

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（一）HMM模型

隐马尔科夫模型HMM（一）HMM模型基础

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

隐马尔科夫模型（Hidden Markov Model，以下简称HMM）是比较经典的机器学习模型了，它在语言识别，自然语言处理，模式识别等领域得到广泛的应用。当然，随着目前深度学习的崛起，尤其是RNN，LSTM等神经网络序列模型的热度，HMM的地位有所下降。但是作为一个经典的模型，学习HMM的模型和对应算法，对我们解决问题建模的能力提高以及算法思路的拓展还是很好的。本文是HMM系列的第一篇，关注于HMM模型的基础。

1. 什么样的问题需要HMM模型

首先我们来看看什么样的问题解决可以用HMM模型。使用HMM模型时我们的问题一般有两个特征：1）我们的问题是基于序列的，比如时间序列，或者状态序列。2）我们的问题中有两类数据，一类序列数据是可以观测到的，即观测序列；而另一类数据是不能观察到的，即隐藏状态序列，简称状态序列。

有了这两个特征，那么这个问题一般可以用HMM模型来尝试解决。这样的问题在实际生活中是很多的。比如：我现在在打字写博客，我在键盘上敲出来的一系列字符就是观测序列，而我实际想写的一段话就是隐藏序列，输入法的任务就是从敲入的一系列字符尽可能的猜测我要写的一段话，并把最可能的词语放在最前面让我选择，这就可以看做一个HMM模型了。再举一个，我在和你说话，我发出的一串连续的声音就是观测序列，而我实际要表达的一段话就是状态序列，你大脑的任务，就是从这一串连续的声音中判断出我最可能要表达的话的内容。

从这些例子中，我们可以发现，HMM模型可以无处不在。但是上面的描述还不精确，下面我们用精确的数学符号来表述我们的HMM模型。

2. HMM模型的定义

对于HMM模型，首先我们假设Q是所有可能的隐藏状态的集合，V是所有可能的观测状态的集合，即：

$$Q = \{q_1, q_2, \dots, q_N\}, V = \{v_1, v_2, \dots, v_M\}$$

其中，N是可能的隐藏状态数，M是所有的可能的观察状态数。

对于一个长度为T的序列，I对应的状态序列，O是对应的观察序列，即：

$$I = \{i_1, i_2, \dots, i_T\}, O = \{o_1, o_2, \dots, o_T\}$$

其中，任意一个隐藏状态 $i_t \in Q$ 任意一个观察状态 $o_t \in V$

HMM模型做了两个很重要的假设如下：

1）齐次马尔科夫链假设。即任意时刻的隐藏状态只依赖于它前一个隐藏状态，这个我们在MCMC(二)马尔科夫链中有详细讲述。当然这样假设有点极端，因为很多时候我们的某一个隐藏状态不仅仅只依赖于前一个隐藏状态，可能是前两个或者是前三个。但是这样假设的好处就是模型简单，便于求解。如果在时刻t的隐藏状态是 $i_t = q_b$ 在时刻t + 1的隐藏状态是 $i_{t+1} = q_j$ 则从时刻t到时刻t + 1的HMM状态转移概率 a_{ij} 可以表示为：

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

这样 a_{ij} 可以组成马尔科夫链的状态转移矩阵A:

$$A = \begin{bmatrix} a_{ij} \end{bmatrix}_{N \times N}$$

2）观测独立性假设。即任意时刻的观察状态仅仅依赖于当前时刻的隐藏状态，这也是一个为了简化模型的假设。如果在时刻t的隐藏状态是 $i_t = q_j$ 而对应的观察状态为 $o_t = v_b$ 则该时刻观察状态 v_b 在隐藏状态 q_j 下生成的概率为 $b_j(k)$ 满足：

$$b_j(k) = P(o_t = v_k | i_t = q_j)$$

这样 $b_j(k)$ 可以组成观测状态生成的概率矩阵B:

$$B = \begin{bmatrix} b_j(k) \end{bmatrix}_{N \times M}$$

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
	25	26	27	28	1	2
	4	5	6	7	8	9
	11	12	13	14	15	16
	18	19	20	21	22	23
	25	26	27	28	29	30
	1	2	3	4	5	6

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

除此之外，我们需要一组在时刻 $t = 1$ 的隐藏状态概率分布 Π :

$$\Pi = \left[\pi(i) \right]_N \text{ 其中 } \pi(i) = P(i_1 = q_i)$$

一个HMM模型，可以由隐藏状态初始概率分布 Π , 状态转移概率矩阵 A 和观测状态概率矩阵 B 决定。 Π, A 决定状态序列, B 决定观测序列。因此，HMM模型可以由一个三元组表示如下：

$$\lambda = (A, B, \Pi)$$

3.一个HMM模型实例

下面我们用一个简单的实例来描述上面抽象出的HMM模型。这是一个盒子与球的模型，例子来源于李航的《统计学习方法》。

假设我们有3个盒子，每个盒子里都有红色和白色两种球，这三个盒子里球的数量分别是：

盒子	1	2	3
红球数	5	4	7
白球数	5	6	3

按照下面的方法从盒子里抽球，开始的时候，从第一个盒子抽球的概率是0.2，从第二个盒子抽球的概率是0.4，从第三个盒子抽球的概率是0.4。以这个概率抽一次球后，将球放回。然后从当前盒子转移到下一个盒子进行抽球。规则是：如果当前抽球的盒子是第一个盒子，则以0.5的概率仍然留在第一个盒子继续抽球，以0.2的概率去第二个盒子抽球，以0.3的概率去第三个盒子抽球。如果当前抽球的盒子是第二个盒子，则以0.5的概率仍然留在第二个盒子继续抽球，以0.3的概率去第一个盒子抽球，以0.2的概率去第三个盒子抽球。如果当前抽球的盒子是第三个盒子，则以0.5的概率仍然留在第三个盒子继续抽球，以0.2的概率去第一个盒子抽球，以0.3的概率去第二个盒子抽球。如此下去，直到重复三次，得到一个球的颜色的观测序列：

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

注意在这个过程中，观察者只能看到球的颜色序列，却不能看到球是从哪个盒子里取出的。

那么按照我们上一节HMM模型的定义，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：

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

4. HMM观测序列的生成

从上一节的例子，我们也可以抽象出HMM观测序列生成的过程。

输入的是HMM的模型 $\lambda = (A, B, \Pi, \text{观测序列的长度} T)$

输出是观测序列 $O = \{o_1, o_2, \dots, o_T\}$

生成的过程如下：

- 1) 根据初始状态概率分布 Π 生成隐藏状态 i_1
- 2) for t from 1 to T

a. 按照隐藏状态 i_t 的观测状态分布 $b_{i_t}(k)$ 生成观察状态 o_t

b. 按照隐藏状态 i_t 的状态转移概率分布 $a_{i_t, i_{t+1}}$ 产生隐藏状态 i_{t+1}

所有的 o_t 一起形成观测序列 $O = \{o_1, o_2, \dots, o_T\}$

5. HMM模型的三个基本问题

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)

HMM模型一共有三个经典的问题需要解决：

- 1）评估观察序列概率。即给定模型 $\lambda = (A, B, \Pi)$ 和观测序列 $O = \{o_1, o_2, \dots, o_T\}$ 计算在模型 λ 下观测序列 O 出现的概率 $P(O|\lambda)$ 这个问题的求解需要用到前向后向算法，我们在这个系列的第二篇会详细讲解。这个问题是HMM模型三个问题中最简单的。
- 2）模型参数学习问题。即给定观测序列 $O = \{o_1, o_2, \dots, o_T\}$ 估计模型 $\lambda = (A, B, \Pi)$ 的参数，使该模型下观测序列的条件概率 $P(O|\lambda)$ 最大。这个问题的求解需要用到基于EM算法的鲍姆-韦尔奇算法，我们在这个系列的第三篇会详细讲解。这个问题是HMM模型三个问题中最复杂的。
- 3）预测问题，也称为解码问题。即给定模型 $\lambda = (A, B, \Pi)$ 和观测序列 $O = \{o_1, o_2, \dots, o_T\}$ 求给定观测序列条件下，最可能出现的对应的状态序列，这个问题的求解需要用到基于动态规划的维特比算法，我们在这个系列的第四篇会详细讲解。这个问题是HMM模型三个问题中复杂度居中的算法。

(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

+加关注

5

0

« 上一篇：[EM算法原理总结](#)
» 下一篇：[隐马尔科夫模型HMM（二）前向后向算法评估观察序列概率](#)

posted @ 2017-06-06 15:01 刘建平Pinard 阅读(3002) 评论(4) 编辑 收藏

评论列表

- # 1楼 2018-02-03 17:31 hapjin

冒昧打扰一下：我很好奇博主写出的这四篇 隐马尔可夫模型 的文章：是参考的某本书籍吗？或者说博主有没有送入NLP相关算法的书籍推荐？

支持(0) 反对(0)
- # 2楼[楼主] 2018-02-11 11:43 刘建平Pinard

@ hapjin
你好，主要参考的是李航的<统计学习方法>。
NLP领域很广，其实我也只是做文本处理比较多。
如果你想学习文本处理这一块的话，个人推荐从<统计自然语言处理基础>这本书开始。

支持(0) 反对(0)
- # 3楼 2018-02-11 13:28 hapjin

@ 刘建平Pinard
明白了，谢谢博主，向博主学习，哈哈.....

支持(0) 反对(0)
- # 4楼 2018-03-05 01:03 木易晚成

留个脚印，感谢

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：

- 写给自学者的入门指南
 - 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...