隐马尔可夫模型可以用于标注，这时状态对应着标记。标注问题是给定观测的序列预测其对应的标记序列。可以假设标注问题的数据是由隐马尔可夫模型生成的。这样我们可以利用隐马尔可夫模型的学习与预测算法进行标注。

下面看一个隐马尔可夫模型的例子。

例 10.1（盒子和球模型） 假设有 4 个盒子，每个盒子里都装有红、白两种颜色的球，盒子里的红、白球数由表 10.1 列出。

表 10.1 各盒子的红、白球数

	盒 子			
	1	2	3	4
红球数	5	3	6	8
白球数	5	7	4	2

按照下面的方法抽球，产生一个球的颜色的观测序列：

- 开始，从 4 个盒子里以等概率随机选取 1 个盒子，从这个盒子里随机抽出 1 个球，记录其颜色后，放回；
- 然后，从当前盒子随机转移到下一个盒子，规则是：如果当前盒子是盒子 1，那么下一盒子一定是盒子 2；如果当前是盒子 2 或 3，那么分别以概率 0.4 和 0.6 转移到左边或右边的盒子；如果当前是盒子 4，那么各以 0.5 的概率停留在盒子 4 或转移到盒子 3；
- 确定转移的盒子后，再从这个盒子里随机抽出 1 个球，记录其颜色，放回；
- 如此下去，重复进行 5 次，得到一个球的颜色的观测序列：

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

在这个过程中，观察者只能观测到球的颜色的序列，观测不到球是从哪个盒子取出的，即观测不到盒子的序列。

在这个例子中有两个随机序列，一个是盒子的序列（状态序列），一个是球的颜色的观测序列（观测序列）。前者是隐藏的，只有后者是可观测的。这是一个隐马尔可夫模型的例子。根据所给条件，可以明确状态集合、观测集合、序列长度以及模型的三要素。

盒子对应状态，状态的集合是：

$$Q=\{\text{盒子 }1,\ \text{盒子 }2,\ \text{盒子 }3,\ \text{盒子 }4\},\quad N=4$$

球的颜色对应观测。观测的集合是：

$$V=\{\text{红},\ \text{白}\},\quad M=2$$

状态序列和观测序列长度 $T=5$。

初始概率分布为

$$\pi=(0.25,\ 0.25,\ 0.25,\ 0.25)^{\mathrm{T}}$$

状态转移概率分布为

$$A=\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

观测概率分布为

$$B=\begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{bmatrix}$$

■

10.1.2 观测序列的生成过程

根据隐马尔可夫模型定义，可以将一个长度为 T 的观测序列 $O=(o_1,o_2,\cdots,o_T)$ 的生成过程描述如下。

算法 10.1（观测序列的生成）

输入：隐马尔可夫模型 $\lambda=(A,B,\pi)$，观测序列长度 T；

输出：观测序列 $O=(o_1,o_2,\cdots,o_T)$。

（1）按照初始状态分布 π 产生状态 i_1；

（2）令 $t=1$；

（3）按照状态 i_t 的观测概率分布 $b_{i_t}(k)$ 生成 o_t；

（4）按照状态 i_t 的状态转移概率分布 $\{a_{i_t i_{t+1}}\}$ 产生状态 i_{t+1}，$i_{t+1}=1,2,\cdots,N$；

（5）令 $t=t+1$；如果 $t<T$，转步 (3)；否则，终止。 ■

10.1.3 隐马尔可夫模型的 3 个基本问题

隐马尔可夫模型有 3 个基本问题：

（1）概率计算问题。给定模型 $\lambda=(A,B,\pi)$ 和观测序列 $O=(o_1,o_2,\cdots,o_T)$，计算在模型 λ 下观测序列 O 出现的概率 $P(O|\lambda)$。