HMM——三个基本问题

上一篇我们讲了什么是 HMM。

和我们之前学习过的几个模型对比，HMM 挺别扭的。其他模型都是直接把特征对应成一个结果，当然，这个结果的取值本身可以是连续的（回归模型）或者离散的（分类模型），不过整个过程里涉及的也就是输入变量（特征）和预测结果两部分数据。

HMM 却是变量自己就分成两类：状态变量和观测变量，而且，模型的运行过程也是来来回回在这两类变量之间转圈圈，这到底有什么用呢？

三个基本问题

在实际运用中，HMM 有三个基本问题。

概率计算问题

问题名称：概率计算问题，又称评价（Evaluation）问题。

已知信息：

- 模型 $\lambda = [A, B, \pi]$
- 观测序列 $O = (o_1, o_2, ..., o_T)$

求解目标：计算在给定模型 λ 下，已知观测序列 O 出现的概率：$P(O|\lambda)$ 。也就是说，给定观测序列，求它和评估模型之间的匹配度。

预测问题

问题名称：预测问题，又称解码（Decoding）问题。

已知信息：

- 模型 $\lambda = [A, B, \pi]$
- 观测序列 $O = (o_1, o_2, ..., o_T)$

求解目标：计算在给定模型 λ 下，使已知观测序列 O 的条件概率 $P(O|S)$ 最大的状态序列 $S = (s_1, s_2, ..., s_T)$ 。即给定观测序列，求最有可能与之对应的状态序列。

学习问题

问题名称：学习（Learning）问题又称训练（Training）问题。

已知信息：

- 观测序列 $O = (o_1, o_2, ..., o_T)$
- **或许**也会给定与之对应的状态序列： $S = (s_1, s_2, ..., s_T)$

求解目标：估计模型$\lambda = [A, B, \pi]$参数，使得该模型下观测序列概率 $P(O|\lambda)$ 最大。也就是训练模型，使其最好地描述观测数据。

前两个问题是模型已经存在之后如何使用模型的问题，而最后一个则是如何通过训练得到模型的问题。

一个例子

抽象描述不太好理解，我们来看一个例子。

背景描述

小马利亚的统治者是宇宙公主（就是下面这位）：

宇宙公主的妹妹月亮公主化身梦魇之月（下面这位），挑战宇宙公主的统治，但是被击败了。最后被囚禁在月亮里。

一千年过去了，小马谷的六个姑娘请求宇宙公主赦免月亮公主。

宇宙公主表示，是否赦免月亮公主应该由神明来决定。

宇宙公主有金银铜三个首饰盒。金首饰盒里有两件红宝石、一件蓝宝石，一件珍珠首饰和一件珊瑚；银首饰盒里有红宝石、珍珠，珊瑚和蓝宝石首饰各一件；铜首饰盒里有珍珠、红宝石和珊瑚首饰各一件。

她准备这样做：把三个首饰盒放在一个轮盘上，随机转动轮盘。停下时，哪个盒子在她眼前，她就从这个首饰盒里随机拿出一件首饰，记录首饰的材质后，把首饰放回原盒中。

宇宙公主将重复上述动作3次，如果3次拿出来的正好依次是红宝石，珍珠和珊瑚，她就赦免月亮公主。

那么请问：月亮公主被赦免的可能性到底有多大呢？

问题分析

我们先来看一下这个问题，按照上面的操作，最终我们希望看到的是一个首饰的观测序列，是：O = (红宝石，珍珠，珊瑚)。

其实，在整个事件中，除了最终的首饰序列之外，还有另一个序列：首饰盒的序列。

首先宇宙公主选了一系列的盒子，然后再从每个盒子里选了一件首饰。

如此一来，首饰盒正好对应状态，而首饰则对应观测。首饰盒序列就是状态序列，首饰序列则是观测序列。

状态序列的不同状态之间是依据时序一对一跳转的，而某一个时刻的观测值也仅与当时的状态值有关。

这种情况下，我们完全可以应用 HMM 来估计 O 出现的概率。

要确定一个 HMM，我们需要两个空间和 λ：

状态空间，也就是状态值的集合是：{金盒子,银盒子,铜盒子}。

观测空间，也就是观测值的集合是：{红宝石,珍珠,珊瑚,蓝宝石}。

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

我们来看看 A、B 和 π 分别是什么。

A 是状态转移矩阵，A 中元素用来反映不同状态之间的跳转概率。

我们现在设定的场景是“轮盘赌”，也就是说一个轮盘随机旋转。

如果每一次停留的位置都是随机的，那么任意一个状态到下面任意一个状态的概率都应该是0.33。

但是为了我们的例子更加清晰，我们假设：这个轮盘是不均衡的。

当前盒子为金时，下一次转到银盒子的概率是0.5，铜盒子为0.4，还是金盒子为0.1；当前为银盒子，则下一轮转到金盒子或者铜盒子的概率都是0.4，还是银盒子的概率为0.2；当前为铜盒子，下一个盒子为金银铜的概率分别为0.5、0.3、0.2。

那么我们用一个表格来反应概率跳转（每一行表示一个当前状态，每一列表示一个下一时刻的状态）：

\	金	银	铜
金	0.1	0.5	0.4
银	0.4	0.2	0.4
铜	0.5	0.3	0.2

我们把金银铜改成数字，分别代表行号和列号：

\	1	2	3
1	0.1	0.5	0.4
2	0.4	0.2	0.4
3	0.5	0.3	0.2

这分明就是一个矩阵嘛，写出来就是：

$$\begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$$

这就是状态转移矩阵：

$$A = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$$

再看各个盒子里面拿不同材质首饰出来的概率：

\	红宝石	珍珠	珊瑚	蓝宝石
金	0.4	0.2	0.2	0.2
银	0.25	0.25	0.25	0.25
铜	0.33	0.33	0.33	0

这个表格对应观测矩阵：

$$B = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.2 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.33 & 0.33 & 0.33 & 0 \end{bmatrix}$$

初始概率分布铜盒子略高，其他两个概率相等：$\pi = (0.3, 0.3, 0.4)^T$ 。

求解三个基本问题的现实意义

在开始时就有了两个空间和一个 $\lambda = (A, B, \pi)$ ， 我们就定义了一个 HMM。

那么运用这个 HMM 解决**概率计算问题**，就可以求出连续拿出三件首饰依次是红宝石、珍珠和珊瑚的可能性有多大。

反过来，通过解决**预测问题**，还可以得到最终观测序列为(红宝石，珍珠，珊瑚)情况下，最有可能出现的状态序列（盒子序列）是什么。

如果假设在开始的时候，我们根本不知道 λ 的取值，则可以通过不断地转动轮盘选盒子，拿首饰，生成一个个状态序列和观测序列。然后用它们作为训练数据，来解决**学习问题**，从而得出 λ。

这里有的同学可能会问，解决这些问题到底有什么意义啊？

学习问题还好说，毕竟可以求λ ， 但是像解决概率计算问题和预测问题，反正都已经知道观测序列了，还去计算它出现的概率干什么呀？

单用这个例子来说，我们可以把上面的情形变得复杂一点——宇宙公主现在提出了两种方案：

【方案-1】如果转三次轮盘，拿出来的首饰依次是红宝石，珍珠和珊瑚，她就赦免月亮公主；

【方案-2】如果转四次，连续4次都拿出红宝石，则赦免月亮公主。

宇宙公主让六匹小马在两个方案之间选择，那么应如何选呢？

这时候就需要分别计算两个序列 $O_1 = (\text{红宝石}, \text{珍珠}, \text{珊瑚})$， $O_2 = (\text{红宝石}, \text{红宝石}, \text{红宝石}, \text{红宝石})$ 的出现概率，然后比大小，选择大的那个所对应的方案。

这三个基本问题到底怎么求解呢？