第11章 条件随机场

给定一组输入的随机变量，另一组输出随机变量的条件概率分布模型

**11.1 概率无向图模型**

11.1.1模型定义

**概率图模型(probablistic graph model)**: 由图表示的概率分布，用节点表示一个随机变量，用边表示随机变量之间的依赖关系

定义无向图表示的随机变量之间存在以下性质：

**成对马尔可夫性**：给定随机变量组Yo(其它所有节点)的条件下**随机变量Yu和Yv(两个随机变量)条件独立**，即为**P(Yu,Yv|Yo)=P(Yu|Yo)P(Yv|Yo)**

**局部马尔可夫性**：给定随机变量组Yw(其它无关的所有节点)的条件下**随机变量Yv(随机变量)和Yo(与随机变量v有边连接的节点)条件独立**，即为**P(Yv,Yo|Yw)=P(Yv|Yw)P(Yo|Yw)，**P(Yo|Yw)>0时，也可以表示为P(Yv,Yo|Yw)=P(Yv|Yw)

全局马尔可夫性：给定随机变量组Yc条件下随机变量组Ya和Yb(被Yc隔开的两个随机变量组)，则我们可以得到**P(Ya,Yb|Yc)=P(Ya|Yc)P(Ya|Y)**

三种马尔可夫性的等价：**概率图上没有连接的点彼此条件独立**

**概率无向图模型：**只有联合概率分布P(Y)，由无向图G=(V,E)表示，在图G中，节点表示随机变量，边表示随机变量之间的依赖关系，如果联合分布P(Y)满足任意一种马尔可夫性，就称词联合概率分布为**概率无向图模型**，或**马尔可夫随机场**

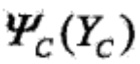
概率无向图的最大特点**：易于因子分解**

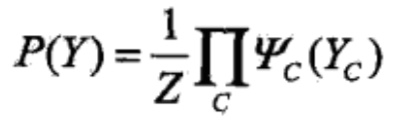
11.1.2 概率无限图模型的因子分解

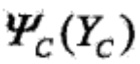
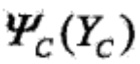
**团**：无向图G中任意两个节点均有边连接的节点子集

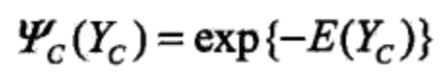
**最大团**：满足上述性质的最大集合

无向图的联合概率分布-分解为其最大团上的随机变量的函数的乘积形式的操作，称为无向图的**因子分解**

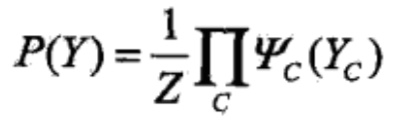
那么概率无向图模型的联合概率分布屏幕快照 2016-08-08 下午2.12.06.png可写作图中所有最大团C上的函数的乘积形式，即

其中，Z是规范化因子（normalization factor),由式

给出。规范化因子保证屏幕快照 2016-08-08 下午2.12.06.png构成一个概率分布。函数称为**势函数**(potential function)。这里要求势函数是严格正的，通常定义为指数函数：

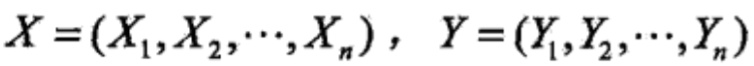
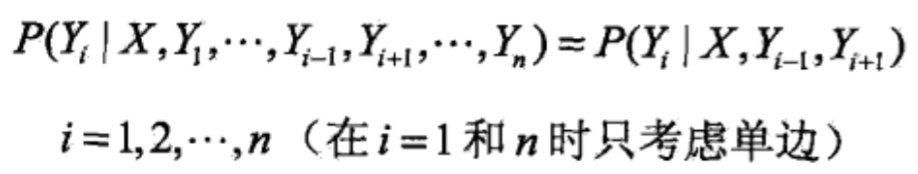


**Hammersley-Clifford定理：**概率无向图模型的联合概率分布可以表示为如下形式：



**11.2 条件随机场的形式定义**

**条件随机场**（conditional random field)：给定随机变量X条件下，随机变量Y的马尔可夫随机场。这里主要介绍定义在线性链上的特殊的条件随机场，称为**线性链条件随机场**（linear chain conditional random field)。

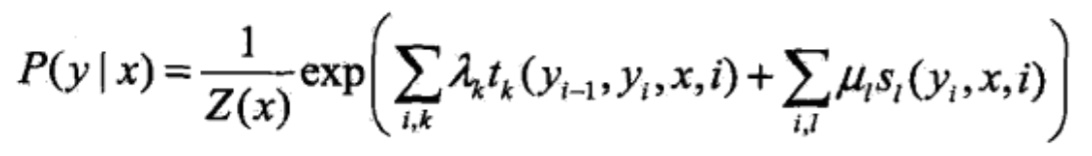
**线性链条件随机场**：设均为线性链表示的随机变量序列，若在给定随机变量序列尤的条件下，随机变量序列Y的条件概率分布IMG_256构成条件随机场，即满足马尔可夫性

11.2.1 条件随机场的定义

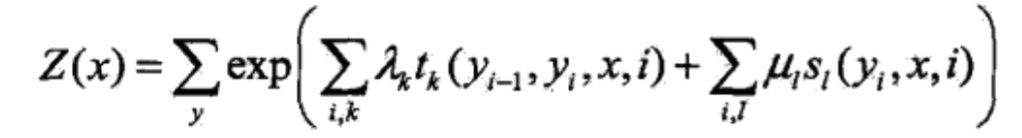
核心是列举说有可能的长度为T的状态I，求各个状态序列I与观测序列O的联合概率，然后对所有可能的状态求和。由P(I|λ)与P(O,I|λ)计算出联合概率P(O,I|λ)，然后再对I求和，得到P(O|λ)

11.2.2 条件随机场的参数化形式

设屏幕快照 2016-08-08 下午3.17.37.png为线性链条件随机场，则在随机变量X取值为x的条件下，随机变量Y取值为y的条件概率具有如下形式：

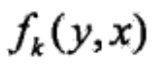


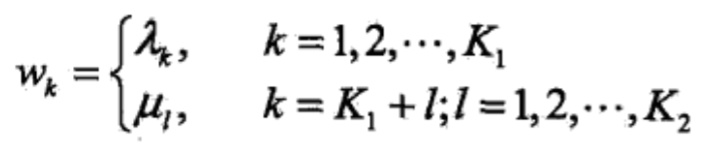
其中，



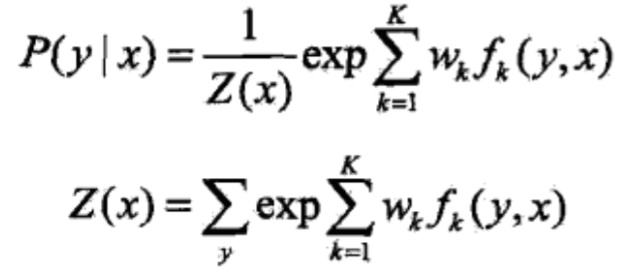
式中，屏幕快照 2016-08-08 下午4.26.31.png和屏幕快照 2016-08-08 下午4.26.44.png是特征函数，屏幕快照 2016-08-08 下午4.27.12.png和屏幕快照 2016-08-08 下午4.27.34.png是对应的权值。

11.2.3 条件随机场的简化形式

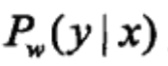
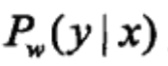
用屏幕快照 2016-08-08 下午4.55.23.png表示特征的权值，即



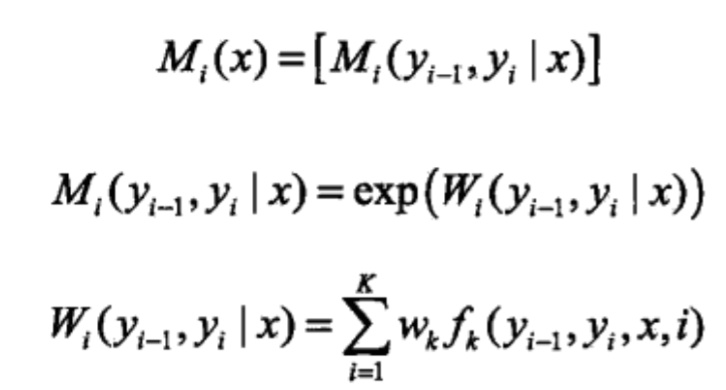
于是，条件随机场可表示为

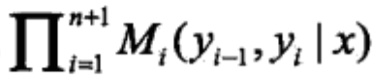
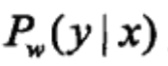


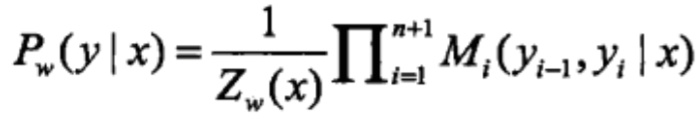
11.2.4 条件随机场的矩阵形式

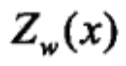
条件随机场还可以由矩阵表示。假设是由内积形式给出的线性链条件随机场，表示对给定观测序列x，相应的标记序列y的条件概率。引进特殊的起点和终点状态标记屏幕快照 2016-08-08 下午5.07.46.png,这时可以通过矩阵形式表示。

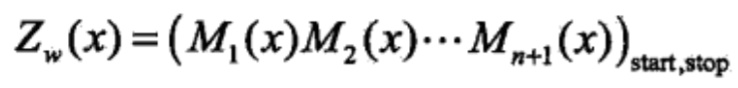
对观测序列x的每一个位置i=1,2,…，n+1，定义一个m阶矩阵（m是标记y\_i取值的个数，因为x是给定的，i-1位置和i位置各有m种可能，所以是m阶的）



这样，给定观测序列x,标记序列y的非规范化概率可以通过n+1个矩阵的乘积表示，于是，条件概率是



其中，为规范化因子，是n+1个矩阵的乘积的(start,stop)元素：



**11.3 条件随机场的概率计算问题**

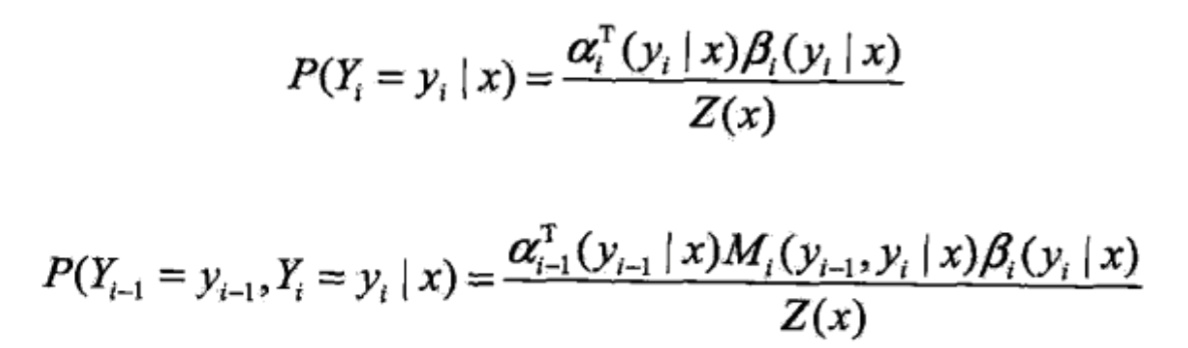
给定条件概率场，求解条件概率

11.3.1 前向-后向算法

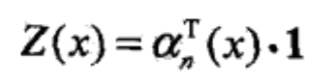
与之前章节讨论的结果一致

11.3.2 概率计算

按照前向-后向向量的定义，很容易计算标记序列在位置i是标记屏幕快照 2016-08-08 下午7.58.41.png的条件概率和在位置i-1与i是标记屏幕快照 2016-08-08 下午8.03.19.png和屏幕快照 2016-08-08 下午7.58.41.png的条件概率：

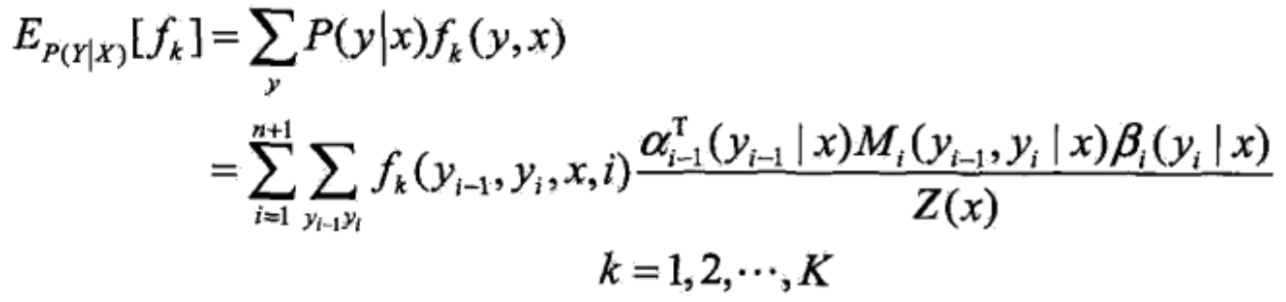


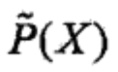
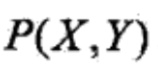
其中，

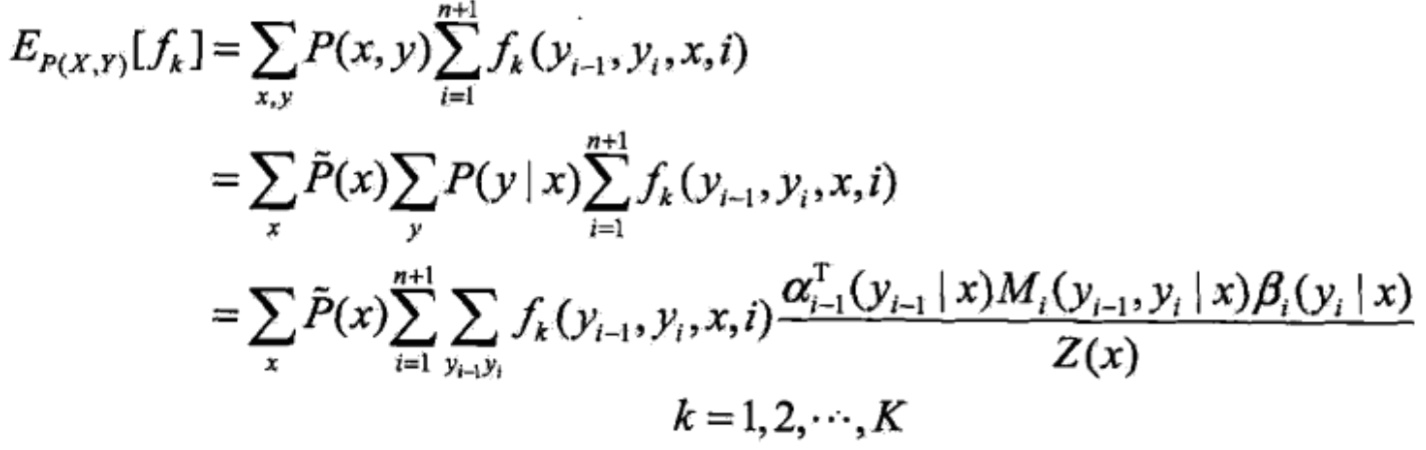


11.3.3 期望值计算

特征函数屏幕快照 2016-08-08 下午8.15.17.png关于条件分布屏幕快照 2016-08-08 下午8.14.46.png的数学期望是



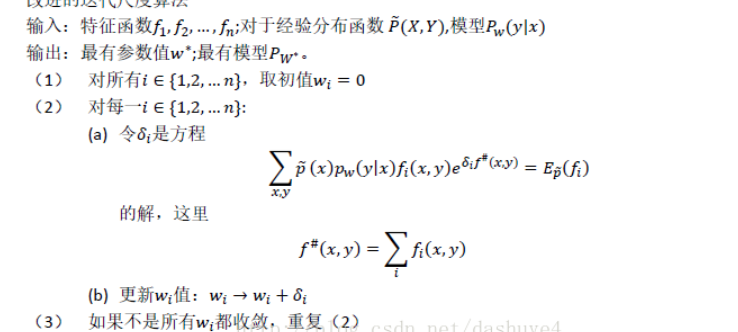
假设经验分布为，特征函数屏幕快照 2016-08-08 下午8.15.17.png关于联合分布的数学期望是



**11.4 条件随机场的学习算法**

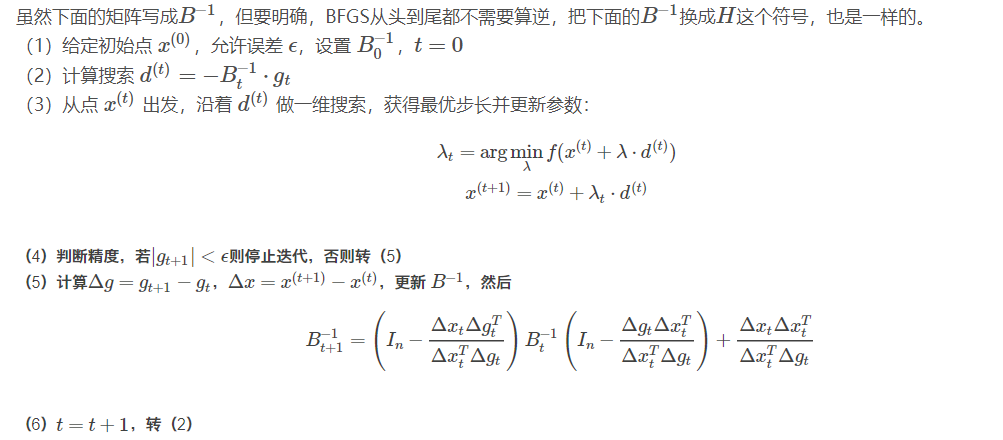
11.4.1 改进的迭代尺度法

这里的目的就是**最大化两次迭代之间的差值**。 差值的最大值如果小于0，说明已经到了最大值，在当前位置向任何方向走，对数似然函数都会变小，我们要求的就是极大似然函数，所以已经得到最大的似然函数，此时的参数即是所求的。



10.4.2 拟牛顿法

牛顿法属于**利用一阶和二阶导数的无约束目标最优化方法**。基本思想是，在每一次迭代中，**以牛顿方向为搜索方向**进行更新。牛顿法对目标的可导性更严格，要求二阶可导，有Hesse矩阵求逆的计算复杂的缺点。



**11.4 条件随机场的预测算法**

条件随机场的预测问题是给定条件随机场P(Y|X)和输入序列（观测序列）x,求条件概率最大的输出序列（标记序列）y\*，即**对观测序列进行标注**。条件随机场的预测算法是著名的**维特比算法**

维特比算法的性质与上一章的讨论一致