HMM中的维特比算法

机器学习的一些算法 动态规划

1. 算法

维特比算法实际上常常被用来求解HMM模型的预测问题。即用动态规划求解概率最大（最优路径）。最后求解出来的是一个状态序列，比如在中文分词中，最后出来的可能是[BMESSS]这样子一个状态序列表。

根据动态规划的原理，最优路径具有这样的特性，如果最优路径在时刻t通过节点i_t^*，那么这一路径中的i_t^*到终点节点i_T^*的这一小段路径记作l1，从节点i_t^*到终点节点i_T^*所有路径记作L，那么l1必然是L中最优的一条路径。因为如果不是这样，那么就存在一条更好的路径l2，而我们将这条路径l2和节点i_1^*到节点i_t^*连接起来就是一条最优路径，这和我们的假设矛盾。

根据这个原理，我们只需要从t=1开始，递推的计算在时刻t状态为i的各条部分路径的最大概率。直至到了时刻t=T，状态为i的各条路径的最大概率。时刻t=T的最大概率即为最优路径的概率p^*，最优路径的终节点i_T^*也同时得到。之后为了找到各个最优的节点，从终节点i_T^*开始，由后向前，逐步得到前面的各个节点，最终得到最优路径 $I^* = \left(i_1^*, i_2^*, \cdots, i_T^*\right)$，这就是维特比算法！

下面我们来看一下递推式都是什么样子的。
首先我们引入一个变量δ，定义为在时刻为t，状态为i的所有单个路径中，概率的最大值。
那么在t时刻，其可以写成：

$$\delta_t(i) = \max_{i_1, i_2, \cdots, i_{t-1}} P\left(i_t = i, i_{t-1}, \cdots, i_1, o_t, \cdots, o_1 | \lambda\right), \quad i = 1, 2, \cdots, N$$

那么在t+1时刻，

$$\delta_{t+1}(i) = \max_{i_1, i_2, \cdots, i_t} P\left(i_{t+1} = i, i_t, \cdots, i_1, o_{t+1}, \cdots, o_1 | \lambda\right) = \max_{1 \leqslant j \leqslant N}\left[\delta_t(j) a_{ji}\right] b_i\left(o_{t+1}\right), \quad i = 1, 2, \cdots, N; t = 1, 2, \cdots, T-1$$

定义在时刻t状态为i的所有单个路径$(i_1, i_2, \cdots, i_{t-1}, i)$中，概率最大的路径的第t-1个节点为：

$$\psi_t(i) = \arg\max_{1 \leqslant j \leqslant N}\left[\delta_{t-1}(j) a_{ji}\right], \quad i = 1, 2, \cdots, N$$

这样我们就得到的递推式。

2. 一个例子

考察盒子球模型，其中状态集合Q={1，2，3}，观测集合V={红，白}，观测向量O="红白红"，试求最优状态序列。
其中已知HMM模型为：

$$\pi = \begin{pmatrix} 0.2 \\ 0.4 \\ 0.4 \end{pmatrix} \quad A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix} \quad B = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix}$$

步骤一：初始化
在t=1时，对于每一个状态i，求状态为i观测到o1=红的概率，记此概率为δ_1(t)
则有：

$$\delta_1(i) = \pi_i b_{io_1} = \pi_i b_{i红}$$

带入具体数值算得：

$$\delta_1(1) = 0.2 \times 0.5 = 0.1$$

$$\delta_1(2) = 0.16$$
$$\delta_1(3) = 0.28$$

并且我们定义$\psi_1(i) = 0$

步骤二：t=2时
在t=2时，对每个状态i，求在t=1时状态为j，观测为红，并且在t=2时状态为i，观测为白的路径的最大概率，记概率为$\delta_2(\mathrm{t})$，则：

$$\delta_{t+1}(i) = \max_{1 \leq j \leq 3} \left(\delta_1(j) a_{ji}\right) b_{io_2}$$
$$= \max_{1 \leq j \leq 3} \left(\delta_1(j) a_{ji}\right) b_{i白}$$

拿一个展开算一下：
当j=1时，$\delta_2(1) = \delta_1(1) a_{11} b1白 = 0.1 \times 0.5 \times 0.5 = 0.025$
当j=2时，$\delta_2(1) = \delta_1(2) a_{21} b1白 = 0.16 \times 0.3 \times 0.5 = 0.024$
当j=3时，$\delta_2(1) = \delta_1(3) a_{31} b1白 = 0.28 \times 0.2 \times 0.5 = 0.028$

$$\delta_2(1) = 0.028, \quad \psi_2(1) = 3$$

$$\delta_2(2) = 0.0504, \quad \psi_2(2) = 3$$
$$\delta_2(3) = 0.042, \quad \psi_2(3) = 3$$

同理，我们可以求得t=3时：

$$\delta_3(1) = 0.00756, \quad \psi_3(1) = 2$$
$$\delta_3(2) = 0.01008, \quad \psi_3(2) = 2$$
$$\delta_3(3) = 0.0147, \quad \psi_3(3) = 3$$

步骤三：最优概率和最优路径终点
很明显，最优路径概率为

$$P^* = \max_{1 \leqslant i \leqslant 3} \delta_3(i) = 0.0147$$

最优路径终点为：

$$i_3^* = \arg\max_i \left[\delta_3(i)\right] = 3$$

步骤四：逆向寻找其他节点
由最优终点，逆向寻找其他节点：
当t=3时，$i_2^* = \psi_3\left(i_3^*\right) = \psi_3(3) = 3$
当t=2时，$i_1^* = \psi_2\left(i_2^*\right) = \psi_2(3) = 3$

从而得到序列是(3,3,3)

3. 算法步骤

输入：模型$\lambda = (A, B, \pi)$和观测$O = \left(o_1, o_2, \cdots, o_T\right)$，
输出：最优路径$I^* = \left(i_1^*, i_2^*, \cdots, i_T^*\right)$
（1）初始化：

$$\delta_1(i) = \pi_i b_i\left(o_1\right), \quad i = 1, 2, \cdots, N$$
$$\psi_1(i) = 0, \quad i = 1, 2, \cdots, N$$

（2）递推，对$t=2,3,\cdots,T$

$$\delta_t(i)=\max_{1\leq j\leq N}\left[\delta_{t-1}(j)a_{ji}\right]b_i\left(o_t\right),\quad i=1,2,\cdots,N$$

$$\psi_t(i)=\arg\max_{1\leqslant j\leqslant N}\left[\delta_{t-1}(j)a_{ji}\right],\quad i=1,2,\cdots,N$$

（3）终止

$$P^*=\max_{1\leqslant i\leqslant N}\delta_T(i)$$
$$i_T^*=\arg\max_{1\leqslant i\leqslant N}\left[\delta_T(i)\right]$$

（4）最优路径回溯，对$t=T-1,T-2,\cdots,1$

$$i_t^*=\psi_{t+1}\left(i_{t+1}^*\right)$$

最终求得最优路径$I^*=\left(i_1^*,i_2^*,\cdots,i_T^*\right)$。

参考

维基百科：维特比算法
统计学习方法