

刘建平Pinard

十年码农，对数学统计学，数据挖掘，机器学习，大数据平台，大数据平台应用开发，大数据可视化感兴趣。

博客园 首页 新随笔 联系 订阅 管理

隐马尔科夫模型HMM（二）前向后向算法评估观察序列概率

隐马尔科夫模型HMM（一）HMM模型

隐马尔科夫模型HMM（二）前向后向算法评估观察序列概率

隐马尔科夫模型HMM（三）鲍姆-韦尔奇算法求解HMM参数

隐马尔科夫模型HMM（四）维特比算法解码隐藏状态序列

在隐马尔科夫模型HMM（一）HMM模型中，我们讲到了HMM模型的基础知识和HMM的三个基本问题，本篇我们就关注于HMM第一个基本问题的解决方法，即已知模型和观测序列，求观测序列出现的概率。

1. 回顾HMM问题一：求观测序列的概率

首先我们回顾下HMM模型的问题一。这个问题是这样的。我们已知HMM模型的参数 $\lambda = (A, B, \Pi)$  其中 $A$ 是隐藏状态转移概率的矩阵， $B$ 是观测状态生成概率的矩阵， $\Pi$ 是隐藏状态的初始概率分布。同时我们也已经得到了观测序列 $O = \{o_1, o_2, \dots, o_T\}$ 现在要求观测序列 $O$ 在模型 $\lambda$ 下出现的条件概率 $P(O|\lambda)$

乍一看，这个问题很简单。因为我们知道所有的隐藏状态之间的转移概率和所有从隐藏状态到观测状态生成概率，那么我们是暴力求解的。

我们可以列举出所有可能出现的长度为 $T$ 的隐藏序列 $I = \{i_1, i_2, \dots, i_T\}$ 分布出这些隐藏序列与观测序列 $O = \{o_1, o_2, \dots, o_T\}$ 的联合概率分布 $P(O, I|\lambda)$ 。这样我们就可以很容易的求出边缘分布 $P(O|\lambda)$ 了。

具体暴力求解的方法是这样的：首先，任意一个隐藏序列 $I = \{i_1, i_2, \dots, i_T\}$ 出现的概率是：

$$P(I|\lambda) = \pi_{i_1}a_{i_1i_2}a_{i_2i_3}\dots a_{i_{T-1}i_T}$$

对于固定的状态序列 $I = \{i_1, i_2, \dots, i_T\}$ 我们要求的观察序列 $O = \{o_1, o_2, \dots, o_T\}$ 出现的概率是：

$$P(O|I, \lambda) = b_{i_1}(o_1)b_{i_2}(o_2)\dots b_{i_T}(o_T)$$

则 $O$ 和 $I$ 联合出现的概率是：

$$P(O, I|\lambda) = P(I|\lambda)P(O|I, \lambda) = \pi_{i_1}b_{i_1}(o_1)a_{i_1i_2}b_{i_2}(o_2)\dots a_{i_{T-1}i_T}b_{i_T}(o_T)$$

然后求边缘概率分布，即可得到观测序列 $O$ 在模型 $\lambda$ 下出现的条件概率 $P(O|\lambda)$ ：

$$P(O|\lambda) = \sum_I P(O, I|\lambda) = \sum_{i_1, i_2, \dots, i_T} \pi_{i_1}b_{i_1}(o_1)a_{i_1i_2}b_{i_2}(o_2)\dots a_{i_{T-1}i_T}b_{i_T}(o_T)$$

虽然上述方法有效，但是如果我们的隐藏状态数 $N$ 非常多的那就麻烦了，此时我们预测状态有 $N^T$ 种组合，算法的时间复杂度是 $O(TN^T)$ 阶的。因此对于一些隐藏状态数极少的模型，我们可以用暴力求解法来得到观测序列出现的概率，但是如果隐藏状态多，则上述算法太耗时，我们需要寻找其他简洁的算法。

前向后向算法就是来帮助我们在较低的时间复杂度情况下求解这个问题的。

2. 用前向算法求HMM观测序列的概率

前向后向算法是前向算法和后向算法的统称，这两个算法都可以用来求HMM观测序列的概率。我们先来看看前向算法是如何求解这个问题的。

前向算法本质上属于动态规划的算法，也就是我们要通过找到局部状态递推的公式，这样一步步的从子问题的最优解拓展到整个问题的最优解。

在前向算法中，通过定义“前向概率”来定义动态规划的这个局部状态。什么是前向概率呢，其实定义很简单：定义时刻 $t$ 时隐藏状态为 $q_i$ ，观测状态的序列为 $o_1, o_2, \dots, o_t$ 的概率为前向概率。记为：

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, i_t = q_i|\lambda)$$

既然是动态规划，我们就要递推了，现在我们假设我们已经找到了在时刻 $t$ 时各个隐藏状态的前向概率，现在我们需要递推出时刻 $t + 1$ 时各个隐藏状态的前向概率。

从下图可以看出，我们可以基于时刻 $t$ 时各个隐藏状态的前向概率，再乘以对应的状态转移概率，即 $\alpha_t(j)a_{ij}$ 就是在时刻 $t$ 观测到 $o_1, o_2, \dots, o_t$ 并且时刻 $t$ 隐藏状态 $q_j$ ，时刻 $t + 1$ 隐藏状态 $q_i$ 的概率。如果将想下面所有的线对应的概率求和，即

公告

★珠江追梦，饮岭南茶，恋鄂北家★  
昵称：刘建平Pinard  
园龄：1年5个月  
粉丝：1057  
关注：13  
+加关注

<	2018年3月						>
日	一	二	三	四	五	六	
25	26	27	28	1	2	3	
4	5	6	7	8	9	10	
11	12	13	14	15	16	17	
18	19	20	21	22	23	24	
25	26	27	28	29	30	31	
1	2	3	4	5	6	7	

常用链接

我的随笔  
我的评论  
我的参与  
最新评论  
我的标签

随笔分类(101)

0040. 数学统计学(4)  
0081. 机器学习(62)  
0082. 深度学习(10)  
0083. 自然语言处理(23)  
0121. 大数据挖掘(1)  
0122. 大数据平台(1)  
0123. 大数据可视化

随笔档案(101)

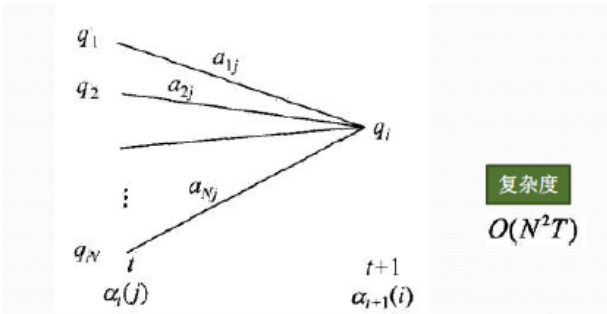
2017年8月 (1)  
2017年7月 (3)  
2017年6月 (8)  
2017年5月 (7)  
2017年4月 (5)  
2017年3月 (10)  
2017年2月 (7)  
2017年1月 (13)  
2016年12月 (17)  
2016年11月 (22)  
2016年10月 (8)

常去的机器学习网站

52 NLP  
Analytics Vidhya

$\sum_{j=1}^N \alpha_t(j) a_{ji}$ 就是在时刻 $t$ 观测到 $o_1, o_2, \dots, o_t$ 并且时刻 $t+1$ 隐藏状态 $q_i$ 的概率。继续一步，由于观测状态 $o_{t+1}$ 只依赖于 $t+1$ 时刻隐藏状态 $q_i$ ，这样 $\left[\sum_{j=1}^N \alpha_t(j) a_{ji}\right] b_i(o_{t+1})$ 就是在时刻 $t+1$ 观测到 $o_1, o_2, \dots, o_t, o_{t+1}$ 并且时刻 $t+1$ 隐藏状态 $q_i$ 的概率。而这个概率，恰恰就是时刻 $t+1$ 对应的隐藏状态 $i$ 的前向概率，这样我们得到了前向概率的递推关系式如下：

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^N \alpha_t(j) a_{ji}\right] b_i(o_{t+1})$$



我们的动态规划从时刻1开始，到时刻 $T$ 结束，由于 $\alpha_T(i)$ 表示在时刻 $T$ 观测序列为 $o_1, o_2, \dots, o_T$ 并且时刻 $T$ 隐藏状态 $q_i$ 的概率，我们只要将所有隐藏状态对应的概率相加，即 $\sum_{i=1}^N \alpha_T(i)$ 就得到了在时刻 $T$ 观测序列为 $o_1, o_2, \dots, o_T$ 的概率。

下面总结下前向算法。

输入：HMM模型 $\lambda = (A, B, \Pi)$  观测序列 $O = (o_1, o_2, \dots, o_T)$

输出：观测序列概率 $P(O|\lambda)$

1) 计算时刻1的各个隐藏状态前向概率：

$$\alpha_1(i) = \pi_i b_i(o_1), \quad i = 1, 2, \dots, N$$

2) 递推时刻2, 3, ..., T时刻的前向概率：

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^N \alpha_t(j) a_{ji}\right] b_i(o_{t+1}), \quad i = 1, 2, \dots, N$$

3) 计算最终结果：

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$$

从递推公式可以看出，我们的算法时间复杂度是 $O(TN^2)$ ，比暴力解法的时间复杂度 $O(TN^T)$ 少了几个数量级。

### 3. HMM前向算法求解实例

这里我们用隐马尔科夫模型HMM（一）HMM模型中盒子与球的例子来显示前向概率的计算。

我们的观察集合是：

$$V = \{\text{红}, \text{白}\}, M = 2$$

我们的状态集合是：

$$Q = \{\text{盒子1}, \text{盒子2}, \text{盒子3}\}, N = 3$$

而观察序列和状态序列的长度为3。

初始状态分布为：

$$\Pi = (0.2, 0.4, 0.4)^T$$

状态转移概率分布矩阵为：

$$A = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

观测状态概率矩阵为：

$$B = \begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \end{pmatrix}$$

球的颜色的观测序列：

$$O = \{\text{红}, \text{白}, \text{红}\}$$

机器学习库

机器学习路线图

深度学习进阶书

深度学习入门书

积分与排名

积分 - 298466

排名 - 614

阅读排行榜

1. 梯度下降（Gradient Descent）小结(94585)
2. 梯度提升树(GBDT)原理小结(45049)
3. 线性判别分析LDA原理总结(30554)
4. scikit-learn决策树算法类库使用小结(27504)
5. 谱聚类（spectral clustering）原理总结(20504)

评论排行榜

1. 梯度提升树(GBDT)原理小结(79)
2. 谱聚类（spectral clustering）原理总结(62)
3. 梯度下降（Gradient Descent）小结(60)
4. 卷积神经网络(CNN)反向传播算法(56)
5. 集成学习之Adaboost算法原理小结(50)

推荐排行榜

1. 梯度下降（Gradient Descent）小结(41)
2. 集成学习原理小结(14)
3. 卷积神经网络(CNN)反向传播算法(14)
4. 集成学习之Adaboost算法原理小结(13)
5. 协同过滤推荐算法总结(11)

按照我们上一节的前向算法。首先计算时刻1三个状态的前向概率：

时刻1是红色球，隐藏状态是盒子1的概率为：

$$\alpha_1(1) = \pi_1 b_1(o_1) = 0.2 \times 0.5 = 0.1$$

隐藏状态是盒子2的概率为：

$$\alpha_1(2) = \pi_2 b_2(o_1) = 0.4 \times 0.4 = 0.16$$

隐藏状态是盒子3的概率为：

$$\alpha_1(3) = \pi_3 b_3(o_1) = 0.4 \times 0.7 = 0.28$$

现在我们可以开始递推了，首先递推时刻2三个状态的前向概率：

时刻2是白色球，隐藏状态是盒子1的概率为：

$$\alpha_2(1) = \left[ \sum_{i=1}^3 \alpha_1(i) a_{i1} \right] b_1(o_2) = [0.1 * 0.5 + 0.16 * 0.3 + 0.28 * 0.2] \times 0.5 = 0.077$$

隐藏状态是盒子2的概率为：

$$\alpha_2(2) = \left[ \sum_{i=1}^3 \alpha_1(i) a_{i2} \right] b_2(o_2) = [0.1 * 0.2 + 0.16 * 0.5 + 0.28 * 0.3] \times 0.6 = 0.1104$$

隐藏状态是盒子3的概率为：

$$\alpha_2(3) = \left[ \sum_{i=1}^3 \alpha_1(i) a_{i3} \right] b_3(o_2) = [0.1 * 0.3 + 0.16 * 0.2 + 0.28 * 0.5] \times 0.3 = 0.0606$$

继续递推，现在我们递推时刻3三个状态的前向概率：

时刻3是红色球，隐藏状态是盒子1的概率为：

$$\alpha_3(1) = \left[ \sum_{i=1}^3 \alpha_2(i) a_{i1} \right] b_1(o_3) = [0.077 * 0.5 + 0.1104 * 0.3 + 0.0606 * 0.2] \times 0.5 = 0.04187$$

隐藏状态是盒子2的概率为：

$$\alpha_3(2) = \left[ \sum_{i=1}^3 \alpha_2(i) a_{i2} \right] b_2(o_3) = [0.077 * 0.2 + 0.1104 * 0.5 + 0.0606 * 0.3] \times 0.4 = 0.03551$$

隐藏状态是盒子3的概率为：

$$\alpha_3(3) = \left[ \sum_{i=1}^3 \alpha_2(i) a_{i3} \right] b_3(o_3) = [0.077 * 0.3 + 0.1104 * 0.2 + 0.0606 * 0.5] \times 0.7 = 0.05284$$

最终我们求出观测序列:  $O = \{\text{红, 白, 红}\}$  的概率为：

$$P(O|\lambda) = \sum_{i=1}^3 \alpha_3(i) = 0.13022$$

## 4. 用后向算法求HMM观测序列的概率

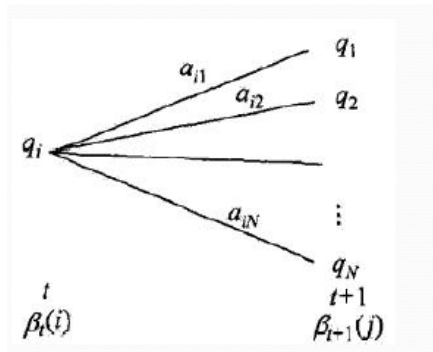
熟悉了用前向算法求HMM观测序列的概率，现在我们来看看怎么用后向算法求HMM观测序列的概率。

后向算法和前向算法非常类似，都是用的动态规划，唯一的区别是选择的局部状态不同，后向算法用的是“后向概率”，那么后向概率是如何定义的呢？

定义时刻 $t$ 时隐藏状态为 $q_i$ ，从时刻 $t+1$ 到最后一时刻 $T$ 的观测状态的序列为 $o_{t+1}, o_{t+2}, \dots, o_T$ 的概率为后向概率。记为：

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | i_t = q_i, \lambda)$$

后向概率的动态规划递推公式和前向概率是相反的。现在我们假设我们已经找到了在时刻 $t+1$ 时各个隐藏状态的后向概率 $\beta_{t+1}(j)$ ，现在我们需要递推出时刻 $t$ 时各个隐藏状态的后向概率。如下图，我们可以计算出观测状态的序列为 $o_{t+2}, o_{t+3}, \dots, o_T$ 时隐藏状态为 $q_i$ ，时刻 $t+1$ 隐藏状态为 $q_j$ 的概率为 $a_{ij}\beta_{t+1}(j)$ 接着可以得到观测状态的序列为 $o_{t+1}, o_{t+2}, \dots, o_T$ 时隐藏状态为 $q_i$ ，时刻 $t+1$ 隐藏状态为 $q_j$ 的概率为 $a_{ij}b_j(o_{t+1})\beta_{t+1}(j)$ 则把下面所有线对应的概率加起来，我们可以得到观测状态的序列为 $o_{t+1}, o_{t+2}, \dots, o_T$ 时隐藏状态为 $q_i$ 的概率为 $\sum_{j=1}^N a_{ij}b_j(o_{t+1})\beta_{t+1}(j)$ 这个概率即为时刻 $t$ 的后向概率。



这样我们得到了后向概率的递推关系式如下：

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

现在我们总结下后向算法的流程,注意下和前向算法的相同点和不同点：

输入：HMM模型  $\lambda = (A, B, \Pi)$  观测序列  $O = (o_1, o_2, \dots, o_T)$

输出：观测序列概率  $P(O|\lambda)$

1) 初始化时刻  $T$  的各个隐藏状态后向概率：

$$\beta_T(i) = 1, \quad i = 1, 2, \dots, N$$

2) 递推时刻  $T-1, T-2, \dots$  时刻的后向概率：

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad i = 1, 2, \dots, N$$

3) 计算最终结果：

$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i)$$

此时我们的算法时间复杂度仍然是  $O(TN^2)$

## 5. HMM常用概率的计算

利用前向概率和后向概率，我们可以计算出HMM中单个状态和两个状态的概率公式。

1) 给定模型  $\lambda$  和观测序列  $O$  在时刻  $t$  处于状态  $q_i$  的概率记为：

$$\gamma_t(i) = P(i_t = q_i | O, \lambda) = \frac{P(i_t = q_i, O | \lambda)}{P(O | \lambda)}$$

利用前向概率和后向概率的定义可知：

$$P(i_t = q_i, O | \lambda) = \alpha_t(i) \beta_t(i)$$

于是我们得到：

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)}$$

2) 给定模型  $\lambda$  和观测序列  $O$  在时刻  $t$  处于状态  $q_i$ ，且时刻  $t+1$  处于状态  $q_j$  的概率记为：

$$\xi_t(i, j) = P(i_t = q_i, i_{t+1} = q_j | O, \lambda) = \frac{P(i_t = q_i, i_{t+1} = q_j, O | \lambda)}{P(O | \lambda)}$$

而  $P(i_t = q_i, i_{t+1} = q_j, O | \lambda)$  可以由前向后向概率来表示为：

$$P(i_t = q_i, i_{t+1} = q_j, O | \lambda) = \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

从而最终我们得到  $\xi_t(i, j)$  的表达式如下：

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{r=1}^N \sum_{s=1}^N \alpha_t(r) a_{rs} b_s(o_{t+1}) \beta_{t+1}(s)}$$

3) 将  $\gamma_t(i)$  和  $\xi_t(i, j)$  在各个时刻  $t$  求和，可以得到：

在观测序列  $O$  下状态  $i$  出现的期望值  $\sum_{t=1}^T \gamma_t(i)$

在观测序列 $O$ 下由状态 $i$ 转移的期望值  $\sum_{t=1}^{T-1} \gamma_t(i)$

在观测序列 $O$ 下由状态 $i$ 转移到状态 $j$ 的期望值  $\sum_{t=1}^{T-1} \xi_t(i, j)$

上面这些常用的概率值在求解HMM问题二，即求解HMM模型参数的时候需要用到。我们在这个系列的第三篇来讨论求解HMM参数的问题和解法。

( 欢迎转载，转载请注明出处。欢迎沟通交流：pinard.liu@ericsson.com )

分类: 0083. 自然语言处理

标签: 自然语言处理

好文要顶

关注我

收藏该文

刘建平Pinard

关注 - 13

粉丝 - 1057

4

0

±加关注

« 上一篇：隐马尔科夫模型HMM（一）HMM模型  
» 下一篇：隐马尔科夫模型HMM（三）鲍姆-韦尔奇算法求解HMM参数

posted @ 2017-06-08 08:47 刘建平Pinard 阅读(2352) 评论(11) 编辑 收藏

评论列表

- # 1楼 2017-09-01 16:32 neptunesong

backward algorithm 中  $\{\beta_t\}$  的定义是不是错了？应该condition on  $\{i\}_t$  而不是联合分布？

支持(0) 反对(0)
- # 2楼[楼主] 2017-09-01 16:48 刘建平Pinard

@ neptunesong  
你好，你是说这个公式吗？

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T, i_t = q_i | \lambda)$$

个人觉得没有什么问题，还请明示。

支持(0) 反对(0)
- # 3楼 2017-09-01 17:09 neptunesong

@ 刘建平Pinard

对，这个公式。我是对比着PRML 621页看的，另外自己用条件概率推了一下，如果是联合概率好像推不下去，条件概率可以。附一下书中的推导

$$\begin{aligned} \beta(z_n) &= p(x_{n+1}, \dots, x_N | z_n) \\ &= \sum_{z_{n+1}} p(x_{n+1}, \dots, x_N, z_{n+1} | z_n) \\ &= \sum_{z_{n+1}} p(x_{n+1}, \dots, x_N | z_n, z_{n+1}) p(z_{n+1} | z_n) \\ &= \sum_{z_{n+1}} p(x_{n+1}, \dots, x_N | z_{n+1}) p(z_{n+1} | z_n) \\ &= \sum_{z_{n+1}} p(x_{n+2}, \dots, x_N | z_{n+1}) p(x_{n+1} | z_{n+1}) p(z_{n+1} | z_n). \end{aligned}$$

Making use of the definition (13.35) for  $\beta(z_n)$ , we then obtain

$$\beta(z_n) = \sum_{z_{n+1}} \beta(z_{n+1}) p(x_{n+1} | z_{n+1}) p(z_{n+1} | z_n). \tag{13.38}$$

你这里的公式是来自于《统计学方法》吗？

支持(0) 反对(0)
- # 4楼[楼主] 2017-09-01 17:20 刘建平Pinard

@ neptunesong  
你好，看到了。  
前向后向算法本质是一个动态规划算法，关键点在于找到一个可以递推的中间状态。看PRML的中间状态的定义和我上面的不同。其实没有关系，只要是可以一步步递推的即可，中间状态定义可以不一样。  
  
我上面其实已经用文字描述了递推的过程，只是没有用PRML那样的公式而已。

另外我这部分内容主要是参考了李航的统计学习方法的中间状态的定义。

支持(0) 反对(0)

#5楼 2017-09-01 17:27 neptunesong

不懂，两者只是符号不一样，但中间状态的概念定义一样吧？时间上你用了符号t，PRML用了符号n。z对应于文中的i，x对用文中的O。你这里只是写成离散化的形式了。如果按照本文中的定义，

“我们可以计算出观测状态的序列为 $o_{t+2}, o_{t+3}, \dots, o_T$ ，t时隐藏状态为 $q_i$ ，时刻t+1隐藏状态为 $q_j$ 的概率为 $a_{ij}\beta_{t+1}(j)$ ”这部分通过条件概率是得不到的？拆不开

另外我不确定《统计学方法》里面的公式是否全部正确。之前推里面的其他章节公式也发现并最终验证是有问题的。

支持(0) 反对(0)

#6楼[楼主] 2017-09-01 17:38 刘建平Pinard

@ neptunesong

你好，我的中间状态递推没用通过条件概率计算，而是根据HMM的定义得到的。

你可以再看下。要是还是觉得不对，我可以再把上面那段写的更详细一些。

当然，你也可以再翻一下李航书里的这部分内容。

支持(0) 反对(0)

#7楼 2017-09-01 18:57 neptunesong

能否把这部分写的再详细些？

“我们可以计算出观测状态的序列为 $o_{t+2}, o_{t+3}, \dots, o_T$ ，t时隐藏状态为 $q_i$ ，时刻t+1隐藏状态为 $q_j$ 的概率为 $a_{ij}\beta_{t+1}(j)$ ”

对于上面的理解我不知道推错了没有，前半句话就写为(1):  
而 $a_{ij}$ 的定义为(2):

$$P(i_{t+1} = q_j | i_t = q_i) = a_{ij}$$

如果 $\beta$ 按照文中的定义，为(3):

$$\beta_{t+1}(j) = P(o_{t+2}, o_{t+3}, \dots, o_T, i_{t+1} = q_j | \lambda)$$

两者相乘，为(4):

$$a_{ij} * \beta_{t+1}(j) = P(i_{t+1} = q_j | i_t = q_i) * P(o_{t+2}, o_{t+3}, \dots, o_T, i_{t+1} = q_j | \lambda)$$

从概率定义上，(1)和(4)是怎么等上的？就是如下这个式子(6)

$$P(i_{t+1} = q_j | i_t = q_i) * P(o_{t+2}, o_{t+3}, \dots, o_T, i_{t+1} = q_j) = P(i_t = q_i, i_{t+1} = q_j, o_{t+2}, \dots, o_T)$$

看上面右边的话，按照条件概率定义：

$$P(A, B) = P(A|B) P(B)$$

如果定义B为(7):

$$B = (i_{t+1} = q_j, o_{t+2}, \dots, o_T)$$

则A为(8):

$$A = \{i_t = q_i\}$$

则 $P(A|B)$ 是(9):

$$P(A|B) = P(i_t = q_i | i_{t+1} = q_j, o_{t+2}, \dots, o_T)$$

而不是定义的 $a_{ij}$ (10):

$$i_{t+1} = q_j | i_t = q_i$$

支持(0) 反对(0)

#8楼 2017-09-01 23:26 neptunesong

另外补充下，从你的第5节公式1也可以看出：

$$P(i_t = q_i, O) = P(O|i_t)P(i_t) = P(O_1 \dots O_{t-1}|i_t)P(O_{t+1} \dots O_T|i_t)P(i_t) = P(O_1 \dots O_{t-1}, i_t)P(O_{t+1} \dots O_T|i_t) = \alpha_i * \beta_i$$

$\beta$ 应该是conditional prob,而不是joint prob

支持(0) 反对(0)

#9楼[楼主] 2017-09-01 23:30 刘建平Pinard

@ neptunesong

你好，刚才回去翻看了之前的学习笔记和李航书。发现李航书里面也是条件概率，是我自己总结的笔记错误，你上面说的是正确的，非常感谢指出错误，后面会修改这部分。

祝好！

#10楼 2018-02-03 10:56 hapjin

博主，感谢您的好文章。关于后向概率，有个不懂的地方：  
为什么后向概率的递归公式在初始化的时候，所有的概率初始化为1？  
能否从后向概率的定义角度出发，解释一下，在时刻T初始化为1的原因？

$$\beta_T(i) = 1, i = 1, 2, \dots, N$$

初始化时刻T的各个隐藏状态后向概率：  
 $\beta_T(i) = 1, i = 1, 2, \dots, N$

支持(0) 反对(0)

#11楼[楼主] 2018-02-11 11:39 刘建平Pinard

@ hapjin  
你好，这个可以从 $\beta_t(i)$ 的定义得到。

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | i_t = q_i, \lambda)$$

注意到 $\beta_T(i)$ ，此时对应的观测序列已经超出时间T的范围了。即此时知道 $q_t$ 后，整个序列已经完全确定，所以对应的后向概率为1.即是一个确定事件。

支持(0) 反对(0)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

【推荐】超50万VC++源码：大型组态工控、电力仿真CAD与GIS源码库！

【缅怀】传奇谢幕，回顾霍金76载传奇人生

【推荐】腾讯云校园拼团福利，1核2G服务器10元/月！

【活动】2050 科技公益大会 - 年青人因科技而团聚



#### 最新IT新闻：

- 微信、支付宝们战火升级，遭殃的是谁
  - 一家周旋于阿里、腾讯、京东、美团之间的服装企业
  - 拼多多为什么能爆红？
  - ICO收割炒币者，交易所收割ICO，谁才是韭菜？
  - 牙买加政府积极探索使用开源软件
- » 更多新闻...



#### 最新知识库文章：

- 写给学者的入门指南
  - 和程序员谈恋爱
  - 学会学习
  - 优秀技术人的管理陷阱
  - 作为一个程序员，数学对你到底有多重要
- » 更多知识库文章...