

# 隐马尔科夫模型HMM的前向算法和后向算法

标签：HMM 隐马尔科夫 概率计算 前向算法 公式推导

2016-03-22 16:43

4477人阅读

评论(0)

收藏

举报

分类：机器学习 (29)

版权声明：本文为博主原创文章，欢迎转载，但请注明出处~

目录(?)

[+]

最近重新看了一遍《统计学习方法》中第十章《隐马尔可夫模型》，更加觉得这本书有浅有深，简洁深刻。HMM模型有三个基本问题：概率计算问题，学习问题和预测问题。今天我就来将其中的**概率计算问题**的一些细节思考总结一下，就当是笔记吧！主要就是概率计算的前向**算法**和后向算法。

## HMM简介

隐马尔可夫模型的参数一般称为其三要素，包括初始状态概率，转移概率和观测概率。其模型的定义建立在两个基本假设的前提上。分别是齐次马尔科夫假设和观测独立性假设。这两个假设是HMM的重点，必须要说一下。

**齐次马尔科夫假设**：通俗地说就是HMM的任一时刻t某一状态只依赖于其前一时刻的状态，与其它时刻的状态及观测无关，也与时刻t无关。

**观测独立性假设**：通俗地说就是任一时刻的观测只依赖于该时刻的马尔科夫链的状态，与其他观测及状态无关。

这两个假设是HMM的核心，之后的公式推导都是依赖这两个假设成立的基础上进行的。

## 概率计算问题

HMM的概率计算问题是HMM三大问题之一。所谓概率计算就是给定一个模型参数已知的HMM和一组观测序列，求这组观测序列由这个HMM所生成的概率。概率计算问题其实评价了模型的好坏，试想如果有两个HMM和一组观测序列，第一个HMM给出的 $P(O|\theta_1)$ 是0.8，第二个HMM给出的 $P(O|\theta_2)$ 是0.9。如果给定多组**测试**观测数据都是这样，那么显然第二个HMM模型更准确一些，性能也更好。HMM的概率计算算法主要有前向算法和后向算法。直接算法虽然理论上可行但计算复杂度过高实际中不可行，我们直接省略掉，下面先介绍前向算法，然后介绍后向算法。

## 前向算法

前向算法定义了一个概念，叫**前向概率**：给定隐马尔科夫模型 $\lambda$ ，定义到时刻t部分观测序列为 $o_1, o_2, \dots, o_t$ 且状态为q的概率为前向概率，记作：

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

有了前向概率，我们就可以递推地求得前向概率 $\alpha_t(i)$ 及观测序列概率 $P(O|\lambda)$ 。具体递推过程为：

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

其中， $a_{ji}$  是转移概率，具体为：

$$a_{ji} = P(i_{t+1} = q | i_t = q)$$

而根据前向概率的定义：

$$\alpha_t(j) = P(o_1, o_2, \dots, o_t, i_t = q | \lambda)$$

根据上面提到的**齐次马尔科夫假设**，有：

$$a_{ji} = P(i_{t+1} = q | i_t = q) = P(i_{t+1} = q | o_1, o_2, \dots, o_t, i_t = q)$$

因此可以将 $\alpha_t(j)a_{ji}$  合并为：

$$P(o_1, o_2, \dots, o_t, i_t = q, i_{t+1} = q | \lambda)$$

在经过求和符号处理之后，有：

$$\sum_{j=1}^N \alpha_t(j) a_{ji} = P(o_1, o_2, \dots, o_t, i_{t+1} = q | \lambda)$$

然后递推式中的 $b(o_{t+1})$ 是观测概率，具体为：

$$b(o_{t+1}) = P(o_{t+1} | i_{t+1} = q) = P(o_{t+1} | o_1, o_2, \dots, o_t, i_{t+1} = q)$$

后面的连等成立是因为应用了前面提到的**观测独立性假设**。这样，最终 $\alpha_{t+1}(i)$ 就可以用概率表示为：

$$\alpha_{t+1}(i) = P(o_1, o_2, \dots, o_t, o_{t+1}, i_{t+1} = q | \lambda)$$

这与我们之前对前向概率的定义是一致的。最后，为了得到 $P(O|\lambda)$ ，我们只要把马尔科夫状态序列的最后一个状态的所有前向概率进行求和便可以了，具体为：

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) = P(o_1, o_2, \dots, o_t, o_{t+1}, \dots, o_T | \lambda)$$

## 后向算法

后向算法与前向算法类似，它定义了一个概念叫后向概率：给定隐马尔科夫模型 $\lambda$ ，定义在时刻 $t$ 状态为 $q$ 的条件下，从 $t+1$ 到 $T$ 的部分观测序列为 $o_{t+1}, o_{t+2}, \dots, o_T$ 的概率为后向概率，记作

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

同样也可以用递推的方法求得后向概率 $\beta_t(i)$ 及观测概率 $P(O|\lambda)$ ，由于和前向算法类似，也用到了HMM的齐次马尔科夫假设和观测独立假设，这里就不详细推导了。

## 前向概率和后向概率的应用

有了前向概率和后向概率的基础，我们就可以利用它们来表示一些HMM概率计算的小问题。比如可以将观测序列概率 $P(O|\lambda)$ 统一写成：

$$P(O|\lambda) = \sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b(o_{t+1}) \beta_{t+1}(j)$$

两个求和符号中的 $\alpha_t(i) a_{ij} b(o_{t+1}) \beta_{t+1}(j)$ 其实经过概率表示等价于：

$$\alpha_t(i) a_{ij} b(o_{t+1}) \beta_{t+1}(j) = P(o_1, o_2, \dots, o_t, i_t = q, i_{t+1} = q | \lambda)$$

两次求和其实就是将状态 $q$ 和状态 $q$ 的所有可能取值都取一遍， $N$ 是所有可能的状态数。

下面给出一些概率和期望值的计算。

### 在时刻 $t$ 处于状态 $q$ 的概率

$$\gamma_t(i) = P(i_t = q | O, \lambda) = \frac{P(i_t = q, O | \lambda)}{P(O | \lambda)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)}$$

### 在时刻 $t$ 处于状态 $q$ 且在时刻 $t+1$ 处于状态 $q$ 的概率

$$\xi_t(i, j) = P(i_t = q, i_{t+1} = q | O, \lambda) = \frac{P(i_t = q, i_{t+1} = q, O | \lambda)}{P(O | \lambda)} = \frac{\alpha_t(i) a_{ij} b(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b(o_{t+1}) \beta_{t+1}(j)}$$

### 利用 $\gamma_t(i)$ 和 $\xi_t(i, j)$ 对各个时刻求和得到有用的期望值

(1) 在观测 $O$ 下状态 $i$ 出现的期望值为

$$\sum_{t=1}^T \gamma_t(i)$$

解释：出现一次的概率是 $\gamma_t(i)$ ，那么长度为 $T$ 的观测下出现的概率就是对 $T$ 求和。

(2) 在观测 $O$ 下由状态 $i$ 转移的期望值为

$$\sum_{t=1}^{T-1} \xi_t(i, j)$$

解释：即除去当前被转移到的状态剩下 $T-1$ 个长度。

( 3 ) 在观测O下由状态i转移到状态j的期望值为

$$\sum_{t=1}^{T-1} \xi_t(i,j)$$

解释：只能有T-1个由状态i转移到状态j的机会。

## 总结

通过这些概率和期望值的计算，可以为后面HMM的其它两个问题即**学习问题**和**预测问题**提供方便。

顶 踩  
3 0

上一篇 浅谈EM算法的两个理解角度

下一篇 HMM的Baum-Welch算法和Viterbi算法公式推导细节