

# 条件随机场 (CRF)

## 1 从例子说起-词性标注问题

词性标注问题就是给一个句子中的每个单词注明词性。

**比如这句话：“Bob drank coffee at Starbucks”，注明每个单词的词性后是这样的：“Bob (名词) drank(动词) coffee(名词) at(介词) Starbucks(名词)”。**

下面，就用条件随机场来解决这个问题。

以上面的话为例，有5个单词，我们将：**(名词，动词，名词，介词，名词)**作为一个标注序列，称为 $l$ ，可选的标注序列有很多种，比如 $l$ 还可以是这样：**(名词，动词，动词，介词，名词)**，我们要在这么多的可选标注序列中，挑选出一个**最靠谱**的作为我们对这句话的标注。

怎么判断一个标注序列靠谱不靠谱呢？

就我们上面展示的两个标注序列来说，第二个显然不如第一个靠谱，因为它把第二、第三个单词都标注成了动词，动词后面接动词，这在一个句子中通常是说不通的。

假如我们给每一个标注序列打分，打分越高代表这个标注序列越靠谱，我们至少可以说，凡是标注中出现了**动词后面还是动词**的标注序列，要给它**负分！！**

上面所说的动词后面还是动词就是一个特征函数，我们可以**定义一个特征函数集合，用这个特征函数集合来为一个标注序列打分，并据此选出最靠谱的标注序列。**

也就是说，每一个特征函数都可以用来为一个标注序列评分，把集合中所有特征函数对同一个标注序列的评分综合起来，就是这个标注序列最终的评分值。

## 2 定义CRF中的特征函数

现在，我们正式地定义一下什么是CRF中的特征函数，所谓特征函数，就是这样的函数，它接受四个参数：

- 句子 $s$ （就是我们要标注词性的句子）
- $i$ ，用来表示句子 $s$ 中第 $i$ 个单词
- $l_i$ ，表示要评分的标注序列给第 $i$ 个单词标注的词性
- $l_{i-1}$ ，表示要评分的标注序列给第 $i-1$ 个单词标注的词性

它的输出值是0或者1,0表示要评分的标注序列不符合这个特征，1表示要评分的标注序列符合这个特征。

**Note:**这里，我们的特征函数仅仅依靠当前单词的标签和它前面的单词的标签对标注序列进行评判，这样建立的CRF也叫作**线性链CRF**，这是CRF中的一种简单情况。

为简单起见，本文中我们仅考虑**线性链CRF**。

## 3 从特征函数到概率

定义好一组特征函数后，我们要给每个特征函数  $f_j$  赋予一个权重  $\lambda_j$ 。

现在，只要有一个句子  $s$ ，有一个标注序列  $l$ ，我们就可以利用前面定义的特征函数集来对  $l$  评分。

$$score(l|s) = \sum_{j=1}^m \sum_{i=1}^n \lambda_j f_j(s, i, l_i, l_{i-1})$$

上式中有两个求和，外面的求和用来求每一个特征函数  $f_j$  评分值的和，里面的求和用来求句子中每个位置的单词的特征值的和。

对这个分数进行**指数化和标准化**，我们就可以得到标注序列的概率值 $p(l|s)$ ，如下所示：

$$p(l|s) = \frac{\exp[\text{score}(l|s)]}{\sum_{l'} \exp[\text{score}(l'|s)]} = \frac{\exp[\sum_{j=1}^m \sum_{i=1}^n \lambda_j f_j(s, i, l_i, l_{i-1})]}{\sum_{l'} \exp[\sum_{j=1}^m \sum_{i=1}^n \lambda_j f_j(s, i, l'_i, l'_{i-1})]}$$

## 4 几个特征函数的例子

前面我们已经举过特征函数的例子，下面我们再看几个具体的例子，帮助增强大家的感性认识。

$$f_1(s, i, l_i, l_{i-1}) = 1$$

当 $l_i$ 是“副词”并且第 $i$ 个单词以“ly”结尾时，我们就让 $f_1 = 1$ ，其他情况 $f_1$ 为0。不难想到， $f_1$ 特征函数的权重 $\lambda_1$ 应当是正的。而且 $\lambda_1$ 越大，表示我们越倾向于采用那些把以“ly”结尾的单词标注为“副词”的标注序列

$$f_2(s, i, l_i, l_{i-1}) = 1$$

如果 $i=1$ ， $l_i$ =动词，并且句子 $s$ 是以“?”结尾时， $f_2=1$ ，其他情况 $f_2=0$ 。同样， $\lambda_2$ 应当是正的，并且 $\lambda_2$ 越大，表示我们越倾向于采用那些把问句的第一个单词标注为“动词”的标注序列。

$$f_3(s, i, l_i, l_{i-1}) = 1$$

当 $l_{i-1}$ 是介词， $l_i$ 是名词时， $f_3 = 1$ ，其他情况 $f_3=0$ 。 $\lambda_3$ 也应当是正的，并且 $\lambda_3$ 越大，说明我们越认为介词后面应当跟一个名词。

$$f_4(s, i, l_i, l_{i-1}) = 1$$

如果 $l_i$ 和 $l_{i-1}$ 都是介词，那么 $f_4$ 等于1，其他情况 $f_4=0$ 。

这里，我们应当可以想到 $\lambda_4$ 是负的，并且 $\lambda_4$ 的绝对值越大，表示我们越不认可介词后面还是介词的标注序列。

好了，一个条件随机场就这样建立起来了，让我们总结一下：

为了建一个条件随机场，我们首先要定义一个特征函数集，每个特征函数都以整个句子  $s$ ，当前位置  $i$ ，位置  $i$  和  $i-1$  的标签为输入。

然后为每一个特征函数赋予一个权重，然后针对每一个标注序列  $l$ ，对所有的特征函数加权求和，必要的话，可以把求和的值转化为一个概率值。

## 5 CRF与逻辑回归的比较

观察公式：

$$p(l|s) = \frac{\exp[\text{score}(l|s)]}{\sum_{l'} \exp[\text{score}(l'|s)]} = \frac{\exp[\sum_{j=1}^m \sum_{i=1}^n \lambda_j f_j(s, i, l_i, l_{i-1})]}{\sum_{l'} \exp[\sum_{j=1}^m \sum_{i=1}^n \lambda_j f_j(s, i, l'_i, l'_{i-1})]}$$

是不是有点逻辑回归的味道？

**事实上，条件随机场是逻辑回归的序列化版本。**

**逻辑回归是用于分类的对数线性模型，条件随机场是用于序列化标注的对数线性模型。**

## 6 CRF与HMM的比较

对于词性标注问题，HMM模型也可以解决。HMM的思路是用生成办法，就是说，在已知要标注的句子  $s$  的情况下，去判断生成标注序列  $l$  的概率，如下所示：

$$p(l, s) = p(l_1) \prod_i p(l_i | l_{i-1}) p(w_i | l_i)$$

这里：

$p(l_i | l_{i-1})$  是转移概率，比如， $l_{i-1}$  是介词， $l_i$  是名词，此时的  $p$  表示介词后面的词是名词的概率。

$p(w_i | l_i)$  表示发射概率（emission probability），比如  $l_i$  是名词， $w_i$  是单词 “ball”，此时的  $p$  表示在是名词的状态下，是单词 “ball” 的概率。

那么，HMM和CRF怎么比较呢？

答案是：CRF比HMM要强大的多，它可以解决所有HMM能够解决的问题，并且还可以解决许多HMM解决不了的问题。事实上，我们可以对上面的HMM模型取对数，就变成下面这样：

$$\log p(l, s) = \log p(l_0) + \sum_i \log p(l_i | l_{i-1}) + \sum_i \log p(w_i | l_i).$$

我们把这个式子与CRF的式子进行比较：

$$\text{score}(l|s) = \sum_{j=1}^m \sum_{i=1}^n \lambda_j f_j(s, i, l_i, l_{i-1})$$

不难发现，如果我们把第一个HMM式子中的log形式的概率看做是第二个CRF式子中的特征函数的权重的话，我们会发现，CRF和HMM具有相同的形式。

换句话说，我们可以构造一个CRF，使它与HMM的对数形式相同。怎么构造呢？

对于HMM中的每一个转移概率 $p(l_i=y|l_{i-1}=x)$ ，我们可以定义这样的一个特征函数：

$$f_{x,y}(s, i, l_i, l_{i-1}) = 1$$

该特征函数仅当 $l_i = y, l_{i-1} = x$ 时才等于1。这个特征函数的权重如下：

$$w_{x,y} = \log p(l_i = y | l_{i-1} = x).$$

同样的，对于HMM中的每一个发射概率，我们也都可以定义相应的特征函数，并让该特征函数的权重等于HMM中的log形式的发射概率。

用这些形式的特征函数和相应的权重计算出来的 $p(l|s)$ 和对数形式的HMM模型几乎是一样的！

用一句话来说明HMM和CRF的关系就是这样：

每一个HMM模型都等价于某个CRF

但是，CRF要比HMM更加强大，原因主要有两点：

- **CRF可以定义数量更多，种类更丰富的特征函数。** HMM模型具有天然具有局部性，就是说，在HMM模型中，当前的单词只依赖于当前的标签，当前的标签只依赖于前

一个标签。

- 这样的局部性限制了HMM只能定义相应类型的特征函数，我们在上面也看到了。但是CRF却可以着眼于整个句子s定义更具有全局性的特征函数，如这个特征函数：

$$f_2(s, i, l_i, l_{i-1}) = 1$$

如果 $i=1$ ,  $l_i$ =动词, 并且句子s是以“?”结尾时,  $f_2=1$ , 其他情况 $f_2=0$ 。

**CRF可以使用任意的权重** 将对数HMM模型看做CRF时, 特征函数的权重由于是log形式的概率, 所以都是小于等于0的, 而且概率还要满足相应的限制, 如

$$0 \leq p(w_i | l_i) \leq 1, \sum_w p(w_i = w | l_i) = 1$$

但在CRF中, 每个特征函数的权重可以是任意值, 没有这些限制。

原文: <http://blog.echen.me/2012/01/03/introduction-to-conditional-random-fields>