Conditional Random Fields

BrightHush

2014 年 12 月 1 日

目录

1 Model

1.1 Linear Chain CRFs

在介绍linear chain conditional random fields之前，首先介绍条件概率$p(y|x)$，这个条件可以可以通过HMM中的联合概率得到。重点在于条件分布实际上是使用特定特征函数的条件随机场。

在这里我们首先重写HMM联合概率公式，当然是以更加通用的方式重写该公式，其中需要注意的是在HMM 中变量**y**表示状态序列，**x**表示观察序列，因此注意下面公式中各变量的含义及公式本身表示的意义。

$$p(\mathbf{y}, \mathbf{x}) = \frac{1}{Z}\Pi_{t=1}^{T} exp\left(\sum_{i,j \in S} \theta_{ij} 1_{y_t=i} 1_{y_{t-1}=j} + \sum_{i \in S} \sum_{o \in O} \mu_{oi} 1_{y_t=i} 1_{x_t=o}\right) \quad (1)$$

其中$\theta = (\theta_{ij}, \mu_{oi})$是该分布的实数参数，$Z$是归一化参数来保证最后的概率之和为1。上式可以认为是HMM模型的一种泛化，通常的HMM可以当成该模型的一个特例。当$\theta_{ij} = logp(y^{'} = i|y = j), \mu_{oi} = logp(x = o|y = i)$的时候，那么就是通常我们所熟知的HMM模型了。

通过引入*feature functions*，我们可以更加紧密的表示上面的公式。每一个特征函数的形式为$y_k(y_t, y_{t-1}, x_t)$。为了能完整的表达上面的概率公式，需要有一个特征$f_{ij}(y, y^{'}, x) = 1_{y=i} 1_{y^{'}=j}$表示每一个状态转移$(i, j)$，需要有一个特征$f_{io}(y, y^{'}, x) = 1_{y=i} 1_{x=o}$表示状态-观察值对$(i, o)$。我们使用$f_k$表示特征函数，$f_k$需要遍历所有的$f_{ij}$和所有的$f_{io}$。因此我们可以写作下面的方

式:

$$p(\mathbf{y},\mathbf{x}) = \frac{1}{Z}\Pi_{t=1}^{T} exp(\sum_{k=1}^{K}\theta_k f_k(y_t, y_{t-1}, x_t)) \tag{2}$$

最后一步，对于给定观察序列$\mathbf{x}$，求状态序列$\mathbf{y}$的条件概率如下：

$$p(y|x) = \frac{p(y,x)}{\sum_{y'} p(y',x)} \tag{3}$$

$$= \frac{\Pi_{t=1}^{T} exp(\sum_{k=1}^{K}\theta_k f_k(y_t, y_{t-1}, x_t))}{\sum_{y'}\Pi_{t=1}^{T} exp(\sum_{k=1}^{K}\theta_k f_k(y'_t, y'_{t-1}, x_t))} \tag{4}$$

这是一个比较特别的linear chain CRF，因为仅仅考虑了单独当前词对应的特征。而对于其他很多的linear chain CRF 会使用丰富的特征，例如当前词的前缀、后缀，或者是周围词的等等。如果我们用更加通用的特征函数来替换当前词相关的特征函数，那么我们可以得到更加通用的linear chain CRF:

Definition : Let Y, X be random vactors. Then a *linear chain conditional random field* is a distribution $p(\mathbf{y}|\mathbf{x})$ that takes the form:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})}\Pi_{t=1}^{T} exp(\sum_{k=1}^{K}\theta_k f_k(y_t, y_{t-1}, x_t)) \tag{5}$$

where $Z(x)$ is an instance specific normalization function

$$Z(\mathbf{x}) = \sum_{\mathbf{y}}\Pi_{t=1}^{T} exp(\sum_{k=1}^{K}\theta_k f_k(y_t, y_{t-1}, x_t)) \tag{6}$$

通过以上我们可以看到，如果对联合概率$p(y,x)$按照HMM的方式进行分解，那么相应的条件概率$p(y|x)$ 就是一个线性的crf模型。举例来说，在HMM中，通过状态i转到状态j会得到相同的分数$logp(y_t = j|y_{t-1} = i)$，而没有参考输入。在CRF中，我们可以令(i,j)转移得到的分数依赖于当前的观察序列，通过增加一个特征$1_{y_t=j}1_{y_{t-1}=i}1_{x_t=0}$。使用这种特征的CRF模型被广泛的应用在文本领域。

最后，需要注意的是归一化常量$Z(x)$是对所有的可能的状态序列进行求和，这是一个指数级增长的数量。然而，这个可以被forward-backward算法高效的计算。

2 References