

2. 观测独立性假设，即当前时刻的观测仅仅依赖于当前时刻的状态，与其它时刻的状态无关。也就是说 t 时刻，单词 i 出现的概率仅仅与它所对应的标签有关，与其它单词和标签无关。因此 $P(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_n)$ 由观测独立性假设就可以得出

$$P(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_n) = P(x_1 | y_1) P(x_2 | y_2) \cdots P(x_n | y_n)$$

在 HMM 中 $P(x_i | y_i)$ 也叫发射概率，意思是在标签为 y_i 的时候，单词 x_i 出现的概率，显然这可以通过训练数据求出来。 $P(y_i | y_{i-1})$ 也叫转移概率，指的是当前单词被标记为 y_{i-1} ，而它的下一个单词被标记为 y_i 的概率。这两个概率都可以通过遍历训练数据集中每一个句子以此来统计相应的频数，用频率来代替概率。