

2 维隐马尔可夫模型的基本问题求解

李玉

(北京市多媒体与智能软件技术重点实验室, 北京工业大学计算机学院, 北京 100022)

摘要: 研究了 2 维隐马尔可夫模型的三个基本问题, 包括概率评估问题、最优状态问题和参数估计问题. 通过把 2 维隐马尔可夫模型行或者列上的状态序列看作一个马尔可夫模型, 从理论上分别给出了解决这三个基本问题的新算法; 计算机仿真对新算法的实现和运行作了进一步的说明.

关键词: 隐马尔可夫模型; 状态矩阵; 观察矩阵; 基本问题求解

中图分类号: TP181 **文献标识码:** A **文章编号:** 0372-2112 (2004) 12-1832-06

Basic Problems Solving for Two-Dimensional Hidden Markov Models

LI Yu-jian

(College of Computer Science and Technology, Beijing University of Technology, Beijing 100022, China)

Abstract: The three basic problems of two-dimensional (2D) hidden Markov models (HMMs) are studied, including probability evaluation, optimal states and parameter estimation. By using the idea that the sequences of states on columns or rows of a 2D HMM can be seen as states of a 1D HMM, several new analytic formulae for solving these three problems are theoretically derived and further demonstrated by computer simulation.

Key words: hidden Markov models; state matrices; observation matrices; basic problems solving

1 引言

隐马尔可夫模型 (Hidden Markov model, 简称 HMM) 自从 Baum 等人在 19 世纪 60 年提出以来^[1-3], 就在语音识别、字符识别、金融数据分析和图像处理等领域的研究中已经获得了广泛的应用^[4-9]. 由于它只能直接对序列数据建模, 因此可以被看作是 1 维的. 如果要用 1 维隐马尔可夫模型来处理矩阵数据 (比如图像数据), 就必须选择由一些特征构成的序列数据来表达矩阵数据, 而这常常会损失掉矩阵数据中所蕴含的某些重要信息, 因为矩阵数据在本质上是 2 维的, 它们和序列数据相比在拓扑结构上通常都存在着本质上的区别. 因此, 为了更好地利用隐马尔可夫模型的思想对矩阵数据建模, 一些研究工作者开展了对 1 维隐马尔可夫模型进行 2 维推广的工作, 比如 Agazzi 和 Kuo 提出的伪 2 维隐马尔可夫模型 (pseudo 2D HMMs)^[10-11]. 伪 2 维隐马尔可夫模型本质上是两个 1 维隐马尔可夫模型的嵌套, 并不是一个真正的 2 维隐马尔可夫模型.

真正的 2 维隐马尔可夫模型 (2D HMM) 最先由 Devijver 提出并应用于图像分割^[12], 通常又被称为隐马尔可夫网格随机场 (Hidden Markov mesh random field, 简称为 HMMRF), 其特征在于它的状态矩阵构成一个马尔可夫网格随机场 (Markov mesh random field, 简称为 MMRF)^[13]. 类似于 1 维隐马尔可夫

模型, 对 2 维隐马尔可夫模型来说, 最重要的也是解决其中的三个基本问题, 即概率评估 (Probability evaluation) 问题、最优状态 (Optimal states) 问题和参数估计 (Parameter estimation) 问题. 虽然这三个基本问题可以利用确定松弛算法 (Deterministic relaxation algorithm) 和向前看技术 (Look-ahead technique)^[14-15] 等方法近似求解, 但是一直没有得到很好的解决. 文献 [16] 通过把 2 维隐马尔可夫模型斜对角线上的状态序列看作一个状态可变的 1 维马尔可夫模型, 利用前向后向类似算法 (Forward-backward like algorithm) 和可变状态 Viterbi 算法 (Variable-state Viterbi algorithm) 首次讨论了 2 维 2 阶隐马尔可夫模型的分析解 (Analytic solution), 同时指出 2 维隐马尔可夫模型行或者列上的状态序列也具有马尔可夫模型的特征, 但是没有作进一步的研究, 本文将利用这一点从理论上给出求解 2 维 2 阶和 3 阶隐马尔可夫模型基本问题的新算法.

2 2 维隐马尔可夫模型的基本结构

设 $I_{M,N} = \{(m, n) | 1 \leq m \leq M, 1 \leq n \leq N\}$ 是一个大小为 $M \times N$ 的整数网格. 对于 $I_{M,N}$ 中任意两个格点 (m_c, n_c) 和 (m, n) , 如果 $m_c < m$ 或者 $m_c = m$ 且 $n_c < n$, 那么称格点 (m_c, n_c) 在格点 (m, n) 之前, 用 $(m_c, n_c) < (m, n)$ 表示. 比如, 在图 1 (a) 中, 非白色格点都在 (m, n) 之前. 此外, 用 $\hat{<}$ 表示 $<$ 的旋转关系, 也就是说, 如果 $n_c < n$ 或者 $n_c = n$ 且 $m_c < m$, 那么

$(m, n) \hat{=}$ (m, n) . 比如, 在图 1 (b) 中, 非白色格点都 $\hat{=}$ (m, n) . 在图 1 (c) 中, 非白色格点都 $<$ $\hat{=}$ (m, n) .

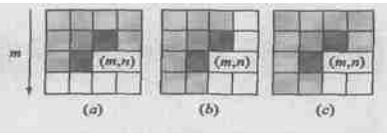


图 1 状态转移之间的马尔可夫性质

令 $W_{m,n}^a, W_{m,n}^b$ 和 $W_{m,n}^c$ 分别表示图 1(a), (b) 和 (c) 中所有非白色格点的集合, 考虑下面两组条件:

$$P(q_{m,n} | q_{k,l}, (k,l) \in W_{m,n}^a) = \begin{cases} P(q_{m,n} | q_{m-1,n}, q_{m,n-1}), & m > 1, n > 1 \\ P(q_{m,n} | q_{m-1,n}), & m > 1, n = 1 \\ P(q_{m,n} | q_{m,n-1}), & m = 1, n > 1 \end{cases} \quad (1)$$

$$P(q_{m,n} | q_{k,l}, (k,l) \in W_{m,n}^b) = \begin{cases} P(q_{m,n} | q_{m-1,n}, q_{m-1,n-1}, q_{m,n-1}), & m > 1, n > 1 \\ P(q_{m,n} | q_{m-1,n}), & m > 1, n = 1 \\ P(q_{m,n} | q_{m,n-1}), & m = 1, n > 1 \end{cases} \quad (2)$$

那么一个 2 维 2 阶和 3 阶马尔可夫模型 $Q = \{q_{m,n} | (m,n) \in I_{M,N}\}$ 可以分别定义为:

定义 1 Q 是一个 2 维 2 阶马尔可夫模型当且仅当 $S = a, b$ 或 c 时条件(1)成立.

定义 2 Q 是一个 2 维 3 阶马尔可夫模型当且仅当 $S = a, b$ 或 c 时条件(2)成立.

事实上, 当 $S = a, b$ 或 c 时条件(1)之间的等价性以及当 $S = a, b$ 或 c 时条件(2)之间的等价性已经被证明^[13]. 因此, 2 维 2 阶和 3 阶马尔可夫模型的第 1 行状态和第 1 列状态必然是一个马尔可夫链, 即满足:

$$P(q_{1,n} | q_{1,n-1}, q_{1,n-2}, \dots, q_{1,1}) = P(q_{1,n} | q_{1,n-1}), 1 < n \leq N \quad (3)$$

$$P(q_{m,1} | q_{m-1,1}, q_{m-2,1}, \dots, q_{1,1}) = P(q_{m,1} | q_{m-1,1}), 1 < m \leq M \quad (4)$$

一个 2 维隐马尔可夫模型通常由状态矩阵 Q 和观察矩阵 O 组成. 设状态矩阵 $Q = \{q_{m,n} | (m,n) \in I_{M,N}\}$, 如果只考虑 2 阶和 3 阶模型, 那么 Q 是 2 阶和 3 阶马尔可夫模型, 其中 $q_{m,n}$ 称为状态. 如果用 $O = \{o_{m,n} | (m,n) \in I_{M,N}\}$ 表示观察矩阵, 那么 $o_{m,n}$ 完全独立地由状态 $q_{m,n}$ 产生, 称为观察符号. 为方便起见, 引入下面的记号:

(1) L , 模型中状态的数目. 不同的状态分别标记为 $\{1, 2, \dots, L\}$, 在格点 (m, n) 处的状态记为 $q_{m,n}$.

(2) R , 每个状态所能产生的不同观察符号数. 不同的符号分别标记为 $\{w_1, w_2, \dots, w_r\}$, 在格点 (m, n) 处的观察符号记为 $o_{m,n}$.

(3) 2 阶状态转移概率分布 $A_2 = \{a_{i,j,k} | 1 \leq i, j, k \leq L\}$, 其中 $a_{i,j,k} = P(q_{m,n} = k | q_{m-1,n} = i, q_{m,n-1} = j), 1 \leq i, j, k \leq L$.

(4) 3 阶状态转移概率分布 $A_3 = \{a_{i,j,k,l} | 1 \leq i, j, k, l \leq L\}$, 其中 $a_{i,j,k,l} = P(q_{m,n} = l | q_{m-1,n} = i, q_{m-1,n-1} = j, q_{m,n-1} = k), 1 \leq i, j, k, l \leq L$.

(5) 第 1 列状态转移概率分布 $A^V = \{a_{i,j}^V | 1 \leq i, j \leq L\}$, 其

中 $a_{i,j}^V = P(q_{m,1} = j | q_{m-1,1} = i), 1 \leq i, j \leq L$.

(6) 第 1 行状态转移概率分布 $A^H = \{a_{i,j}^H | 1 \leq i, j \leq L\}$, 其中 $a_{i,j}^H = P(q_{1,n} = j | q_{1,n-1} = i), 1 \leq i, j \leq L$.

(7) 观察符号的概率分布 $B = \{b_k(r)\}$, 其中 $b_k(r) = b_k(w_r) = P(o_{m,n} = w_r | q_{m,n} = k), 1 \leq r \leq R$ 定义状态 k 的符号分布, $k = 1, 2, \dots, L$.

(8) 初始概率分布 $0 = \{P_j | 1 \leq j \leq L\}$, 其中

$$P_j = P(q_{1,1} = j), 1 \leq j \leq L$$

综上所述, 一个完整的 2 维 2 阶和 3 阶隐马尔可夫模型可以分别用 $\#_2$ 和 $\#_3$ 简单地表示为:

$$\#_2 = (A^V, A^H, A_2, B, 0), \#_3 = (A^V, A^H, A_3, B, 0)$$

3 2 维隐马尔可夫模型的三个基本问题

与 1 维隐马尔可夫模型类似, 对 2 维隐马尔可夫模型来说, 同样也有三个基本问题需要解决, 即:

问题 1: 给定观察矩阵 O 和模型 $\#_2$ 或 $\#_3$, 如何有效地计算在 $\#_2$ 或 $\#_3$ 的条件下产生 O 的概率 $P(O | \#_2)$ 或 $P(O | \#_3)$? 该问题称为概率评估问题.

问题 2: 给定观察矩阵 O 和模型 $\#_2$ 或 $\#_3$, 如何选择一個相应的状态矩阵 Q , 在某种意义上最优地产生观察矩阵 O ? 该问题称为最优状态问题.

问题 3: 如何调整模型 $\#_2$ 或 $\#_3$ 的参数使 $P(O | \#_2)$ 或 $P(O | \#_3)$ 最大? 该问题称为参数估计问题.

文献[16] 证明了 2 维隐马尔可夫模型斜对角线上的 $\{q_{m,1}, q_{m-1,2}, \dots, q_{1,m}\}$ 构成一个状态可变的 1 维马尔可夫模型, 讨论了 2 维隐马尔可夫模型的分析解, 并指出 2 维隐马尔可夫模型行或者列上的状态也可看作一个 1 维马尔可夫模型, 但未作深入研究. 本节将利用这一点导出解决上述三个基本问题的新算法.

如果给定模型 $\#_2$, 把 Q 的第 n 列状态表示为 $T_n = \{q_{1,n}, q_{2,n}, \dots, q_{M,n}\}, 1 \leq n \leq N$, 那么 T_n 具有下面的基本性质:

$$P(T_n | T_{n-1}, \dots, T_1) = P(q_{1,n}, q_{2,n}, \dots, q_{M,n} | T_{n-1}, \dots, T_1) = P(q_{1,n} | T_{n-1}, \dots, T_1) \# P(q_{2,n} | q_{1,n}, T_{n-1}, \dots, T_1) \# \dots \# P(q_{M,n} | q_{M-1,n}, \dots, q_{1,n}, T_{n-1}, \dots, T_1) = P(q_{1,n} | q_{1,n-1}) P(q_{2,n} | q_{1,n}, q_{2,n-1}) \dots P(q_{M,n} | q_{M-1,n}, q_{M,n-1})$$

这表明 $P(T_n | T_{n-1}, \dots, T_1) = P(T_n | T_{n-1})$, 所以 2 维隐马尔可夫模型各列上的状态 T_n 构成一个 1 维马尔可夫模型; 因此, $\#_2$ 实际上等价于一个 1 维隐马尔可夫模型 $K_2 = (A_{\#_2}^V, B_{\#_2}, 0_{\#_2})$, K_2 的参数详细描述如下:

(1) K_2 的状态数是 L^M , 单个状态记为 II , 其中 $8 = \{II | I = (i_1, i_2, \dots, i_k, \dots, i_M), 1 \leq i_k \leq L, 1 \leq k \leq M\}$.

(2) 每个状态所能产生的不同观察符号数是 R^M , 单个符号记为 WI , 其中

$$E = \{W | W = (w_{r_1}, w_{r_2}, \dots, w_{r_k}, \dots, w_{r_M}), 1 \leq r_k \leq R, 1 \leq k \leq M\}$$

(3) 状态转移概率分布是 $A_{\#_2}^V = \{a_{II'}^V\}$, 其中

$$a_{II'}^V = P(T_{n+1} = I' | T_n = I)$$

$$= a_{i_1, j_1}^H a_{i_1, i_2, j_2} a_{j_2, i_3, j_3} \dots a_{i_{M-1}, i_M, j_M}^H, I, J \text{ I } 8$$

(4) 观察符号概率分布是 $B_{\#_2} = \{b_j(W)\}$, 其中

$$b_j(W) = P(W = (w_{r_1}, w_{r_2}, \dots, w_{r_M}) | T_n = J) \\ = b_{j_1}(w_{r_1}) \# b_{j_2}(w_{r_2}) \dots b_{j_M}(w_{r_M}), W \text{ I } E$$

定义状态 J 产生观察符号的分布 $J = (j_1, j_2, \dots, j_M) \text{ I } 8$.

(5) 初始概率分布 $0_{\#_2} = \{P_j\}$, 其中

$$P_j = P(T_1 = J) = P_{j_1} a_{j_1, j_2}^V a_{j_2, j_3}^V \dots a_{j_{M-1}, j_M}^V, J \text{ I } 8$$

同理可证, 2 维 3 阶隐马尔可夫模型 $\#_3$ 也等价于一个 1 维隐马尔可夫模型 $K_3 = (A_{\#_3}, B_{\#_3}, 0_{\#_3})$, 其中 $B_{\#_3} = B_{\#_2}, 0_{\#_3} = 0_{\#_2}, A_{\#_3} = \{a_{i,j}\}$, 并且

$$a_{i_1, j}^3 = P(T_{n+1} = J | T_n = I) \\ = a_{i_1, j_1}^H a_{i_1, i_2, j_2} a_{j_2, i_3, j_3} \dots a_{i_{M-1}, i_M, j_M}^H, I, J \text{ I } 8$$

用行代替列, 也可以得到 2 维 2 阶 (或 3 阶) 隐马尔可夫模型的一个相应的等价 1 维隐马尔可夫模型. 由于问题的对称性和相似性, 在这里先只考虑利用 $K_2 = (A_{\#_2}, B_{\#_2}, 0_{\#_2})$ 来解决 2 维 2 阶隐马尔可夫模型的三个基本问题. 但是在实际计算时, 应使用行数或列数较小的模型.

3.1 问题 1 的解决

在给定观察矩阵 O 和模型 $\#_2$ 的条件下计算观察矩阵 O 的概率 $P(O | \#_2)$, 等价于在 $K_2 = (A_{\#_2}, B_{\#_2}, 0_{\#_2})$ 的条件下求 $P(W_1, W_2, \dots, W_N | K_2)$, 其中 $W_n = \{o_{1, n}, \dots, o_{M, n}\}$.

把模型 K_2 的前向变量 $A_n^2(I)$ 和后向变量 $B_n^2(I)$ 各定义为 $A_n^2(I) = P(W_1, W_2, \dots, W_n, T_n = I | K_2) \quad (5a)$

$$B_n^2(I) = P(W_{n+1}, W_{n+2}, \dots, W_N | T_n = I, K_2) \quad (5b)$$

$A_n^2(I)$ 和 $B_n^2(I)$ 可以递归计算如下:

(1) 初始化:

$$A_1^2(I) = P(b_1(W_1), I) \text{ I } 8 \quad (6a)$$

$$B_N^2(I) = 1, I \text{ I } 8 \quad (6b)$$

(2) 递归过程:

$$A_{n+1}^2(J) = \left[\sum_{I \text{ I } 8} A_n^2(I) A_{i_1, j}^2 \right] b_j(W_{n+1}), \\ J \text{ I } 8, 1 \leq n < N \quad (7a)$$

$$B_n^2(I) = \sum_{J \text{ I } 8} a_{i_1, j}^2 b_j(W_{n+1}) B_{n+1}^2(J), \\ I \text{ I } 8, n = N-1, N-2, \dots, 1 \quad (7b)$$

最后利用 $A_n^2(I)$ 计算 $P(O | \#_2)$ 如下:

$$P(O | \#_2) = P(W_1, W_2, \dots, W_N | K_2) = \sum_{I \text{ I } 8} A_N^2(I) \quad (8)$$

3.2 问题 2 的解决

有多种可能的方法解决问题 2, 即找到多个与观察矩阵相关联的“最优”状态矩阵. 困难在于最优状态矩阵的定义, 因为有多种可能的标准. 对该问题来说, 比较合理的标准有下面三个:

标准 1: 在给定观察矩阵 O 和模型 $\#_2$ 的条件下, 使整个状态矩阵最优, 即

$$Q^* = \arg \max_Q P(Q | O, \#_2) = \arg \max_Q P(Q | O, K_2) \quad (9)$$

标准 2: 在给定观察矩阵 O 和模型 $\#_2$ 的条件下, 使状态

矩阵的每列最优, 即

$$T_n^* = \arg \max_{I \text{ I } 8} P(T_n = I | O, \#_2) = \arg \max_{I \text{ I } 8} P(T_n = I | O, K_2) \quad (10)$$

标准 3: 在给定观察矩阵 O 和模型 $\#_2$ 的条件下, 使整个状态矩阵的每个格点状态最优, 即

$$q_{m, n}^* = \arg \max_{I \text{ I } 8, k \text{ I } 8} P(q_{m, n} = k | O, \#_2) = \arg \max_{I \text{ I } 8, k \text{ I } 8} P(q_{m, n} = k | O, K_2) \quad (11)$$

3.2.1 基于标准 1 的解决 首先将观察矩阵 O 看作一个关于列的序列, 即 $O = (W_1, W_2, \dots, W_N)$, 然后定义变量:

$$D_n^1(I) = \max P(T_1, T_2, \dots, T_{n-1}, T_n = I, W_1, W_2, \dots, W_n | K_2) \quad (12)$$

再使用 Viterbi 算法计算 $D_n^1(I)$:

(1) 初始化

$$D_1^1(I) = P(b_1(W_1), I) \text{ I } F \text{ I } F \text{ N} \quad (13a)$$

$$U_1^1(I) = 0, \text{ 其中 } 0 \text{ 表示空列} \quad (13b)$$

(2) 递归过程

$$D_{n+1}^1(J) = \max_{I \text{ I } 8} \{D_n^1(I) a_{i_1, j}^2 \# b_j(W_{n+1}), 1 \leq n < N-1, J \text{ I } 8\} \quad (14a)$$

$$U_{n+1}^1(J) = \arg \max_{I \text{ I } 8} \{D_n^1(I) a_{i_1, j}^2 \# b_j(W_{n+1}), 1 \leq n < N-1, J \text{ I } 8\} \quad (14b)$$

(3) 终止过程

$$P^* = \max_{I \text{ I } 8} \{D_N^1(I)\} \quad (15a)$$

$$T_N^* = \arg \max_{I \text{ I } 8} \{D_N^1(I)\} \quad (15b)$$

(4) 路径 (状态列) 回朔

$$T_n^* = U_{n+1}^1(T_{n+1}^*), n = N-1, N-2, \dots, 1 \quad (16)$$

3.2.2 基于标准 2 的解决 在该标准下求解问题 2, 需要定义一个关于状态列的后验概率变量:

$$C_n^2(I) = P(T_n = I | O, K_2) \quad (17a)$$

显然, $C_n^2(I)$ 可以改写成:

$$C_n^2(I) = \frac{P(O, T_n = I | K_2)}{P(O | K_2)} = \frac{A_n^2(I) B_n^2(I)}{\sum_{I \text{ I } 8} A_n^2(I) B_n^2(I)} \quad (17b)$$

于是, 利用 $C_n^2(I)$ 得到的每个最优状态列为:

$$T_n^* = \arg \max_{I \text{ I } 8} \{C_n^2(I)\}, 1 \leq n < N \quad (18)$$

3.2.3 基于标准 3 的解决 此时求解问题 2, 需要定义一个关于单个状态的后验概率变量:

$$G_{m, n}^2(k) = P(q_{m, n} = k | O, \#_2) \quad (19a)$$

显然, $G_{m, n}^2(k)$ 可以改写成:

$$G_{m, n}^2(k) = \sum_{I \text{ I } 8, I(m) = k} P(T_n = I | O, K_2) = \sum_{I \text{ I } 8, I(m) = k} C_n^2(I) \quad (19b)$$

其中, $I(m)$ 表示列状态 I 的第 m 个分量, 即如果 $I = (i_1, i_2, \dots, i_m, \dots, i_M)$, 则 $I(m) = i_m$.

于是, 利用 $G_{m, n}^2(k)$ 得到的单个最优状态为:

$$q_{m, n}^* = \arg \max_{I \text{ I } 8, k \text{ I } 8} G_{m, n}^2(k), 1 \leq m < M, 1 \leq n < N \quad (20)$$

3.3 问题 3 的解决

定义 $N_n^2(I, J)$ 为在给定观察矩阵 O 和模型 K_2 的条件下, 第 n 列状态为 I 且第 n+1 列状态为 J 的概率, 即,

$$N_n^2(I, J) = P(T_n = I, T_{n+1} = J | O, K_2), I, J \text{ I } 8 \quad (21a)$$

利用前向变量 $A_n^2(I)$ 和后向变量 $B_n^2(I)$, 不难把 $N_n^2(I, J)$

改写成下面的形式:

$$N_n^2(I, J) = \frac{A_n^2(I) a_{i,j}^2 b_j (W_{n+1}) B_{n+1}^2(J)}{\sum_{i=1}^M \sum_{j=1}^N A_n^2(I) a_{i,j}^2 b_j (W_{n+1}) B_{n+1}^2(J)} \quad (21b)$$

利用 $G_{m,n}^2(k)$, $C_n^2(I)$ 和 $N_n^2(I, J)$ 可以得到一个对 2 维 2 阶隐马尔可夫模型 #2 进行参数重估的方法. 如果把重估后得到的模型用 #2 = (A^V, A^H, A₂, B, 0) 表示, 那么各部分的重估公式描述如下:

$$P_j = G_{1,1}^2(j) \quad (22a)$$

$$a_{i,j}^V = \frac{\sum_{m=1}^{M-1} \sum_{l(1)=1}^{l(1)=j} \sum_{l(2)=1}^{l(2)=j} C_l^2(I)}{\sum_{m=1}^{M-1} G_{m,1}^2(i)} \quad (22b)$$

$$a_{i,j}^H = \frac{\sum_{n=1}^{N-1} \sum_{l(1)=1}^{l(1)=j} \sum_{l(2)=1}^{l(2)=j} N_n^2(I, J)}{\sum_{n=1}^{N-1} \sum_{l(1)=1}^{l(1)=i} \sum_{l(2)=1}^{l(2)=j} C_n^2(I)} \quad (22c)$$

$$a_{i,j,k}^2 = \frac{\sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \sum_{l(m)=i}^{l(m)=j} \sum_{l(m+1)=k}^{l(m+1)=k} N_n^2(I, J)}{\sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \sum_{l(m)=i}^{l(m)=j} \sum_{l(m+1)=k}^{l(m+1)=k} C_n^2(I)} \quad (22d)$$

$$b_k(r) = \frac{\sum_{m=1}^M \sum_{n=1}^N G_{m,n}^2(k) D(O_{m,n}, W_r)}{\sum_{m=1}^M \sum_{n=1}^N G_{m,n}^2(k)} \quad (22e)$$

其中 $1 \leq i, j, k \leq L$, $1 \leq r \leq R$, $D(O_{m,n}, W_r) = \begin{cases} 1, & O_{m,n} = W_r, \\ 0, & O_{m,n} \neq W_r \end{cases}$, $I(m)$ 和 $J(m)$ 的涵义见 3.1.2.3.

上述重估公式的收敛性证明可以借助 EM 算法的基本思想^[17], 参照 1 维隐马尔可夫模型的参数估计公式推导过程^[5]来完成, 其核心在于证明和满足如下关系 (见附录 1):

$$P(O | \#_2) \geq P(O | \#_1) \quad (23)$$

因此, 如果不断使用 #2 代替 #1, 并重复上述估计的计算过程, 最终得到的模型将使从模型 #2 产生观察矩阵 O 的概率 P(O | #2) 达到一个局部极大值.

前面对如何解决 2 维 2 阶隐马尔可夫模型的三个基本问题进行了详细的描述. 事实上, 几乎只需将上述公式中与 2 阶 O 有关的 2 替换成与 3 阶 O 有关的 3, 就可以得到 2 维 3 阶隐马尔可夫模型的求解方法, 关键的不同之处在于重估公式是:

$$a_{i,j,k,l}^3 = \frac{\sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \sum_{l(m)=i}^{l(m)=j} \sum_{l(m+1)=k}^{l(m+1)=k} \sum_{l(m+2)=l}^{l(m+2)=l} N_n^3(I, J)}{\sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \sum_{l(m)=i}^{l(m)=j} \sum_{l(m+1)=k}^{l(m+1)=k} \sum_{l(m+2)=l}^{l(m+2)=l} N_n^3(I, J)} \quad (24)$$

4 仿真举例

本节将构造一个 2 维 2 阶隐马尔可夫模型的计算实例对第 3 部分中几个算法的实现和运行作进一步说明. 取 L = 3, R = 4, #2 = (A^V, A^H, A₂, B, 0), 其中 P_j, a_{i,j}^V, a_{i,j}^H, a_{i,j,k}, b_k(r) 的

具体数值见表 1.

令观察矩阵 O = {O_{m,n} = w_l(m, n) | I ∈ I_{5,6}}, 那么用式(8)计算可得 P(O | #2) = 21922385 × 10⁻¹⁶.

基于式(9), (10) 和(11) 得到的最优状态矩阵都是:

$$Q = \{q_{m,n} = 1 | (m, n) \in I_{5,6}\}$$

此外, 如果以模型 #2 的数据作为初始概率, 用(22a)~(22e) 训练模型, 那么最后 P(O | #2) = 110, 其中 P_j, a_{i,j}^V, a_{i,j}^H, a_{i,j,k}, b_k(r) 的计算结果 (保留两位小数) 见表 2.

表 1 初始模型参数

#2 的参数	j= 1	j= 2	j= 3
P _j	0.3	0.4	0.3
a _{1,j} ^V , a _{1,j} ^H	0.6, 0.6	0.2, 0.2	0.2, 0.2
a _{2,j} ^V , a _{2,j} ^H	0.2, 0.2	0.2, 0.2	0.6, 0.6
a _{3,j} ^V , a _{3,j} ^H	0.2, 0.2	0.6, 0.6	0.2, 0.2
a _{1,j,1} , a _{1,j,2} , a _{1,j,3}	0.5, 0.3, 0.2	0.3, 0.2, 0.5	0.2, 0.5, 0.3
a _{2,j,1} , a _{2,j,2} , a _{2,j,3}	0.3, 0.2, 0.5	0.2, 0.5, 0.3	0.5, 0.3, 0.2
a _{3,j,1} , a _{3,j,2} , a _{3,j,3}	0.2, 0.5, 0.3	0.5, 0.3, 0.2	0.3, 0.2, 0.5
b _j (1)	0.4	0.3	0.2
b _j (2)	0.3	0.2	0.1
b _j (3)	0.2	0.1	0.4
b _j (4)	0.1	0.4	0.3

表 2 重估的新模型参数

#2 的参数	j= 1	j= 2	j= 3
P _j	0.59	0.25	0.15
a _{1,j} ^V , a _{1,j} ^H	0.76, 0.77	0.14, 0.13	0.10, 0.10
a _{2,j} ^V , a _{2,j} ^H	0.36, 0.36	0.20, 0.20	0.44, 0.44
a _{3,j} ^V , a _{3,j} ^H	0.32, 0.33	0.54, 0.54	0.13, 0.13
a _{1,j,1} , a _{1,j,2} , a _{1,j,3}	0.63, 0.25, 0.12	0.44, 0.21, 0.35	0.29, 0.50, 0.21
a _{2,j,1} , a _{2,j,2} , a _{2,j,3}	0.44, 0.21, 0.35	0.28, 0.51, 0.21	0.61, 0.27, 0.12
a _{3,j,1} , a _{3,j,2} , a _{3,j,3}	0.29, 0.50, 0.21	0.61, 0.26, 0.12	0.43, 0.21, 0.36
b _j (1)	1.00	1.00	1.00
b _j (2)	0.00	0.00	0.00
b _j (3)	0.00	0.00	0.00
b _j (4)	0.00	0.00	0.00

5 总结

本文研究了 2 维 2 阶 (或 3 阶) 隐马尔可夫模型的三个基本问题, 即概率评估问题、最优状态问题和参数估计问题, 从理论上给出了解决这三个基本问题的新算法, 并通过计算机仿真对算法的实现和运行作了进一步的说明. 这些算法不仅更加简洁, 而且更加易于实现. 尽管从本质上说这些算法具有指数复杂性, 但是对行数或者列数较小的情况, 其计算复杂性与 1 维隐马尔可夫模型相当.

附录 1 不等式 P(O | #2) ≥ P(O | #1) 和重估公式的推导

设 #c 和 # 是任意两个 2 维 2 阶隐马尔可夫模型的参数集, 首先构造下面的辅助函数:

$$5(\#c, \#) = \sum_Q P(O, Q | \#c) \log P(O, Q | \#)$$

其中 $P(O, Q | \#)$ 可以表达为:

$$P(O, Q | \#) = P_{q_{1,1}} b_{q_{1,1}}(o_{1,1}) \prod_{m=2}^M (a_{q_{m-1,1}, q_{m,1}}^V b_{q_{m,1}}(o_{m,1})) \\ \# \prod_{n=2}^N (a_{q_{1,n-1}, q_{1,n}}^H b_{q_{1,n}}(o_{1,n})) \\ \# \prod_{m=2}^M \prod_{n=2}^N (a_{q_{m-1,n}, q_{m,n}}^V b_{q_{m,n}}(o_{m,n}))$$

不难把 $5(\#c, \#)$ 改写为:

$$5(\#c, \#) = 5_0(\#c, 0) + \sum_{i=1}^L 5_{A_i^V}(\#c, A_i^V) + \sum_{j=1}^L 5_{A_j^H}(\#c, A_j^H) \\ + \sum_{i=1}^L \sum_{j=1}^L 5_{A_{i,j}}(\#c, A_{i,j}) + \sum_{k=1}^L 5_{B_k}(\#c, B_k)$$

其中:

$$0 = (P_1, P_2, \dots, P_L), A_i^V = (a_{i,1}^V, a_{i,2}^V, \dots, a_{i,L}^V), \\ A_j^H = (a_{j,1}^H, a_{j,2}^H, \dots, a_{j,L}^H), A_{i,j} = (a_{i,j,1}, a_{i,j,2}, \dots, a_{i,j,L}), \\ B_k = (b_k(1), b_k(2), \dots, b_k(R)),$$

$$5_0(\#c, 0) = \sum_{k=1}^L P(q_{1,1} = k, O | \#c) \log P_k,$$

$$5_{A_i^V}(\#c, A_i^V) = \sum_{j=1}^L \sum_{m=2}^M P(q_{m-1,1} = i, q_{m,1} = j, O | \#c) \log a_{i,j}^V,$$

$$5_{A_j^H}(\#c, A_j^H) = \sum_{i=1}^L \sum_{n=2}^N P(q_{1,n-1} = j, q_{1,n} = i, O | \#c) \log a_{j,i}^H,$$

$$5_{A_{i,j}}(\#c, A_{i,j}) = \sum_{k=1}^L \sum_{m=2}^M \sum_{n=2}^N P(q_{m-1,n} = i, q_{m,n} = j, \\ q_{m,n} = k, O | \#c) \log a_{i,j,k},$$

$$5_{B_k}(\#c, B_k) = \sum_{r=1}^R \left[\sum_{m=1}^M \sum_{n=1}^N P(q_{m,n} = k, O | \#c) D(o_{m,n}, w_r) \right] \log b_k(r)$$

由此可见, $5(\#c, \#)$ 可以分解为若干独立项的求和, 并且

在每个独立项最大时 $5(\#c, \#)$ 达到最大。由于 $\sum_{k=1}^L P_k =$

$$1, \sum_{j=1}^L a_{i,j}^V = 1, \sum_{i=1}^L a_{j,i}^H = 1, \sum_{k=1}^L a_{i,j,k} = 1, \sum_{r=1}^R b_k(r) = 1, \text{ 所以每个}$$

独立项都具有形式: $z = \prod_{t=1}^T w_t \log y_t$, 其中 $\prod_{t=1}^T y_t = 1, y_t \in (0, 1]$, z 是 $\{y_t\}_{t=1}^T$ 的函数。使用拉格朗日乘法, 易知 z 达到全局最大值的条件是

$$y_t = \frac{w_t}{\sum_{t=1}^T w_t}, t = 1, 2, \dots, T$$

所以最大化后, $5(\#c, \#) \leq 5(\#c, \#c)$, 从而

$$\log P(O | \#) - \log P(O | \#c) = \sum_Q P(Q | O, \#c) [\log P(O | \#) - \log P(O | \#c)]$$

$$= \sum_Q P(Q | O, \#c) \left[\log \frac{P(O, Q | \#)}{P(Q | O, \#)} - \log \frac{P(O, Q | \#c)}{P(Q | O, \#c)} \right]$$

$$= \frac{1}{P(O | \#c)} [5(\#c, \#) - 5(\#c, \#c)] + \sum_Q P(Q | O, \#c)$$

$$\# \log \frac{P(Q | O, \#)}{P(Q | O, \#c)} \leq \sum_Q P(Q | O, \#c) \log \frac{P(Q | O, \#)}{P(Q | O, \#c)} \leq 0$$

因此, $P(O | \#) \leq P(O | \#c)$ 。在辅助函数 $5(\#c, \#)$ 中用 $\#_2$ 代替 $\#$, 用 $\#_1$ 代替 $\#c$, 可得 $P(O | \#_2) \leq P(O | \#_1)$ 。故可得

重估公式如下:

$$P_j = \frac{P(q_{1,1} = j, O | \#_2)}{\sum_{j=1}^L P(q_{1,1} = j, O | \#_2)} \quad (25a)$$

$$a_{i,j}^V = \frac{\sum_{m=1}^{M-1} P(q_{m,1} = i, q_{m+1,1} = j, O | \#_2)}{\sum_{j=1}^L \sum_{m=1}^{M-1} P(q_{m,1} = i, q_{m+1,1} = j, O | \#_2)} \quad (25b)$$

$$a_{i,j}^H = \frac{\sum_{n=1}^{N-1} P(q_{1,n} = i, q_{1,n+1} = j, O | \#_2)}{\sum_{j=1}^L \sum_{n=1}^{N-1} P(q_{1,n} = i, q_{1,n+1} = j, O | \#_2)} \quad (25c)$$

$$a_{i,j,k}^2 = \frac{\sum_{m=1}^{M-1} \sum_{n=1}^{N-1} P(q_{m,n+1} = i, q_{m+1,n} = j, q_{m+1,n+1} = k, O | \#_2)}{\sum_{k=1}^L \sum_{m=1}^{M-1} \sum_{n=1}^{N-1} P(q_{m,n+1} = i, q_{m+1,n} = j, q_{m+1,n+1} = k, O | \#_2)} \quad (25d)$$

$$b_k(r) = \frac{\sum_{m=1}^M \sum_{n=1}^N P(q_{m,n} = k, O | \#_2) D(o_{m,n}, w_r)}{\sum_{m=1}^M \sum_{n=1}^N P(q_{m,n} = k, O | \#_2)} \quad (25e)$$

重估公式(25a~ e)经改写即得(22a~ e)。

$$b_k(r) = \frac{\sum_{m=1}^M \sum_{n=1}^N P(q_{m,n} = k, O | \#_2) D(o_{m,n}, w_r)}{\sum_{m=1}^M \sum_{n=1}^N P(q_{m,n} = k, O | \#_2)} \quad (25e)$$

参考文献:

[1] Baum L E, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains[J]. Ann. Math. Stat, 1966, 37(7): 1554-1563.

[2] Baum L E, Eagon J A. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology[J]. Bull. Amer. Math. Stat, 1967, 37(2): 360- 363.

[3] Baum L E. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov process[J]. Inf. equalities, 1972, 3(1): 1- 8.

[4] Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE, 1989, 77(2): 257- 286.

[5] Rabiner L R, Juang B H. Fundamentals of Speech Recognition[M]. USA: Prentice-Hall, 1993. 321- 389.

[6] Chen M Y, Kundu A, Zhou J. Offline handwritten word recognition using a hidden Markov model type stochastic network[J]. IEEE Transactions on Pattern Recognition and Machine Intelligence, 1994, 16(5): 481- 496.

[7] Feng B, Ding X Q. Offline handwritten Chinese character recognition with hidden Markov models[A]. 5th International Conference on Signal Processing Proceedings[C]. China: WCC2000, 2000. 3. 1542-1545.

[8] Yoshua Bengio, Vincent Philippe Lauzon, R Jean Ducharme. Experiments on the application of ICHMMs to model financial returns series [J]. IEEE Transactions on Neural Networks, 2001, 12(1): 113- 123.

[9] Florent Perronnin, Jean-Luc Dugelay, Ken Rose. Iterative decoding of two-dimensional hidden Markov models[A]. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Hong

- Kong, 2003. 3. 329- 332.
- [10] Agazzi O E, Kuo S S. Hidden Markov model based optical character recognition in the presence of deterministic transformations[J]. Pattern Recognition, 1993, 26(12): 1813- 1826.
- [11] Kuo S S, Agazzi O E. Keyword spotting in poorly printed documents using pseudo 2D hidden Markov models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(8): 842- 848.
- [12] Devijver P A. Segmentation of binary images using third order Markov mesh image models[A]. Proc. 8th Int. Conf. Pattern Recogn[C]. Paris, France, Oct. 1986. 259- 261.
- [13] Jeng F C, Woods J W. On the relationship of the Markov mesh to the NSHP Markov chain[J]. Pattern Recognition Lett, 1987, 5(4): 273- 279.
- [14] Devijver P A. Hidden Markov mesh random field models in image analysis[A]. K V Mardia, G K Kanji, eds. Advances in Applied Statistics (Statistics and Images 1)[M]. Carfax, Abingdon, 1993. 187- 227.
- [15] Park H S, Lee S W. A truly 2D hidden Markov model for offline handwritten character recognition[J]. Pattern Recognition, 1998, 31(2): 1894- 1864.
- [16] Li J, Najmi A, Gray R M. Image classification by a two-dimensional hidden Markov model[J]. IEEE Transactions on Signal Processing, 2000, 48(2): 517- 533.
- [17] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm[J]. J R Stat Soc, 1977, 39(1): 1- 21.

作者简介:



李玉 男, 1968 年 10 月生于湖南邵东, 1999 年中国科学院半导体研究所获工学博士学位, 2001 年北京邮电大学博士后出站, 现为北京工业大学计算机学院副教授, 主要研究方向包括自然语言处理, 模式识别和生物信息学.