隐藏状态：想得到的属性

从观测值推测出隐藏状态序列

一、模型简介

- 可用于序列标注问题的统计学模型，描述了由隐马尔可夫链随机生成观测序列的过程，属于生成模型。
- 是关于时序的概率模型，描述由一个隐藏的马尔可夫链随机生成不可观测的状态随机序列，再由各个状态生成一个观测值，从而做种产生观测序列的过程。

关键概念

- 状态序列：隐藏的马尔可夫链随机生成的状态的序列。
- 观测序列：每个状态生成一个观测，而由此产生的观测的随机序列称作观测序列。
- 序列的每一个位置又可以看作是一个时刻。

> 状态序列：文字（状态空间）
>
> 观测序列：拼音

- 状态集合为$\mathbb{Q} = \{q_1, q_2, \dots, q_Q\}$，观测值集合为$\mathbb{V} = \{v_1, v_2, \dots, v_V\}$
  - Q和V分别表示状态数量和观测值数量
- 设$\mathbf{I} = (i_1, i_2, \dots, i_T)$是长度为$T$的状态序列，　$\mathbf{O} = (o_1, o_2, \dots, o_T)$是对应的观测序列
  - $i_t \in \{1, \dots, Q\}$是一个随机变量，代表一个可能的状态值 q_{i_t}
  - $o_t \in \{1, \dots, V\}$是一个随机变量，代表一个可能的观测值 v_{o_t}

$$P(i_1, i_2, \dots, i_T, o_1, o_2, \dots, o_T), \qquad t = 1,2,\dots,T$$

二、模型表示

隐马尔可夫模型由初始状态概率 π 、状态转移矩阵 $\mathbf{A}$ 、以及观测概率矩阵 $\mathbf{B}$ 决定。一个隐马尔可夫模型可用三元符号表示：$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$

- 初始状态概率 π 和状态转移矩阵 $\mathbf{A}$ 确定了隐藏的马尔可夫链，生成了不可观测的状态序列；
- 观测概率矩阵 $\mathbf{B}$ 确定了如何从状态生成观测值，与状态序列一起确定了如何产生观测序列。

1、齐次性假设--状态转移矩阵

隐藏马链任意时刻的状态只依赖于前一时刻的状态，与其他时刻状态和观测无关，与时刻t也无关

$$P(i_t|i_1, o_1, \dots, i_{t-1}, o_{t-1}) = P(i_t|i_{t-1}), \qquad t = 1,2, \dots, T$$

- a_{i,j}表示从qi状态转移到qj状态的概率

其中$a_{i,j} = P(i_{t+1} = q_j | i_t = q_i)$，表示在 t 时刻处于状态 q_i 的条件下，在 $t+1$ 时刻转移到 q 的概率

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,Q} \\ a_{1,2} & a_{2,2} & \dots & a_{2,Q} \\ \vdots & \vdots & \vdots & \vdots \\ a_{Q,1} & a_{Q,2} & \dots & a_{Q,Q} \end{bmatrix}$$

2、观测独立性假设--观测概率矩阵

- 任意时刻观测值只和当前时刻马链状态有关，和其他观测和状态无关

$$P(o_t|i_1, o_1, \dots, i_{t-1}, o_{t-1}, i_t) = P(o_t|i_t), \qquad t = 1,2, \dots, T$$

- b_{j}（k）表示从qj状态生成观测值vk的概率

其中 $b_j(k) = P(o_t = v_k | i_t = q_j)$，表示 t 时刻处于状态 q_j 的条件下生成观测值 v_k 的概率

$$\mathbf{B} = \begin{bmatrix} b_1(1) & b_1(2) & \dots & b_1(V) \\ b_2(1) & b_2(2) & \dots & b_2(V) \\ \vdots & \vdots & \vdots & \vdots \\ b_Q(1) & b_Q(2) & \dots & b_Q(V) \end{bmatrix}$$

3、状态初始化\pi

π为初始的状态概率：$\pi = (\pi_1, \pi_2, \dots, \pi_Q)$

- $\pi_i = P(i_1 = q_i)$ 表示开始时刻 $t = 1$ 时处于状态 q_i 的概率
- $\sum_{i=1}^{Q} \pi_i = 1$

三、运算

1、生成观测序列

- 输入：隐马尔可夫模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$和观测序列长度$T$
- 输出：观测序列$\mathbf{O} = (o_1, o_2, \ldots, o_T)$
- 算法步骤：
  - 按照初始状态分布 π 产生状态 i_1
  - 令 $t = 1$ ，开始迭代。迭代条件：$t \leq T$。迭代步骤为：
    - 按照状态 i_t 的观测概率分布 $b_j(k)$ 生成观测值 o_t
    - 按照状态 i_t 的状态转移分布 $a_{i,j}$ 产生状态 i_{t+1}
    - 令 $t = t + 1$

2、基本问题

1、学习问题（极大似然）

- 已知序列，求未知的参数使序列可能性最大
- 如果训练数据有观测数据和对应的状态序列，则可监督学习来学习隐马

> 搜狗新闻语料，python包转为拼音，得到对应关系

- 如果训练数据仅有观测序列，无监督学习

2、概率计算问题

> 输入法项目用不到

给定模型参数，求某一观测序列产生的概率

3、预测解码问题

给定模型参数，求某一观测序列对应的最大可能状态序列

- 学习问题
  - 已知观测序列$\mathbf{O}=(o_1,o_2,\dots,o_T)$，估计模型$\lambda=(\mathbf{A},\mathbf{B},\pi)$的参数，使得该模型下观测序列的概率 $P(\mathbf{O};\lambda)$最大。
- 概率计算问题
  - 给定模型 $\lambda=(\mathbf{A},\mathbf{B},\pi)$ 和观测序列 $\mathbf{O}=(o_1,o_2,\dots,o_T)$，计算观测序列 $\mathbf{O}$ 出现的概率 $P(\mathbf{O};\lambda)$
- 预测问题（解码问题）
  - 已知模型$\lambda=(\mathbf{A},\mathbf{B},\pi)$ 和观测序列 $\mathbf{O}=(o_1,o_2,\dots,o_T)$，求该观测序列对应的最可能的状态序列 $\mathbf{I}=(i_1,i_2,\dots,i_T)$

四、例题-学习问题-监督学习

- 用频率估计概率得到三个重要参数

- 假设数据集为$\mathbb{D}=\{(\mathbf{O}_1,\mathbf{I}_1),(\mathbf{O}_2,\mathbf{I}_2),\dots,(\mathbf{O}_N,\mathbf{I}_N)\}$，其中：
  - $\mathbf{O}_1,\dots,\mathbf{O}_N$为$N$个观测序列； $\mathbf{I}_1,\dots,\mathbf{I}_N$为对应的$N$个观测序列。
  - 序列$\mathbf{O}_k$，$\mathbf{O}_k$的长度为T_k。
- 估计转移概率 $a_{i,j}$
  - 设样本中前一时刻处于状态 q_i 、且后一时刻处于 q_j 的频数为 $A_{i,j}$，则转移概率 $a_{i,j}$ 的估计是：

$$a_{i,j}=\frac{A_{i,j}}{\sum_{u=1}^{Q}A_{i,u}},\qquad i=1,2,\dots,Q;\ \ j=1,2,\dots,Q$$

- 估计观测概率 $b_j(k)$
  - 设样本中状态为 q_j 且其对应观测值为 v_k 的频数为 $B_{j,k}$，则状态为 q_j 并且观测值为 v_k 的概率 $b_j(k)$ 的估计为：

$$b_j(k)=\frac{B_{j,k}}{\sum_{v=1}^{V}B_{j,v}},\qquad j=1,2,\dots,Q;\ \ k=1,2,\dots,V$$

- 估计初始状态概率 π_i
  - 设样本中初始时刻（$t=1$）处于状态 q_i 的频数为 C_i，则初始状态概率 π_i 的估计为：

$$\pi_i=\frac{C_i}{\sum_{j=1}^{Q}C_j},\qquad i=1,2,\dots,Q;$$

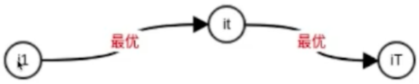
$$\mathrm{argmax}_{i_1,i_2,\dots,i_T}P(i_1,i_2,\dots,i_T,o_1,o_2,\dots,o_T)=\mathrm{argmax}_{i_1,i_2,\dots,i_T}\pi_{i_1}\sum_{t=1}^{T}a_{i_t,i_{t+1}}b_{i_t}(o_t)$$

计算复杂度：$O(T\times Q^T)$

- 下山：2拓展到q，动态规划降低复杂度

五、预测问题---维特比算法

- 根据动态规划原理，最优路径具有这样的特性：如果最优路径在时刻 t 通过结点 i_t^*，则这一路径从结点 i_t^* 到终点 i_T^* 的部分路径，对于从 i_t^* 到 i_T^* 的所有可能路径来说，也必须是最优的。



- 只需要从时刻 $t=1$ 开始，递推地计算从时刻 1 到时刻 t 且时刻 t 状态为 i $(i=1,2,\dots,Q)$ 的各条部分路径的最大概率（以及取最大概率的状态）。于是在时刻 $t=T$ 的最大概率即为最优路径的概率 P^* ，最优路径的终结点 i_T^* 也同时得到。
- 之后为了找出最优路径的各个结点，从终结点 i_T^* 开始，由后向前逐步求得结点 $i_{T-1}^*,\dots,i_1^*$，得到最优路径 $\mathbf{I}^*=(i_1^*,i_2^*,\dots,i_T^*)$。

- 符号定义
  - t 时刻状态为 q_i 的所有可能的路径 $(i_1,i_2,\dots,i_t)$ 中概率最大值为：

$$\delta_t(i)=\max_{i_1,\dots,i_{t-1}}P(i_1,\dots,i_{t-1},i_t=q_i,o_1,\dots,o_t),\qquad i=1,2,\dots,Q$$

  - 则根据定义，得到变量 δ 的递推公式

$$\delta_t(i)=\max_{i_1,\dots,i_{t-1}}P(i_1,\dots,i_{t-1},i_t=q_i,o_1,\dots,o_t)=\max_{1\le j\le Q}\delta_{t-1}(j)\times a_{j,i}\times b_i(o_t)$$
$$i=1,2,\dots,Q;\quad t=2,\dots,T$$

  - t 时刻状态为 q_i 的所有单个路径中概率最大的路径的第 $t-1$ 个结点为：

$$\Psi_t(i)=\operatorname{argmax}_{1\le j\le Q}\delta_{t-1}(j)a_{j,i}\ ,\qquad i=1,2,\dots,Q$$

- 输入：隐马尔可夫模型$\lambda=(\mathbf{A},\mathbf{B},\pi)$，观测序列$\mathbf{O}=(o_1,o_2,\dots,o_T)$
- 输出：最优的状态路径 $\mathbf{I}^*=(i_1^*,i_2^*,\dots,i_T^*)$
- 算法流程：
  - 初始化：　$\delta_1(i)=\pi_i b_i(o_1),\Psi_1(i)=0,\qquad i=1,2,\dots,Q$
  - 递推：　$\delta_t(i)=\max_{1\le j\le Q}\delta_{t-1}(j)\times a_{j,i}\times b_i(o_t)$
    $\Psi_t(i)=\operatorname{argmax}_{1\le j\le Q}\delta_{t-1}(j)a_{j,i}\qquad i=1,2,\dots,Q;\quad t=2,\dots,T$
  - 终止：　$P^*=\max_{1\le i\le Q}\delta_T(i),\qquad i_T^*=\operatorname{argmax}_{1\le j\le Q}\delta_{t-1}(j)a_{j,i}$
  - 最优路径回溯：　$i_t^*=\Psi_{t+1}(i_{t+1}^*),\ \ t=T-1,\dots,1$
  - 获得最优路径 $\mathbf{I}^*=(i_1^*,i_2^*,\dots,i_T^*)$。

做拼音时，要存top 10序列和相应的概率