条件随机场

条件随机场用途：

求序列的隐状态

条件随机场思想：

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

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。

线性链 crf：

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，这是 CRF 中的一种简单情况。
也可以用非线性链条件随机场：
针对上面公式举例子：
1）.虽然 score(l|s)的函数的自变量包含 4 个，但是真正用的时候并不是四个都要用。
　　如：当 l_i 是“副词”并且第 i 个单词以“ly”结尾时，我们就让 f1 = 1，其他情况 f1 为 0。不难想到，f1 特征函数的权重λ1 应当是正的。而且λ1 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。
　　这个例子就仅仅使用了 li 和 i 和 s 这三个变量，和前面一个词对应的隐变量没关系
2）i=1，l_i=动词，并且句子 s 是以“？”结尾时，f2=1，其他情况 f2=0。同样，λ2 应当是正的，并且λ2 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。
　　这个例子就是用了两个变量，s 和 li
3）当 l_i-1 是介词，l_i 是名词时，f3 = 1，其他情况 f3=0。λ3 也应当是正的，并且λ3 越大，说明我们越认为介词后面应当跟一个名词。
　　Emmm........

和 hmm 的比较

hmm 是 crf 的一种特殊情况。Crf 远比 hmm 要强大。
Crf 可以做到，但是 hmm 做不到的例子：
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。
譬如上面的：i=1，l_i=动词，并且句子 s 是以“？”结尾时，f2=1，其他情况 f2=0。
这个例子 hmm 就做不到。但是 crf 可以做到。
Crf 转化为 hmm 的一个例子：
Hmm:
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，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。

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

给 hmm 加 log

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

下面是 crf：

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

做比较可知：
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：

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

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。

和逻辑回归的比较

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

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

条件随机场参数求解方法

就是求解逻辑回归的参数的方法。

条件随机场如何解码

条件随机场更多基础知识

图

是由节点和连接节点的边构成的集合。
节点表示随机变量，边表示随机变量之间的依赖关系。

概率图模型

由图表示的概率分布

成对马尔可夫性

局部马尔可夫性

全局马尔可夫性

以上三个性质非常绕，可以放弃。并不影响 crf 的理解。

概率无向图模型

首先无向图就是没有方向的图。
如果联合概率分布 P(Y)满足成对、局部、全局马尔可夫性就是概率无向图模型，也叫条件随机场。
为什么要满足这三个条件?
因为概率无向图的联合概率分布较为复杂，只有满足了这三个条件，会更加方便求解参数，联合概率分布可以分解出来。

团

任何两个节点均有边连接。就是两个随机变量会有依赖关系，如条件概率、发射概率。
一个团中往往有多个节点。

最大团

一个团中不能再添加边了，就是最大团。

概率无向图的因子分解

上面那么多概念都是为了做这一步，都是为了联合概率分布分解成子问题，才能求解出来。
将概率无向图的联合概率分布表示为：最大团上的随机变量的函数乘积形式。
C 构成一个最大团，C 对应的随机变量是 Yc
Z 是归一化因子

$$P(Y)=\frac{1}{Z}\prod_{C}\Psi_C(Y_C)$$

Z 是规范化因子（normalization factor），由式

$$Z=\sum_{Y}\prod_{C}\Psi_C(Y_C)$$

因子分解的原因

无向图不同于有向图，有向图可以计算条件概率。无向图无法计算条件概率，因为这个条件会发生改变。
因子分解就是说将无向图所描述的联合概率分布表达为若干个子联合概率的乘积，从而便于模型的学习和计算。

条件随机场

输入是一组随机变量，输出是另外一组随机变量
w~v 表示在图 G 中与结点 v 有边连接的所有结点 w，w≠v 表示结点 v 以外的所有结点。Yv，Yu 和 Yw 是结点 v，u 和 w 对应的随机变量。

$$P(Y_v \mid X, Y_w, w \neq v) = P(Y_v \mid X, Y_w, w \sim v)$$

线性链条件随机场

一般假定 X 和 Y 有相同图结构，就是每个 Y 变量对应的图结构比较相似。
其最大团是相邻两个结点的集合。具体定义：设 X=（X1，...，Xn），Y=（Y1，...，Yn）均为线性链表示的随机变量序列，若在给定 X 的条件下，Y 的条件概率分布 P(Y|X)构成条件随机场，满足：

$$P(Y_i \mid X, Y_1, \cdots, Y_{i-1}, Y_{i+1}, \cdots, Y_n) = P(Y_i \mid X, Y_{i-1}, Y_{i+1})$$

$$i = 1, 2, \cdots, n \text{（在 } i = 1 \text{ 和 } n \text{ 时只考虑单边）}$$

则称 P(Y|X)为线性链条件随机场。

线性链条件随机场的形式

参考资料的时候，发现线性链有多种表达形式

参数化形式

tk 和 sl 是特征函数，通常取 1 或 0。其中，tk 是转移函数，依赖与当前和前一个位置；sl 是状态特征，依赖于当前位置。λk 和 ul 是对应的权值，Z(x)是规范化因子。

$$P(y \mid x) = \frac{1}{Z(x)} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i)\right)$$

$$Z(x) = \sum_y \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i)\right)$$

简化形式

观察上面公式发现 s 和 t 两个函数的自变量是一种包含关系，所以两者可以合并为 1 个。

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