HMM-MEMM-CRF

Bright

2015 年 6 月 1 日

目录

1 Why this article

在做序列标注的时候，我们首先想到要用的方法是CRF，因为这个是目前来说效果最好的方法。但是我们知道，在CRF之前，HMM和MEMM同样可以用来做序列标注，为什么前两者的效果不如CRF呢？下面我们就从建模的角度来看看为什么会这样。

2 Hidden Markov Model

在隐马尔科夫模型中，我们知道存在两种矩阵，一种是状态转移矩阵，我们用A表示；一种是发射矩阵，表示从一个隐藏状态到一个观察值的概率，我们用B表示。

下面的讨论中，我们约定几个变量的使用。输出变量用x表示，即$x = \{x_1, x_2, ..., x_T\}$，每一个$x_i$的取值为$V = \{v_1, v_2, ..., v_{|V|}\}$之一，状态变量用$z = \{z_1, z_2, ..., z_T\}$ 表示，每个状态的取值为$S = \{s_1, s_2, ..., s_{|S|}\}$之一。

HMM的建模过程中，存在两个假设：

a. 一阶马尔科夫假设，即$p(z_i|z_{i-1}, ..., z_1) = p(z_i|z_{i-1})$，当前状态的出现，仅仅依赖于前一个状态

b. 输出独立性假设，即$p(x_i|x_{i-1}, ..., x_1, z_i, ..., z_1) = p(x_i|z_i)$，即当前输出仅仅依赖于当前状态

于是对于一个已知的状态序列和观察序列，我们很容易表示其联合概率如下

$$
\begin{aligned}
p(x, z) &= p(x|z)p(z) && (1) \\
&= p(x_1, ..., x_T|z)p(z_1, ..., z_T) && (2) \\
&= \Pi_{i=1}^{T} p(x_i|z_i) \Pi_{i=1}^{T} p(z_i|z_{i-1}) && (3) \\
&= \Pi_{i=1}^{T} p(x_i|z_i) p(z_i|z_{i-1}) && (4)
\end{aligned}
$$

从上面的公式我们可以看到，在HMM中，我们建模的是联合概率，并且在序列中，每一时刻的观察值是相互独立的，并且仅仅依赖于当前时刻的状态。因此这里的假设是非常强的，而且建模联合概率通常也不是我们所希望的，因此HMM的局限性比较大。

3 Maximum Entropy Markov Model

最大熵模型，想必大家都有所耳闻，MEMM只是将最大熵模型扩展到了Markov链上。最大熵模型的思想是，在满足约束条件的基础上，我们保证熵最大，也就是我们并不对概率模型做任何的假设。

在MEMM中，其建模的是条件概率，并且根据Markov假设的阶数不同，建模也会有所不同。下面以Trigram MEMMs为例，也就是说，当前的

输出，仅仅和前两个输出有关。

$$p(y_1,...,y_T|x_1,...,x_T) = \Pi_{i=1}^{T} p(y_i|x_1,...,x_T,y_{i-2},y_{i-1}) \tag{5}$$

$$p(y_i|x_1,...,x_T,y_{i-2},y_{i-1}) = \frac{exp(\sum_{k=1}^{K}\theta_k f_k(x_1,...,x_T,y_{i-2},y_{i-1},y_i)}{\sum_y exp(\sum_{k=1}^{K}\theta_k f_k(x_1,...,x_T,y_{i-2},y_{i-1},y)} \tag{6}$$

从上面的公式我们可以看到，当前输出的y_i不仅依赖于所有的x_i，还依赖于前两阶的(y_{i-2},y_{i-1})，那么任意特征函数的输入就可以是y_i所依赖的内容。

4 Conditional Random Field

在CRF中建模的也是条件概率，这个时候$p(y|x)$并不是对单个的点进行连乘，CRF中$p(y|x)$表示的就是归一化的概率。对于给定观察序列**x**，求状态序列**y**的条件概率如下：

$$p(y|x) = \frac{p(y,x)}{\sum_{y'} p(y',x)} \tag{7}$$

$$= \frac{\Pi_{t=1}^{T} exp(\sum_{k=1}^{K}\theta_k f_k(y_t,y_{t-1},x))}{\sum_{y'} \Pi_{t=1}^{T} exp(\sum_{k=1}^{K}\theta_k f_k(y'_t,y'_{t-1},x))} \tag{8}$$

于是我们知道在CRF中的$p(y|x)$依赖于整体归一化的结果，而不是像MEMM中依赖于逐个的点进行连乘。所以CRF考虑的更加像是全局最优解，因此CRF能解决像MEMM的转移偏置的问题，当然也比HMM更加的通用。

5 References

1. Hidden Markov Models Fundamentals, Daniel Ramage, CS229 Section Notes, Stanford University.

2. Chapter 8, MEMMS(Log-Linear Tagging Models)