机器学习笔记–隐马尔可夫模型 *(HMM)*

空修菜

1 隐马尔可夫模型 (HMM)

1. HMM 模型可以解决什么问题. 隐马尔可夫模型一般用于自然语言处理 (NLP) 中的标注问题. 即分词、词性标注、实体识别等方面.

   (1) 分词. 将一个句子切开, 切分完成的句子是一个句子向量, 每一个元素都是一个词;

   (2) 词性标注. 即判断一个词是形容词、名词还是介词等;

   (3) 实体识别. 即识别出句子中的人名、机构名、公司名等实体名称.

2. HMM 模型是生成模型. HMM 模型生成一个不可观测的状态链, 然后再由该状态链生成一个可以观测到观测链. 表示的是状态序列链和观测序列的联合分布.

1.1 HMM 模型的概念

1. HMM 模型的定义. HMM 模型是一个关于时许的概率模型. 它描述了一个由隐藏的马尔可夫链随机生成的不可观测的状态随机序列, 再由每一个状态生成一个观测从而生成一个观测序列的过程.

2. 一个 HMM 模型包含由以下概念.

   (1) **状态集合和观测集合**. 状态集合是不见 (观测) 的状态的所有可能的值, 即状态序列可能的取值. 观测集合则是每一个状态可能对应的观测值, 比如 q_1 对应 v_1. 假设有 N 种可能的状态, 有 M 个可能的观测可能, 则

   $$Q = \{q_1, q_2, \cdots, q_N\} \quad V = \{v_1, v_2, \cdots, v_M\}.$$

   (2) **状态序列和观测序列**. 假设一个马尔可夫链共有 T 个时刻, 即从 $t = 1$ 到 $t = T$. 则该马尔可夫过程对应的状态序列 I 和观测序列 O 分别为:

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

其中, $i_k \in Q,\ o_r \in V,\ 1 \le k \le N,\ 1 \le r \le M$. 状态序列 I 是隐藏的, 能看到的是隐藏状态序列对应的观测序列.

(3) **转移概率矩阵和观测概率矩阵 (发射矩阵)**. 转移状态矩阵 A 描述的是时刻 t 的状态 q_i, 转移到下一个时刻 $t+1$ 的状态 q_j 的概率, 由于状态有 N 种可能, 转移概率矩阵 A 是 $N \times N$ 的.

观测概率矩阵 B 的元素是在 t 时刻, 状态 j 对应 t 的观测值 o_k 的概率.

$$A = [a_{ij}]_{N\times N}, \quad B = [b_j(k)]_{N\times M},$$

其中,

$$a_{ij} = P(i_{t+1} = q_j \mid i_t = q_i), \quad 1 \le i, j \le N.$$
$$b_j(k) = P(o_t = v_k \mid i_t = q_j),\ 1 \le j \le N, 1 \le k \le M.$$

(4) **初始状态概率向量**. 初始状态概率向量是在 $t = 1$ 时, 状态序列的元素 i_1 的可能取值的概率,

$$\pi = (\pi_1, \pi_2, \cdots, \pi_T),$$

其中, $\pi_i = P(i_1 = q_i), 1 \le i \le N$.

(5) 隐马尔可夫模型 λ 由: 初始状态概率向量 π、状态转移概率矩阵 A、观测概率矩阵决定 B,

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

3. **HMM 模型的假设**. HMM 有两个假设:

(1) 马尔可夫齐次性. 马尔可夫模型的下一个状态 i_{t+1} 只与前一个状态 i_t 有关,

$$P(i_{t+1} \mid i_t, o_t, \cdots, i_1, o_1) = P(i_{t+1} \mid i + t).$$

(2) 观测独立性. t 时刻的观测 o_t 只与 t 时刻的状态 i_t 有关,

$$P(o_t \mid i_T, o_T, \cdots, i_1, o_1) = P(o_t \mid i_t = q_i).$$

4. **HMM 模型的三个基本问题**.

(1) 概率计算问题. 给定模型 $\lambda = (A, B, \pi)$ 和观测序列 $O = (o_1, o_2, \cdots, o_T)$, 计算观测序列 O 在模型 λ 中出现的概率 $P(O \mid \lambda)$. 可应用于中文分词.

(2) 学习问题 (参数估计). 给出观测序列 $O=(o_1,o_2,\cdots,o_T)$, 估计模型的参数.

(3) 预测问题. 给定模型 $\lambda=(A,B,\pi)$ 和观测序列 $O=(o_1,o_2,\cdots,o_T)$, 求使得 $P(I\mid O)$ 最大的状态序列 I, 即求隐藏状态序列. 可以使用于词性标注、实体识别等, 在词性标注中, 状态序列是动词、名次、介词等词性标签, O 则是一个句子. 预测结果是该句子的词性.

1.2 概率计算问题

1. 前向概率的定义. λ 是 HMM 模型. 到时刻 t 时, 部分观测序列为 $o_1,o_2,\cdots,o_t$ 且 t 时刻状态为 q_i 的概率为 $\alpha_t(i)$:

$$\alpha_t(i)=P(o_1,o_2,\cdots,o_t,i_t=q_i\mid\lambda).$$

有了以上的 t 时刻, 观测序列为 $o_1,o_2,\cdots,o_t$, 且状态为 q_j 的概率 $\alpha_t(j)$, 可以递归地得出下一个时刻的递归关系:

$$\begin{aligned}\alpha_{t+1}(i)&=P(o_1,o_2,\cdots,o_t,o_{t+1},i_t=q_j,i_{t+1}=q_i\mid\lambda)\\&=\alpha_t(j)a_{ji}b_j(o_{t+1}).\end{aligned}$$

2. **前向概率算法流程**

输入: HMM 模型 λ; 观测序列 O.

输出: 观测序列的概率 $P(O\mid\lambda)$.

(1) 初始化 $\alpha_1(i)$:

$$\alpha_1(i)=\pi_ib_i(o_1),\quad 1\le i\le N.$$

(2) 递推, 对 $t=2,3,\cdots,T-1$,

$$\alpha_{t+1}(i)=\Big(\sum_{j=1}^{N}\alpha_t(j)a_{ji}\Big)b_i(o_{t+1}),\quad 1\le i\le N.$$

(3) 终止:

$$P(O\mid\lambda)=\sum_{i=1}^{N}\Big(\sum_{j=1}^{N}\alpha_{t-1}(j)a_{ji}\Big)b_i(o_t)=\sum_{i=1}^{N}\alpha_t(i).$$

- 由前向概率的计算方式可知, 下一时刻的前向概率, 也就是观测序列为 $o_1,\cdots,o_t,o_{t+1}$ 状态由 $i_t=q_j$ 转变为 $i_{t+1}=q_i$ 的概率 $\alpha_{t+1}(i)$.

- 比如下一刻的状态是 i, 那么就需要考虑前一时刻 t 的所有状态转变为状态 i 的可能.

3. 后向概率的定义. λ 是 HMM 模型. 到时刻 t 时, t 时刻之后的部分观测序列为 $o_{t+1}, o_{t+2}, \cdots, o_T$ 且 t 时刻状态为 q_i 的概率为 $\beta_t(i)$:

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \cdots, o_T \mid i_t = q_i, \lambda).$$

假设 t 时刻的状态是 i, $t+1$ 时刻的状态是 j. 由此, 可以得到后向的 t 时刻的递推公式:

$$\begin{aligned}\beta_t(i) &= P(o_{t+1}, o_{t+2}, \cdots, o_T \mid i_t = q_i, i_{t+1} = q_j, \lambda) \\ &= \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j).\end{aligned}$$

4. **后向概率算法流程**.

输入: HMM 模型 λ; 观测序列 O.

输出: 观测序列的概率 $P(O \mid \lambda)$.

(1) 初始化:

$$\beta_T(i) = 1, \quad 1 \le i \le N.$$

(2) 对 $t = T-1, T-2, \cdots, 1$,

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \quad 1 \le i \le N.$$

(3) 终止

$$P(O \mid \lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(i).$$

1.3 学习问题

1.4 预测问题和维比特算法

假设状态集合 Q 只有三种取值, 即 $Q = \{q_1, q_2, q_3\}$, 令 $\mu_t(i)$ 是 t 时刻状态是 i 的概率. θ_t 是 $\mu_t(i)$ 中的最大值.

1. 维特比算法使用的动态规划算法. 如 Figure 1.4, 维特比算法的大致思路是:

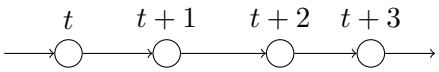

Figure 1.1: 状态链

(1) 在 t 时刻, 计算这一时刻每一个状态的概率 $\mu_t(i)$, $1 \le i \le 3$, 并且取其中的最大概率值对应的状态为当前时刻的最终状态,

$$\theta_t = \max\{\mu_t(i),\ 1 \le i \le 3\},$$

所以, 此时要确定对应的最大概率的状态 θ_t 的前一时刻的状态:

$$\arg\theta_t,$$

即要找到上一时刻 $t-1$ 的最大概率状态.

(2) 计算 $t+1$ 时刻的最大概率状态 θ_{t+1}, 需要先将 $t+1$ 时刻每一个可能的状态都计算一遍, 然后从其中取最大值. 即需要计算 $\mu_{t+1}(i)$, $1 \le i \le 3$. 同时, 每一个 $\mu_{t+1}(i)$ 依赖于上一时刻 t 的 θ_t.

(3) 当计算到最后一个时刻 T 时, 得到 θ_T 以及 $\arg\theta_T$.

(4) 从 T 时刻开始往前找, 逐步确定最优状态链. $\arg\theta_T$ 是 T 时刻时, $T-1$ 时刻的状态; $\arg\theta_{T-1}$ 是 $T-1$ 时刻时, $T-2$ 时刻的状态,$\cdots$,$\arg\theta_2$ 是 2 时刻时, 1 时刻的状态. 因此, 最优的概率状态链条是

$$(\arg\theta_1, \arg\theta_2, \cdots, \arg\theta_T),$$

因为 $t=1$ 时, 没有前一时刻, 所以令 $\arg\theta_1 = 0$.

2. 关于最大概率状态的定义和前一时刻状态的定义.

(1) 在时刻 t 时, 状态为 i 的所有单个路径 $(i_1, i_2, \cdots, i_t)$ 中的最大概率值是:

$$\delta_t(i) = \max_{i_1, i_2, \cdots, i_{t-1}} P(i_t = i, i_{t-1}, \cdots, t_1, o_t, \cdots, o_1 \mid \lambda).$$

$$\delta_{t+1}(i) = \max_{1 \le j \le N} \Big(\delta_t(j) a_{ji}\Big) b_i(o_{t+1}),\ 1 \le i \le N,\ 1 \le t \le T-1.$$

REMARK 1. *$\delta_{t+1}(i)$ 的意思是: 若 $t+1$ 时刻的状态是 i, 那么从 t 时刻的最可能状态 j, 转移为 $t+1$ 时刻的状态 j, 且观测值为 o_t 的最大概率.*

(2) 在时刻 t 时, 状态为 i 的所有单个路径 $(i_1, i_2, \cdots, i_t)$ 中的最大概率值路径的第 $t-1$ 状态 (节点) 是:

$$\Psi_t(i) = \arg \max_{1 \le j \le N} \Big(\delta_{t-1}(j) a_{ji} \Big), \ 1 \le i \le N.$$

REMARK 2. $\Psi_t(i)$ 的意思是: t 时刻的状态是 i, 然后计算 $t-1$ 时刻的所有状态 j 转变为 t 时刻的状态 i 的转移概率, 并且取其中最大的转移概率对应的状态为 $t-1$ 时刻的状态.

3. **维特比算法流程**.

输入: HMM 模型 λ; 观测序列 $O = (o_1, o_2, \cdots, o_T)$.

输出: 最优路径 $I^* = (i_1^*, i_2^*, \cdots, i_T^*)$.

(1) 初始化:

$$\delta_1(i) = \pi_i b_i(o_1), \ \Psi_1(i) = 0, \ 1 \le i \le N.$$

(2) 递推, 对 $t = 2, 3, \cdots, T$:

$$\delta_t(i) = \max_{1 \le j \le N} \Big(\delta_{t-1}(j) a_{ji} \Big) b_i(o_t),$$
$$\Psi_t(i) = \arg \max_{1 \le j \le N} \Big(\delta_{t-1}(j) a_{ji} \Big), \quad 1 \le i \le N.$$

(3) 终止:

$$P^* = \max_{1 \le i \le N} \delta_T(i),$$
$$i_T^* = \arg \max_{1 \le i \le N} \delta_T(i).$$

(4) 最优回溯路径, 对 $t = T-1, \cdots, 1$:

$$i_t^* = \Psi_{t+1}(i_{t+1}^*),$$

最优路径为 $I^* = (i_1^*, i_2^*, \cdots, i_T^*)$.