第四章 隐马尔可夫模型与贝叶斯网络

苏智勇
可视计算研究组
南京理工大学
suzhiyong@njust.edu.cn
https://zhiyongsu.github.io

主要内容

4.1 贝叶斯网络的基本概念

4.2 隐马尔可夫模型（HMM）

4.3 朴素贝叶斯分类器

4.1 贝叶斯网络的基本概念

- **随机变量**
- **随机变量的条件独立性**
  - 两个随机变量x和y独立，当且仅当 $p(x,y)=p(x)p(y)$
  - 随机变量间的**条件独立性**：如果随机变量x，y，z满足 $p(x,y|z)=p(x|z)p(y|z)$，则称x和y关于z条件独立，

$$p(x|y,z)=\frac{p(x,y,z)}{p(y,z)}=\frac{p(x,y|z)p(z)}{p(y|z)p(z)}=p(x|z)$$

4.1 贝叶斯网络的基本概念

- **随机过程**
  - **分布**（联合分布）是对一个随机变量（随机向量）的刻画，而**过程**是对一族随机变量的刻画！
  - **定义**：随机过程是定义在$\Omega \times T$上的二元函数$X(\omega, t)$。对于固定的时间 t，$X(\omega, t)$为**随机变量**，简记为$X(t)$或X_t；对于固定的ω，$X(\omega, t)$为时间 t 的一般函数，称为**样本函数**或样本轨道，简记为$x(t)$。
    - 随机变量和样本函数是两个具有不同定义域和值域的单值函数。
    - 随机过程既可看成是“所有随机变量的集合”，也可看成为“所有样本函数的集合”。

4.1 贝叶斯网络的基本概念

- **随机过程**
  - 随机过程示意图
    - 随机变量$X(t)$：对于固定时间t，蓝色虚线圈圈表示随机变量
    - 样本函数$X(\omega, t)$：对于固定的ω， $X(\omega, t)$为时间 t 的样本函数（3个）
  - 股票的涨跌过程
    - 今天红-明天红-后天绿
    - 今天绿-明天绿-后天绿
    - …

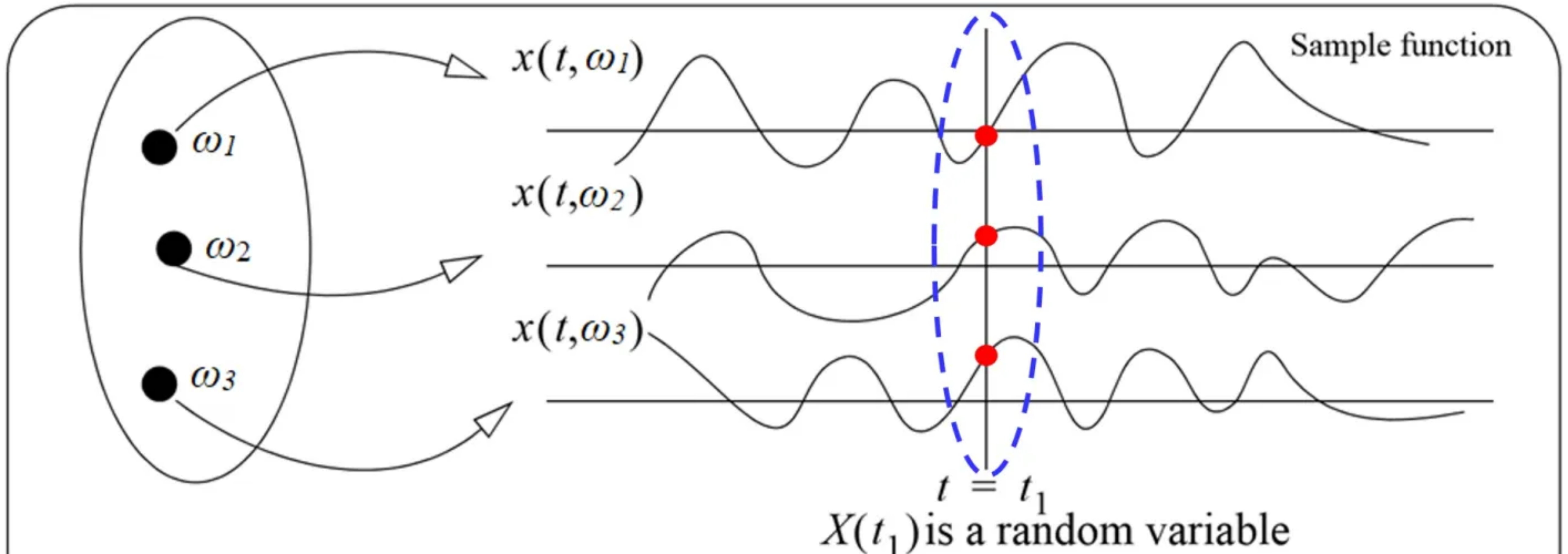


Figure 1:Representation of a stochastic process - relation to random variables

4.1 贝叶斯网络的基本概念

- **马尔科夫性质**
  - 也叫做无后效性、无记忆性，即过去只能影响现在，不能影响将来。
  - 如果$X(t)$，$t > 0$为一个随机过程，则马科尔夫性质可以符号化成如下形式：

$$Pr[X(t+h) = y \,|\, X(s) = x(s), s \leq t] = Pr[X(t+h) = y \,|\, X(t) = x(t)], \forall h > 0$$

    - 过去(s)并不影响将来$(t+h)$的状态，但当前状态蕴含着以往所有的状态信息。

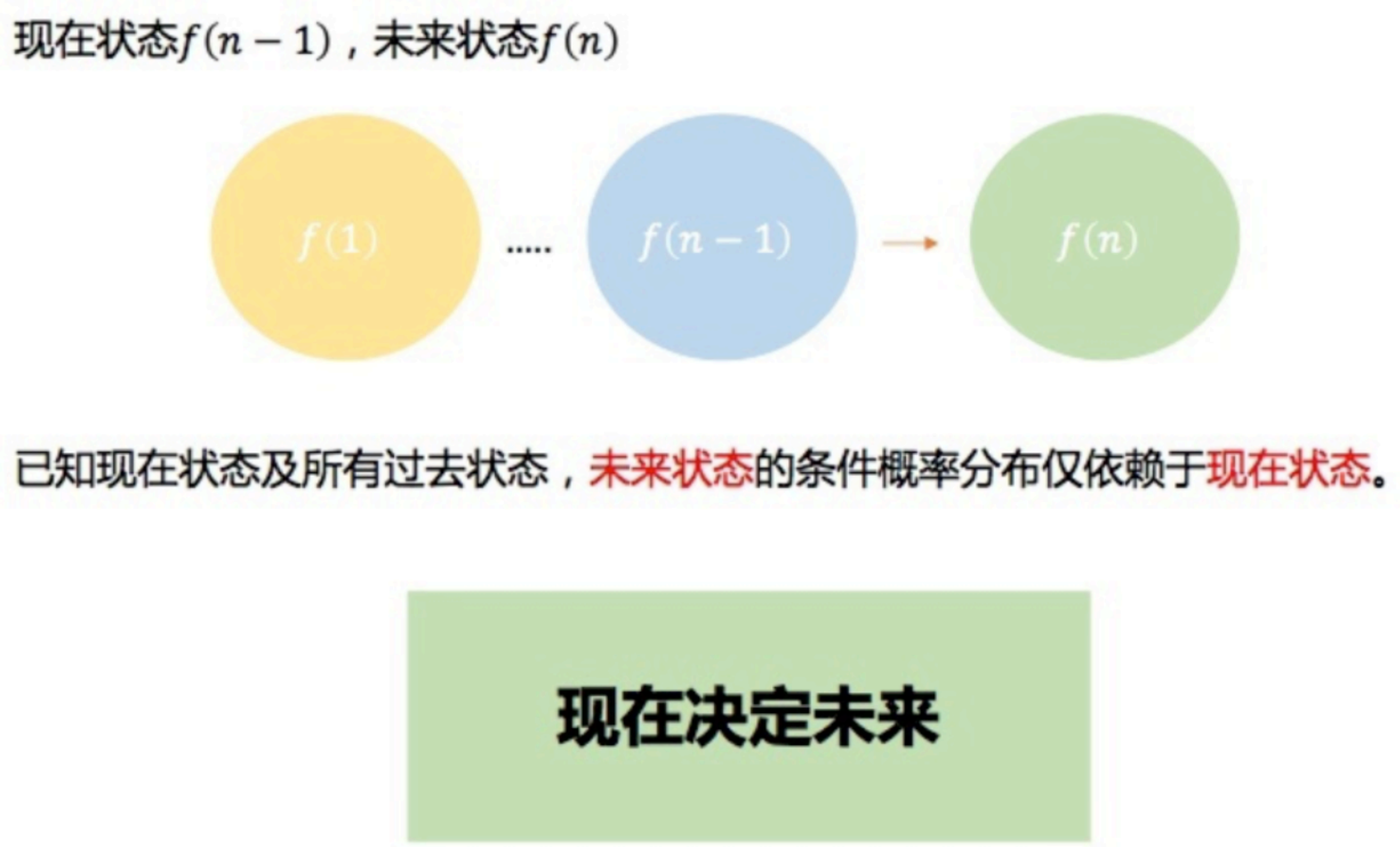
4.1 贝叶斯网络的基本概念

- **马尔科夫模型（链）**

  定义：具有**马尔可夫性质**、并以**随机过程**为基础模型的随机过程/随机模型被统称为马尔可夫模型（链）。

  – 任性的过程：它将来的状态分布只取决于现在，跟过去无关！ **（一阶）**

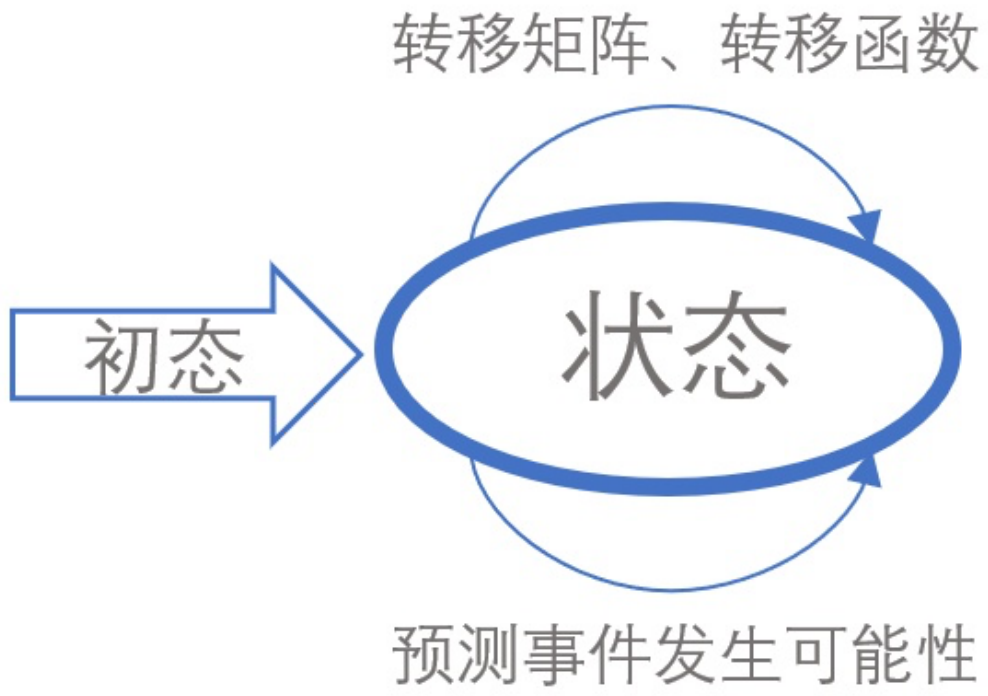
  – Life is like a Markov chain, your future only depends on what you are doing now, and independent of your past.



4.1 贝叶斯网络的基本概念

- **马尔科夫模型（链）基本要素**
  - **状态空间**：$X_n = i$表示随机过程在n时刻处在i状态，所有状态的取值构成的集合称为“状态空间”，以符号I表示
  - **转移概率**：把在当前时刻状态到下一时刻某状态的条件概率称作转移概率

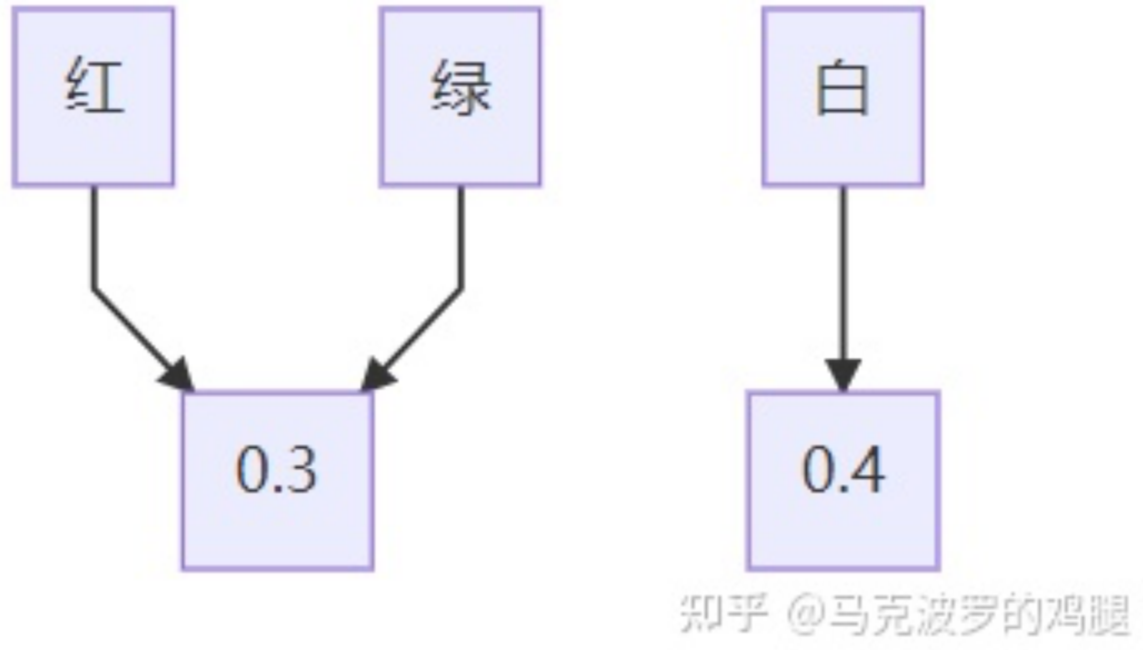
$$P_{ij} = P(x_n = j \mid X_{n-1} = i)$$



4.1 贝叶斯网络的基本概念

- 马尔科夫模型（链）基本要素
  - **转移概率矩阵**：所有状态之间的转移概率组成一个矩阵。**状态转移矩阵不随时间的变化。**

$$P=\begin{Bmatrix} p_{11} & p_{12} & \cdots & p_{1I} \\ p_{21} & p_{22} & \cdots & p_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ p_{I1} & p_{I2} & \cdots & p_{II} \end{Bmatrix}$$



  - **初始状态**：即初始分布
    - $p_0=(p_0(1),p_0(2),p_0(3))=(0.3,0.4,0.3)$
    - $p_n(j)=P\left(X_n=j\right),\ j=1,2,3$

4.1 贝叶斯网络的基本概念

- **马尔科夫模型（链）基本要素**
  - **转移方程**：对于m步转移，则有转移方程如下

$$p_{ij}^{(n+m)} = \sum_{k=1}^{I} p_{ik}^{(n)} p_{kj}^{(m)}$$

    - 今天：$p_0 = (0.3, 0.4, 0.3)$
    - 明天：$p_1 = p_0 \times P^1$
    - 后天：$p_2 = p_1 \times P = p_0 \times P^2$

4.1 贝叶斯网络的基本概念

- **链式法则**

$$\boxed{P(AB) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)}$$

  - 高维随机变量（随机向量）的联合分布：

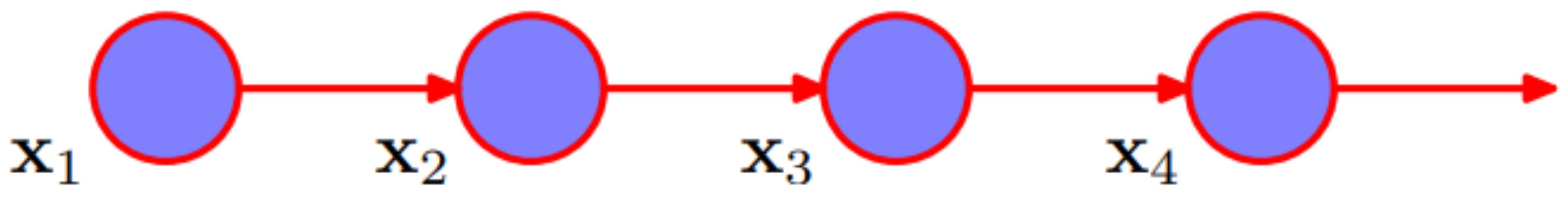
$$p(x_1, x_2, \ldots, x_d) = p(x_1, x_2, \ldots, x_{d-1}) p(x_d \mid x_1, x_2, \ldots, x_{d-1})$$

$$p(x_1, x_2, \ldots, x_d) = p(x_1) p(x_2 \mid x_1) p(x_3 \mid x_1, x_2) \ldots p(x_d \mid x_1, x_2, \ldots, x_{d-1})$$

$$p(x_1, \ldots, x_d) = p(x_1) \prod_{n=2}^{d} p(x_n \mid x_1, \ldots, x_{n-1})$$

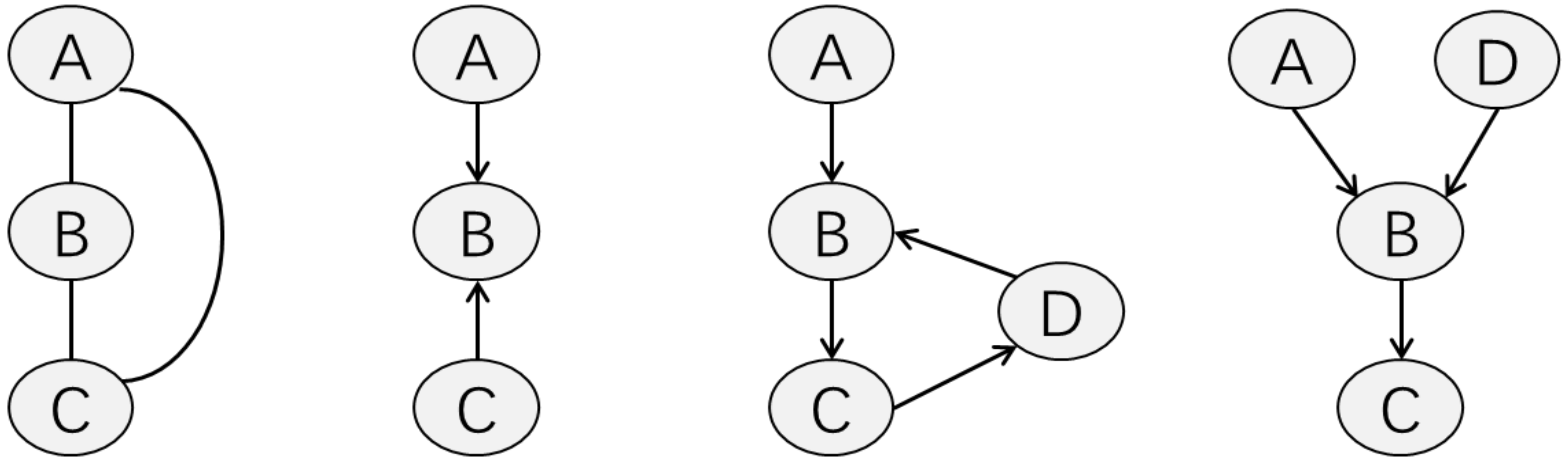
  - 如果x_{t+1}与$x_1, x_2, \cdots, x_{t-1}$关于$x_t$条件独立，则上式可以简化为

$$p(x_1, \ldots, x_d) = p(x_1) \prod_{n=2}^{d} p(x_n \mid x_{n-1})$$



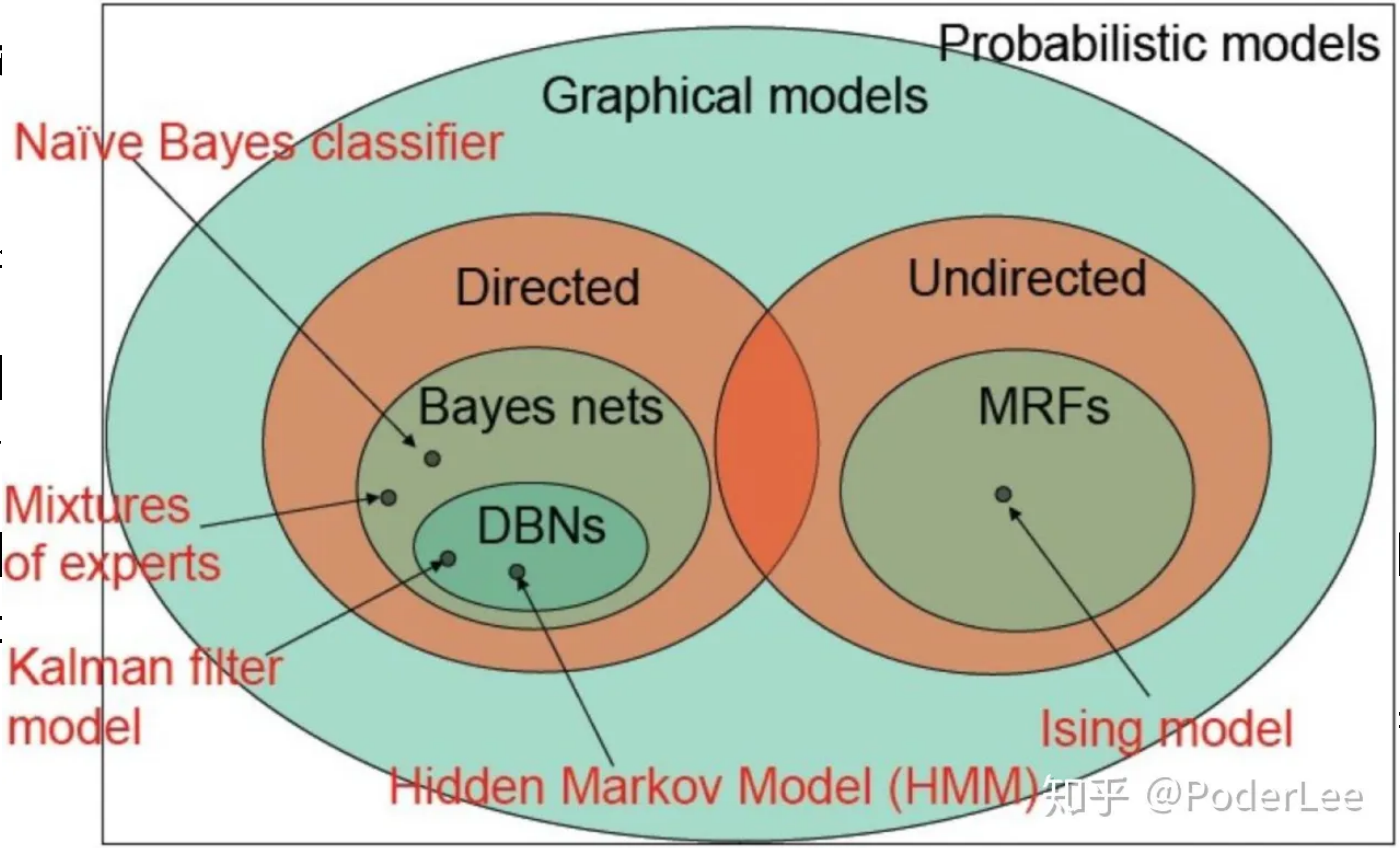
4.1 贝叶斯网络的基本概念

- **条件概率的图表示**
  - 图的基本概念
    - 节点、边；有向图、无向图；有环图、无环图
  - 条件概率$P(X=a\,|\,Y=b)$、联合概率$P(X=a,Y=b)$、边缘概率$P(X=a)$



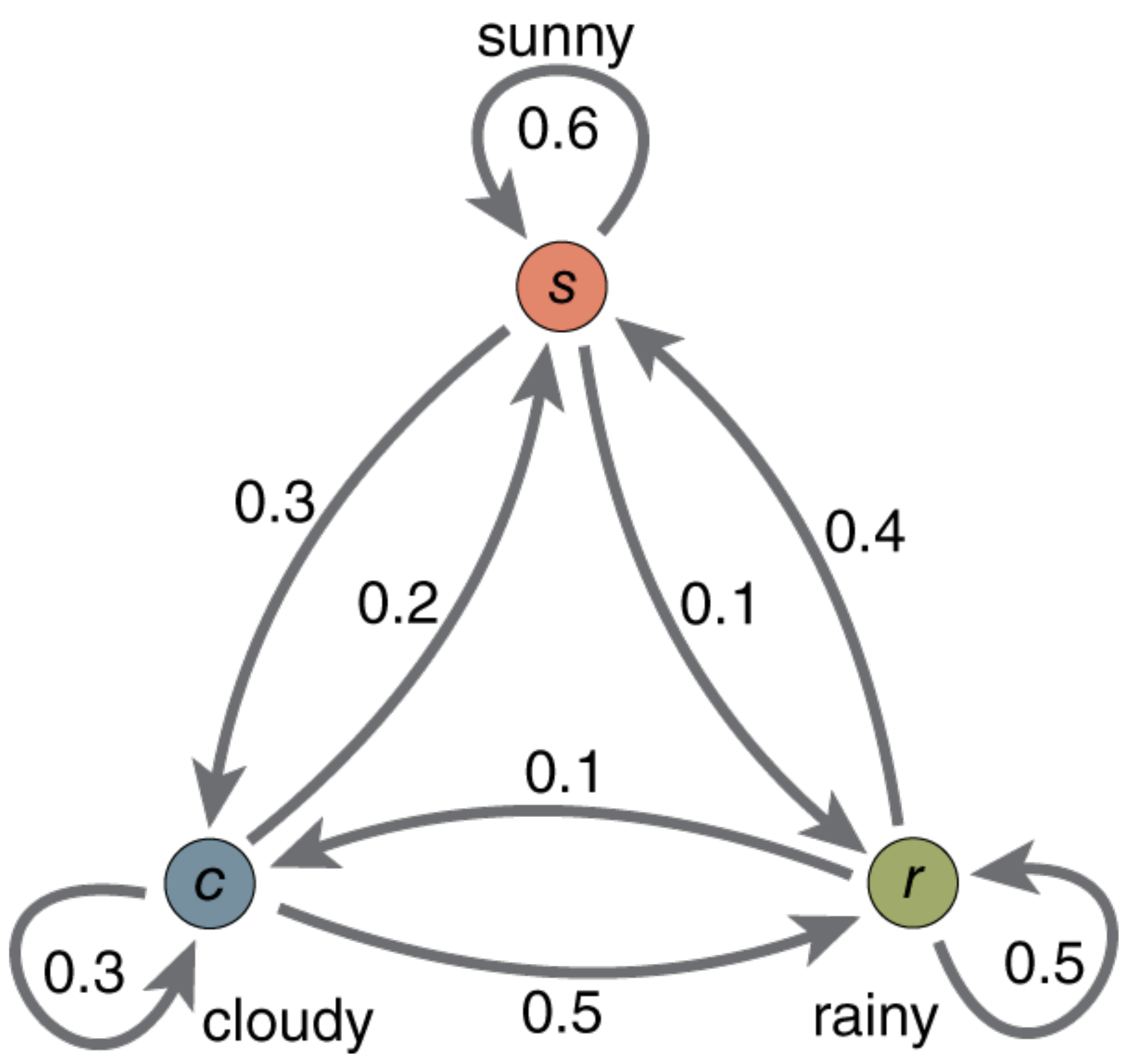
4.1 贝叶斯网络的基本概念

- **贝叶斯网**
  - 一种用
    - 概率
    - **有向**
      netw
    - **无向** 子图
      (fac
  - 通过图 的计算



4.2 隐马尔科夫模型

- 一种特殊的动态贝叶斯网络
- **HMM案例**
  - **已知：**
    - 天气：状态空间（Sunny，Rainy，Cloudy）、天气转移概率
    - 行为：散步、购物、内务
    - A、B跨国恋，A的对象B通过A的行为猜测可能的天气状态

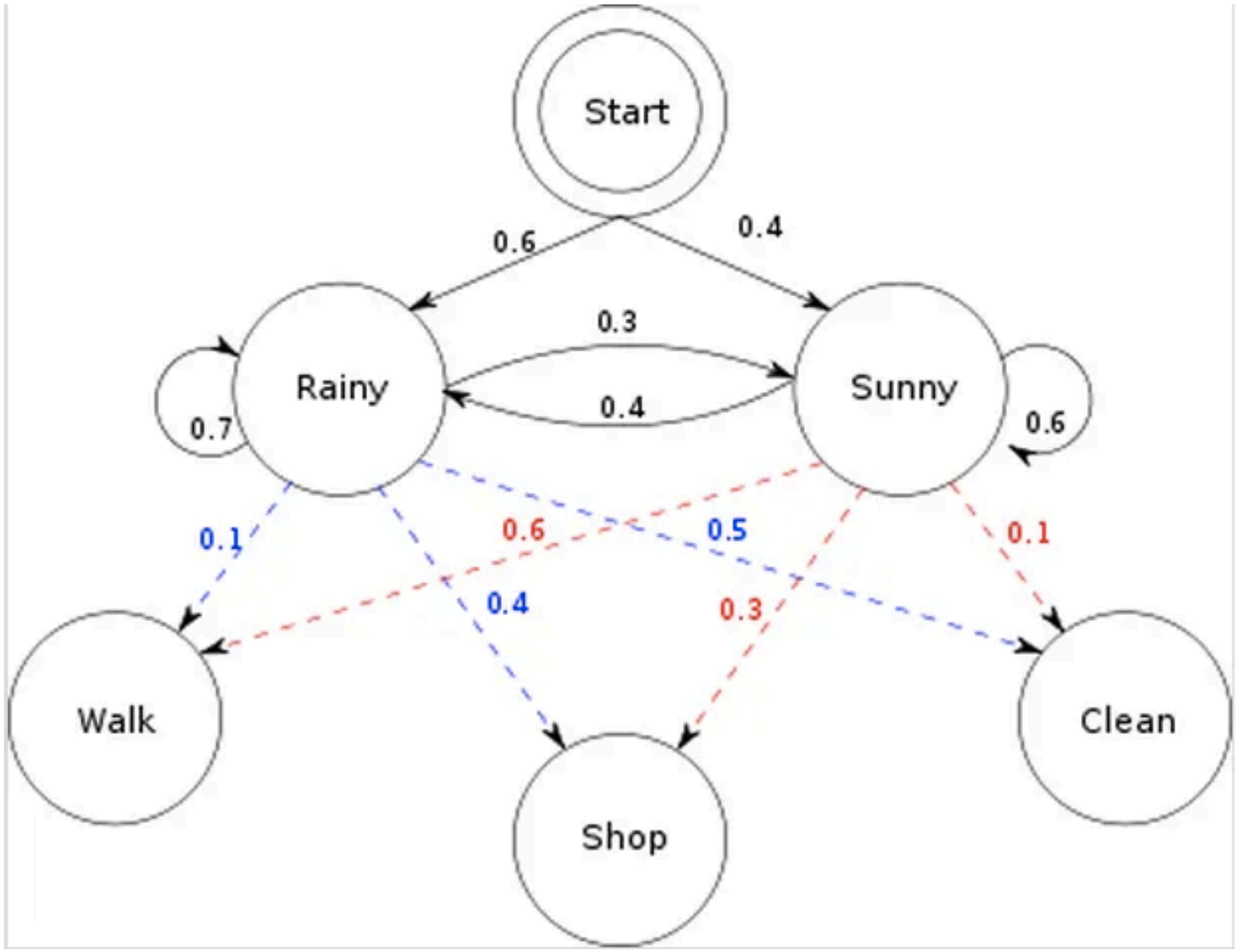


4.2 隐马尔科夫模型

- **HMM案例**
  - 条件：
    - **隐状态序列**：天气状况（B无法直接观测到）
    - **观测序列**：A的行为（A告知B）
    - **行为概率**：以晴天和雨天两种情况为例

✦晴天：散步，购物，内务的概率分别是0.6，0.3，0.1

✦雨天：散步，购物，内务的概率分别是0.1，0.4，0.5

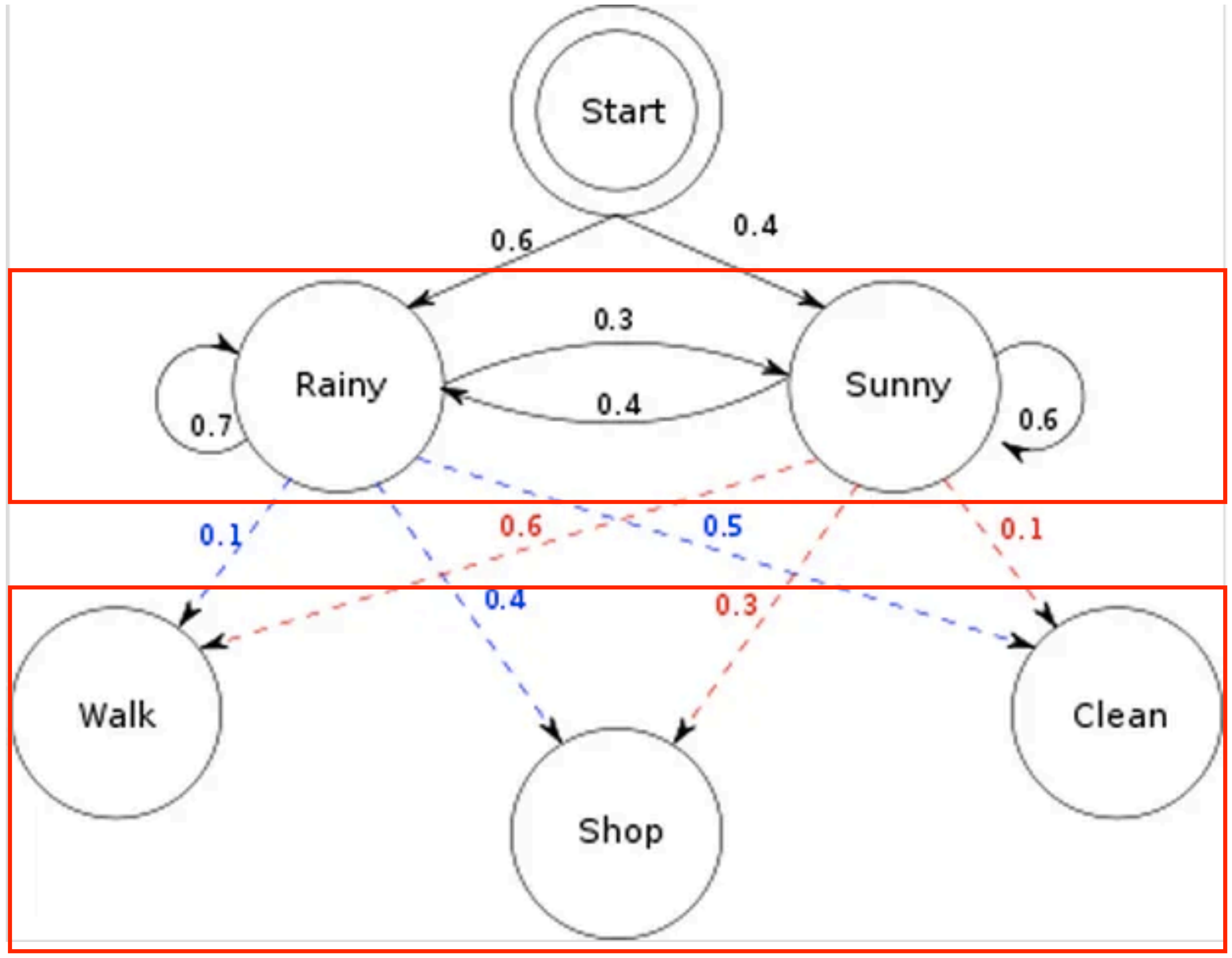


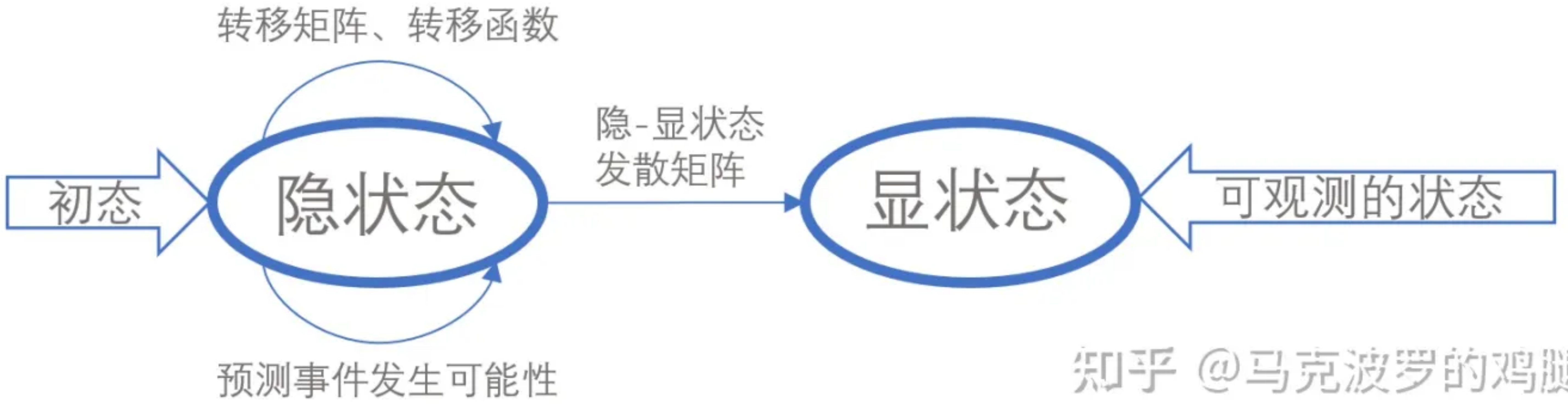
4.2 隐马尔科夫模型

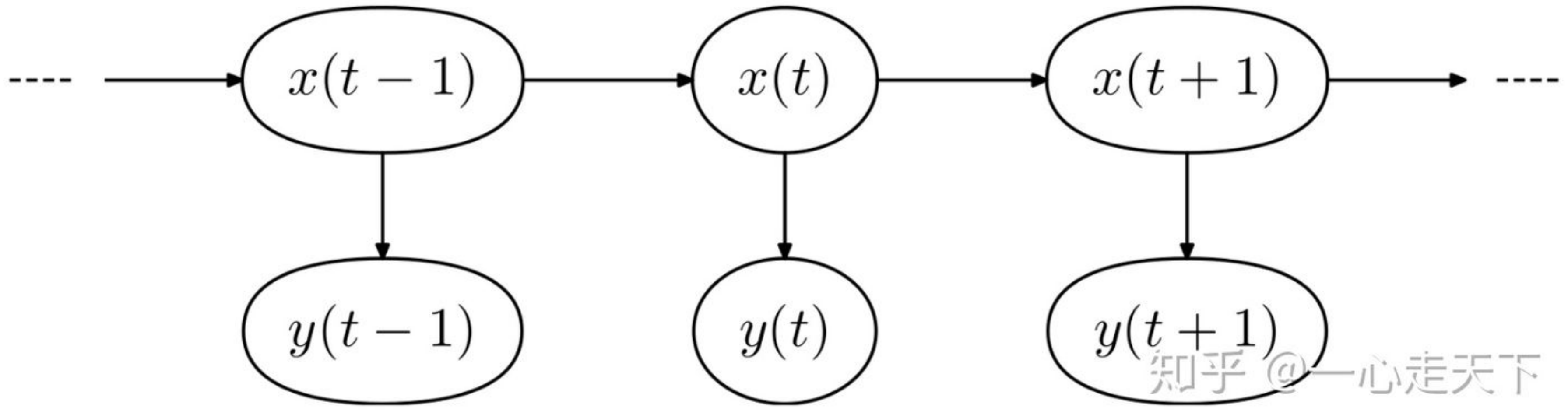
- **HMM案例**
  - **问题1**：已知整个模型，连续三天，A下班后做的事情分别是：散步，购物，内务。那么，B根据模型，计算产生这些行为的概率是多少。
  - **问题2**：同样知晓这个模型，同样是这三件事，A要B猜，这三天她下班后北京的天气是怎么样的。这三天怎么样的天气才最有可能让她做这样的事情。
  - **问题3**，最复杂的，A只告诉B这三天她分别做了这三件事，而其他什么信息我都没有。A要B建立一个模型，晴雨转换概率，第一天天气情况的概率分布，根据天气情况她选择做某事的概率分布。（惨绝人寰）

4.2 隐马尔科夫模型

- **基本概念**
  - 系统的状态取值服从马尔可夫链
  - 系统状态的变化无法直接观察到
  - 可能的状态和观测是贝叶斯网络的**隐节点**和**可观测节点**
  - **转移概率**为隐节点之间的边
  - **发射概率**为隐节点到可观测节点之间的边

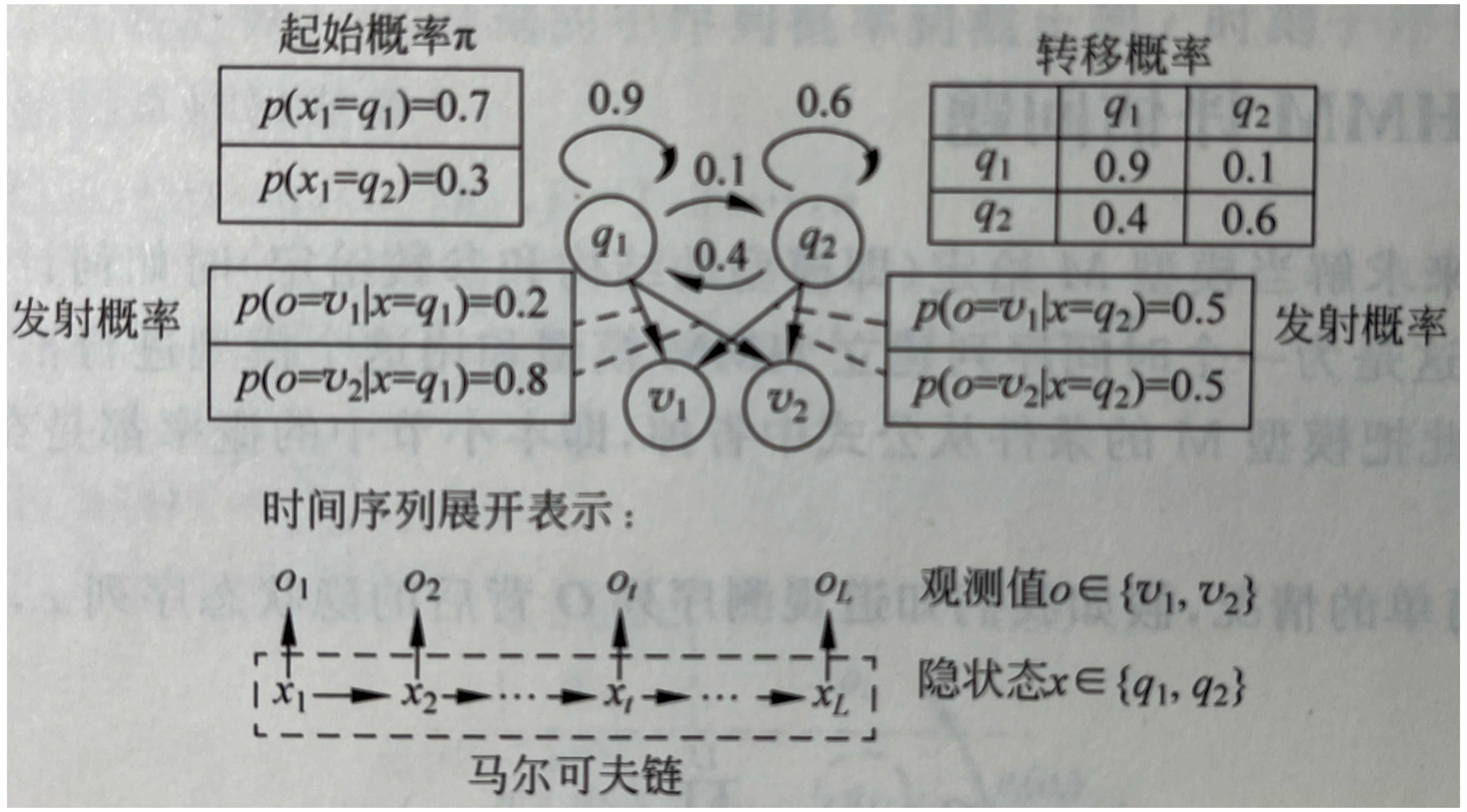






4.2 隐马尔科夫模型

- HMM三要素 $M = (A, E, \pi)$
  - **状态转移矩阵 A，发射概率矩阵E，初始概率分布 π**



变量表示	说明
$Q = \{q_1, q_2, \cdots, q_n\}$	模型含有n个隐状态
$V = \{v_1, v_2, \cdots, v_V\}$	观测值的取值范围
$\boldsymbol{A} = [a_{ij}]_{n \times n}$	状态转移概率矩阵，a_{ij}表示从状态i转到状态j的概率，满足$\sum_{j=1}^{n} a_{ij} = 1, \forall i$
$\boldsymbol{O} = o_1 o_2 \cdots o_L$	长度为L的观测序列，o_t的取值为V中某个值
$\boldsymbol{x} = x_1 x_2 \cdots x_L$	长度为L的隐状态序列，x_t的取值为Q中某个值
$\boldsymbol{E} = [e_{ij}]_{n \times V}$	发射概率矩阵，$e_{ij} = p(o = v_j \| x = q_i)$表示模型隐状态取值$q_i$时观测到$v_j$的概率，满足$\sum_{j=1}^{V} e_{ij} = 1, \forall i$
$\boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_n]$	初始概率分布，π_i表示马氏链从该状态起始的概率，满足$\sum_{i=1}^{n} \pi_i = 1$

4.2 隐马尔科夫模型

- 隐马尔科夫模型的**三种典型问题**：解决时间序列的决策问题
  - **模型评估问题**（Evaluation）

    概率计算问题。已知HMM模型M和观测序列O，求当前**观测序列**出现的概率$p(O\,|\,M)$（似然度）。
  - **隐状态推断问题**（Decoding）

    预测问题。已知HMM模型M和观测序列O，求产生该序列的最有可能的**隐状态序列**
    $\boldsymbol{x} = \arg\max_{x} p(\boldsymbol{x}, O\,|\,M)$。
  - **模型学习问题**（Learning）

    已知观测序列O和隐层状态序列$\boldsymbol{x}$，求HMM的**模型参数** $\arg\max_{M} p(\boldsymbol{x}, O\,|\,M)$或者

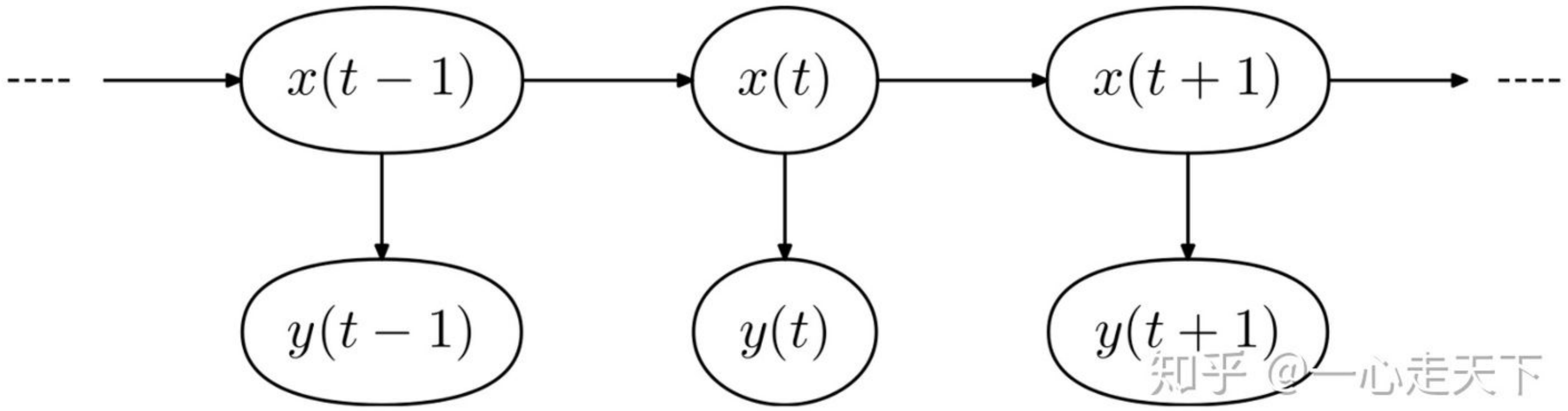
    $\arg\max_{M} p(O\,|\,M)$。

4.2.1 HMM模型评估问题

已知HMM模型M（模型结构和参数）和观测序列O，求当前**观测序列**出现的概率$p(O|M)$（似然度）。

- 已知隐状态序列$\boldsymbol{x}$，即$p(\boldsymbol{x})$已知
  - 似然度：$p(O|\boldsymbol{x}) = \prod_{i=1}^{L} p(o_i|x_i)$
  - 观测序列 O 和隐状态序列 $\boldsymbol{x}$ 的联合概率，即$p(O|M)$：

$$p(O, \mathbf{x}) = p(O \mid \boldsymbol{x})p(\boldsymbol{x}) = \prod_{i=1}^{L} p\left(o_i \mid x_i\right)\left(\pi_{x_1}\prod_{i=2}^{L} p\left(x_i \mid x_{i-1}\right)\right)$$



4.2.1 HMM模型评估问题

已知HMM模型M（模型结构和参数）和观测序列O，求当前**观测序列**出现的概率$p(O|M)$（似然度）。

- 未知隐状态序列$\boldsymbol{x}$，即$p(\boldsymbol{x})$未知
  - 概率分解：穷举，求解当前观测序列在所有可能隐状态序列下的加权概率和

$$p(O) = \sum_{\boldsymbol{x}} p(O, \boldsymbol{x}) = \sum_{\boldsymbol{x}} p(O \mid \boldsymbol{x}) p(\boldsymbol{x})$$

  - 计算量巨大：n^L 种隐状态序列取值组合

4.2.1 HMM模型评估问题

- **前向（向前）算法**：Forward Algorithm
  - 已知当前状态，要得到结果，往后能怎么“走”？（求果）
  - 从前往后，迭代求解：长度为t的观测子序列$o_1o_2\ldots o_t$在t时刻隐变量取值q_j的概率$\alpha_t(j)$
    - $\alpha_t(j) = p\left(o_1o_2\ldots o_t, x_t = q_j\right)$：$t$时刻，隐状态为$q_j$时，观测子序列(1-t)出现的概率
    - $\alpha_t(j) = \sum_{i=1}^{n} \alpha_{t-1}(i)a_{ij}e_j(o_t) = e_j(o_t)\sum_{i=1}^{n} \alpha_{t-1}(i)a_{ij}$，$n$为隐状态数
    - $\alpha_1(j)$，$\alpha_2(j)$，…，$\alpha_L(j)$
  - 终止结果
    - $p(O) = \sum_{i=1}^{n} \alpha_L(i)$：将最终结点的（不同走法）的概率求和，就是产生目前观测的可能情况

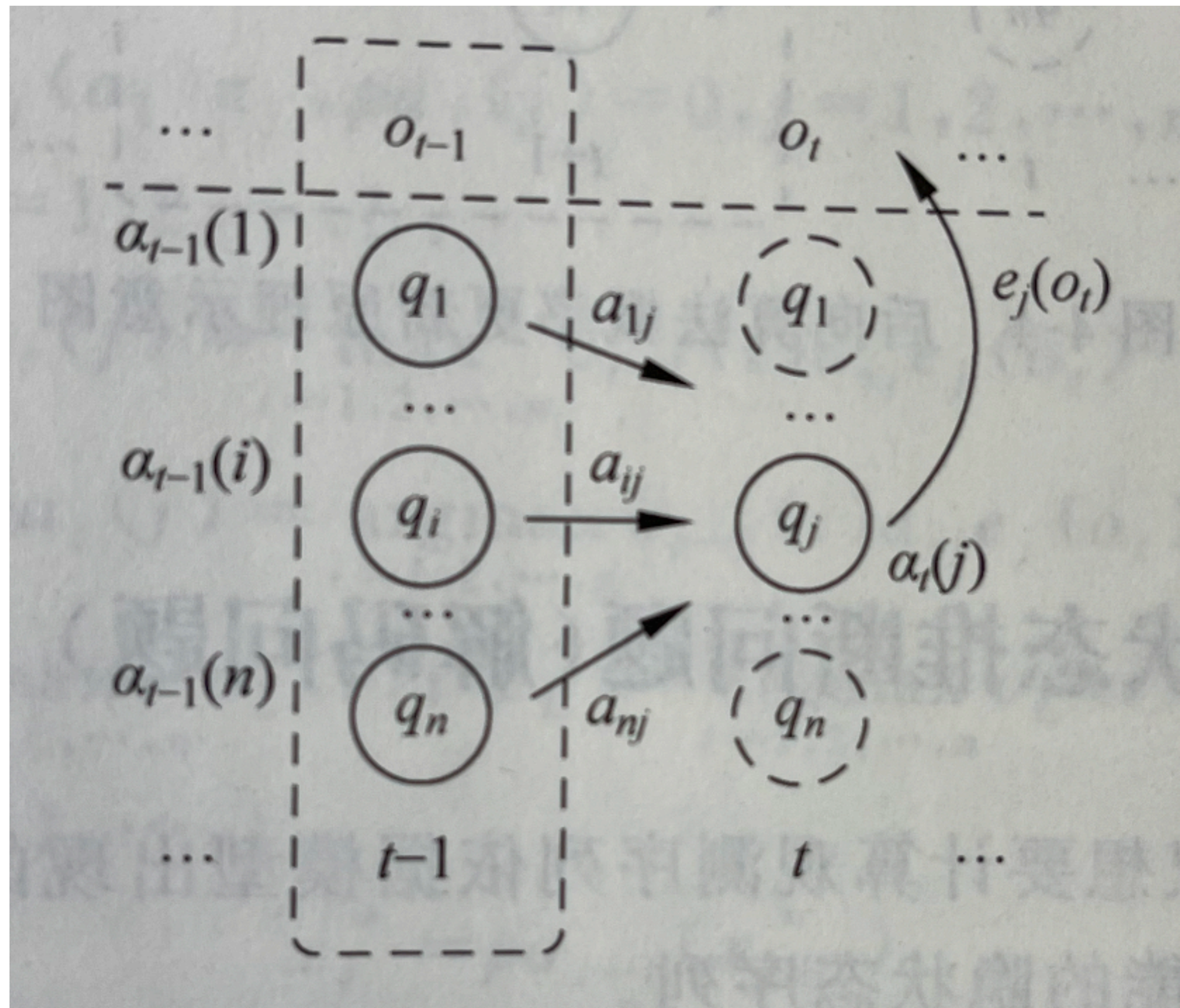


图 4-3　前向算法概率更新原理示意图

4.2.1 HMM模型评估问题

- **前向算法举例**

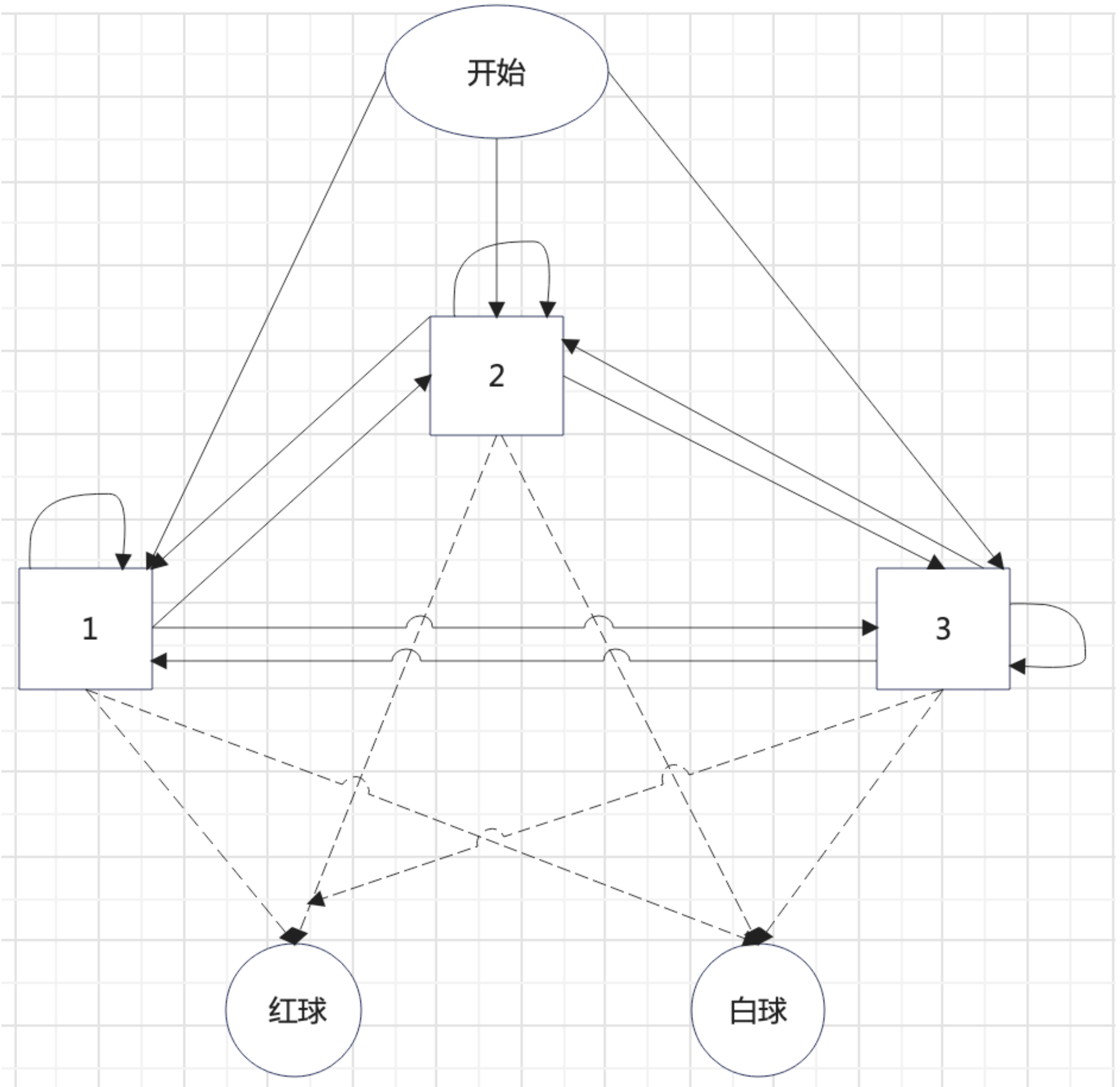
例如，李航《统计学习方法》例10.2的模型，观测序列为O= {红，白，红}，提供转移矩阵A，发射矩阵B，初始矩阵 π ,求最优的状态序列（也就是说这红，白，红最可能是通过怎样从1、2、3号盒中拿球得到的）

$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

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

$$\pi = (0.2, 0.4, 0.4)^T$$

求给定这个模型条件下，观测为O = (红，白，红) 的概率



4.2.1 HMM模型评估问题

$$\boxed{\alpha_t(j) = \sum_{i=1}^{n} \alpha_{t-1}(i) a_{ij} e_j(o_t) = e_j(o_t) \sum_{i=1}^{n} \alpha_{t-1}(i) a_{ij}}$$

- **前向算法举例**

计算初值，第一局取的是红球，可能从1，2，3号盒取：

- 红球是从1盒取的： $\alpha_{red}(1) = \pi_1 b_1(o_{red}) = 0.2 \times 0.5 = 0.1$
- 红球是从2盒取的： $\alpha_{red}(2) = \pi_2 b_2(o_{red}) = 0.4 \times 0.4 = 0.16$
- 红球是从3盒取的： $\alpha_{red}(3) = \pi_3 b_3(o_{red}) = 0.4 \times 0.7 = 0.28$

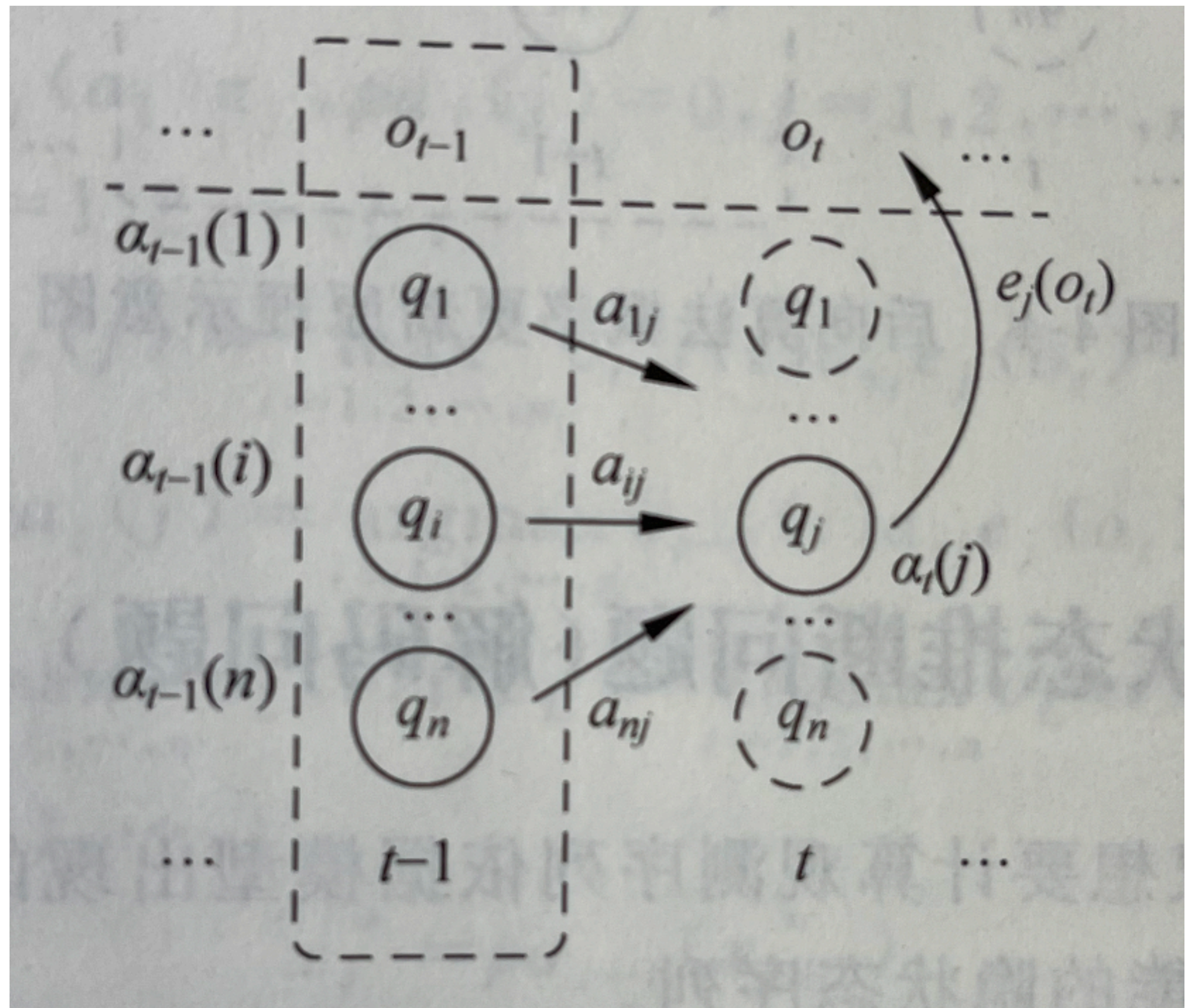


图 4-3 前向算法概率更新原理示意图

第二局取的是白球，可能从1，2，3号盒取

- 白球从1号盒取的概率：

$$\alpha_{white}(1) = (\alpha_{red}(1) \cdot a_{11} + \alpha_{red}(2) \cdot a_{21} + \alpha_{red}(3) \cdot a_{31}) \cdot b_1(O_{white})$$

$$= (0.1 \times 0.5 + 0.16 \times 0.3 + 0.28 \times 0.2) \times 0.5 = 0.154 * 0.5 = 0.077$$

- 白球从2号盒取的概率：

$$\alpha_{white}(2) = (\alpha_{red}(1) \cdot a_{12} + \alpha_{red}(2) \cdot a_{22} + \alpha_{red}(3) \cdot a_{32}) \cdot b_2(O_{white})$$

$$= (0.1 \times 0.2 + 0.16 \times 0.5 + 0.28 \times 0.3) \times 0.6 = 0.184 * 0.6 = 0.1104$$

- 白球从3号盒取的概率

$$\alpha_{white}(3) = (\alpha_{red}(1) \cdot a_{13} + \alpha_{red}(2) \cdot a_{23} + \alpha_{red}(3) \cdot a_{33}) \cdot b_3(O_{white})$$

$$= (0.1 \times 0.3 + 0.16 \times 0.2 + 0.28 \times 0.5) \times 0.7 = 0.202 * 0.3 = 0.0606$$

4.2.1 HMM模型评估问题

- **前向算法举例**

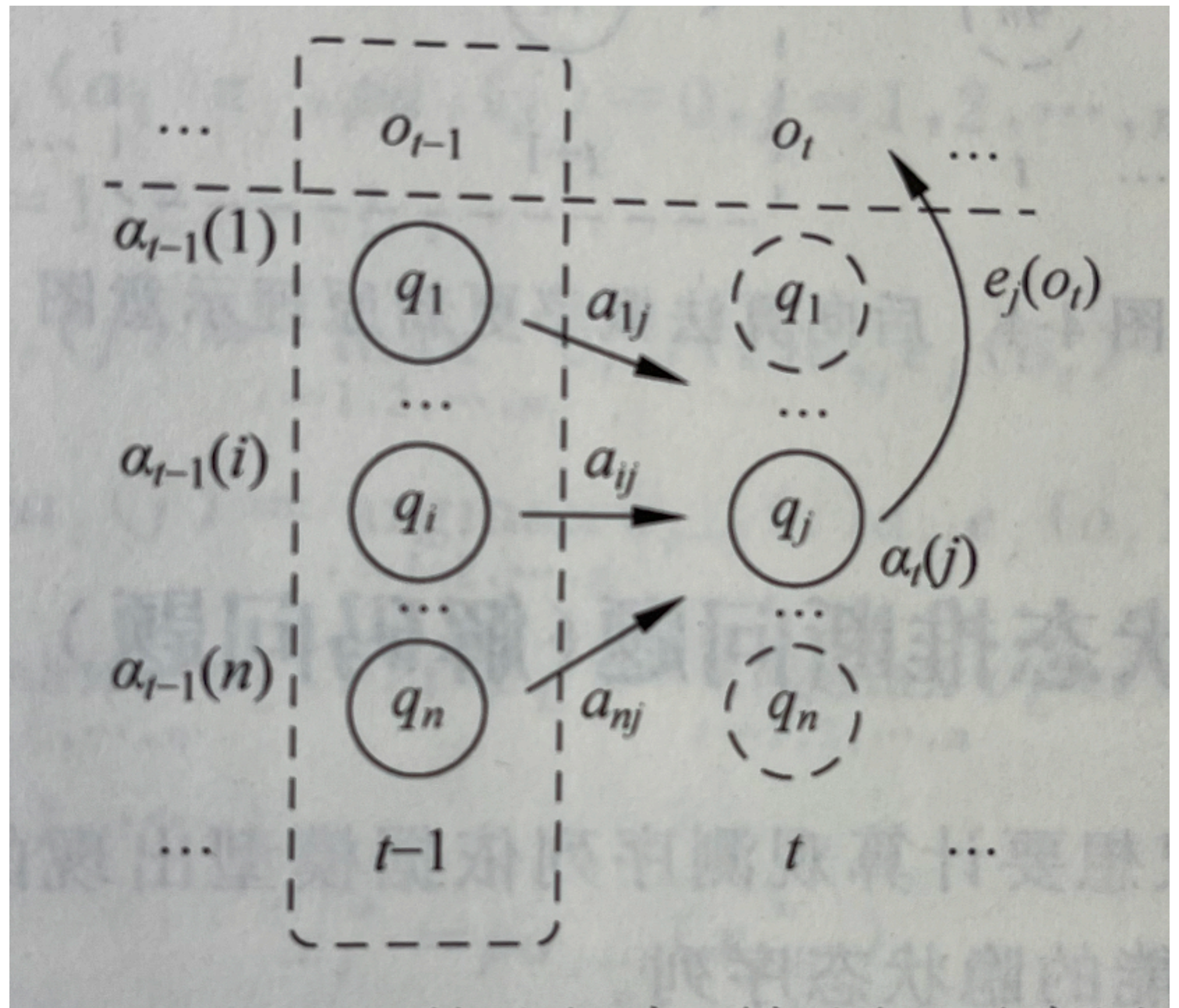


图 4-3　前向算法概率更新原理示意图

第三局取的是红球，可能从1，2，3号盒取

- 红球从1号盒取的概率：
  $\alpha_{red}(1) = (\alpha_{white}(1) \cdot a_{11} + \alpha_{white}(2) \cdot a_{21} + \alpha_{white}(3) \cdot a_{31}) \cdot b_1(O_{red})$

$= (0.077 \times 0.5 + 0.1104 \times 0.3 + 0.0606 \times 0.2) \times 0.5 = 0.08374 * 0.5 = 0.04187$

- 红球从2号盒取的概率：
  $\alpha_{red}(2) = (\alpha_{white}(1) \cdot a_{12} + \alpha_{white}(2) \cdot a_{22} + \alpha_{white}(3) \cdot a_{32}) \cdot b_2(O_{red})$

$= (0.077 \times 0.2 + 0.1104 \times 0.5 + 0.0606 \times 0.3) \times 0.4 = 0.08878 * 0.4 = 0.03551$

- 红球从3号盒取的概率

$\alpha_{red}(3) = (\alpha_{white}(1) \cdot a_{13} + \alpha_{white}(2) \cdot a_{23} + \alpha_{white}(3) \cdot a_{33}) \cdot b_3(O_{red})$

$= (0.077 \times 0.3 + 0.1104 \times 0.2 + 0.0606 \times 0.5) \times 0.7 = 0.08878 * 0.7 = 0.05284$

终止：

$P(O|\lambda) = \alpha_{red}(1) + \alpha_{red}(2) + \alpha_{red}(3) = 0.04187 + 0.03551 + 0.05284 = 0.13022$

以上过程其实是**前向算法**

4.2.1 HMM模型评估问题

- **后向（向后）算法**：Backward Algorithm
  - 已知结果，是当初怎么“走”造成的？（追因）
  - 从后往前求解：t时刻隐状态$x_t = q_j$时，观察到后续观测值$o_{t+1}o_{t+2}\ldots o_L$的概率$\beta_t(j)$
    - $\beta_t(j) = p(o_{t+1}o_{t+2}\cdots o_L \mid x_t = q_j)$：$t$时刻，隐状态为$q_j$时，观测子序列(t+1 - L)出现的概率
    - $\beta_t(j) = \sum_{i=1}^{n} \beta_{t+1}(i)e_i(o_{t+1})a_{ji}$
    - $\beta_L(j) = 1$，$\beta_{L-1}(j)$，…，$\beta_1(j)$
  - 终止结果
    - $p(O) = \sum_{i=1}^{n} \pi_i e_i(o_1)\beta_1(i)$

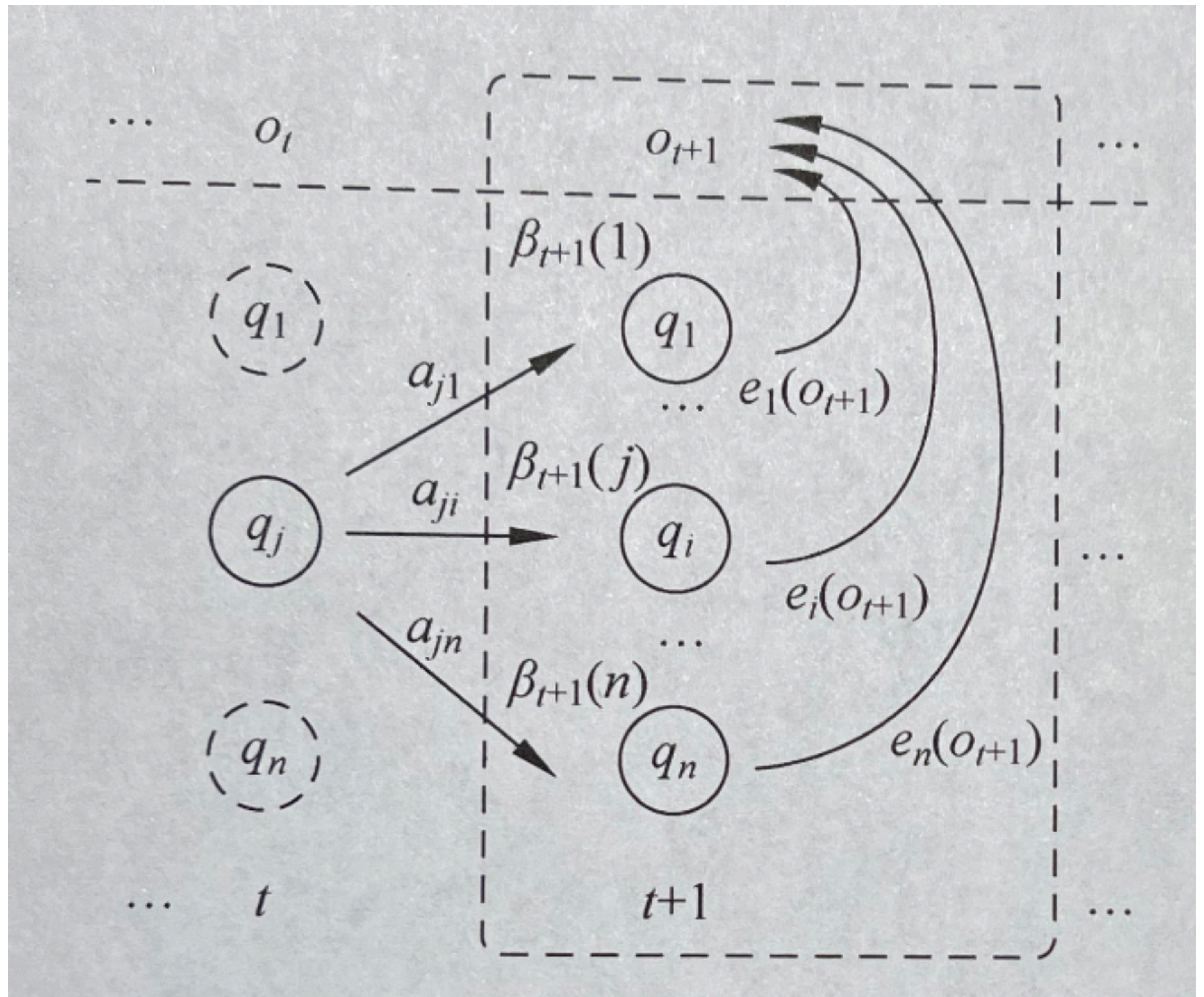
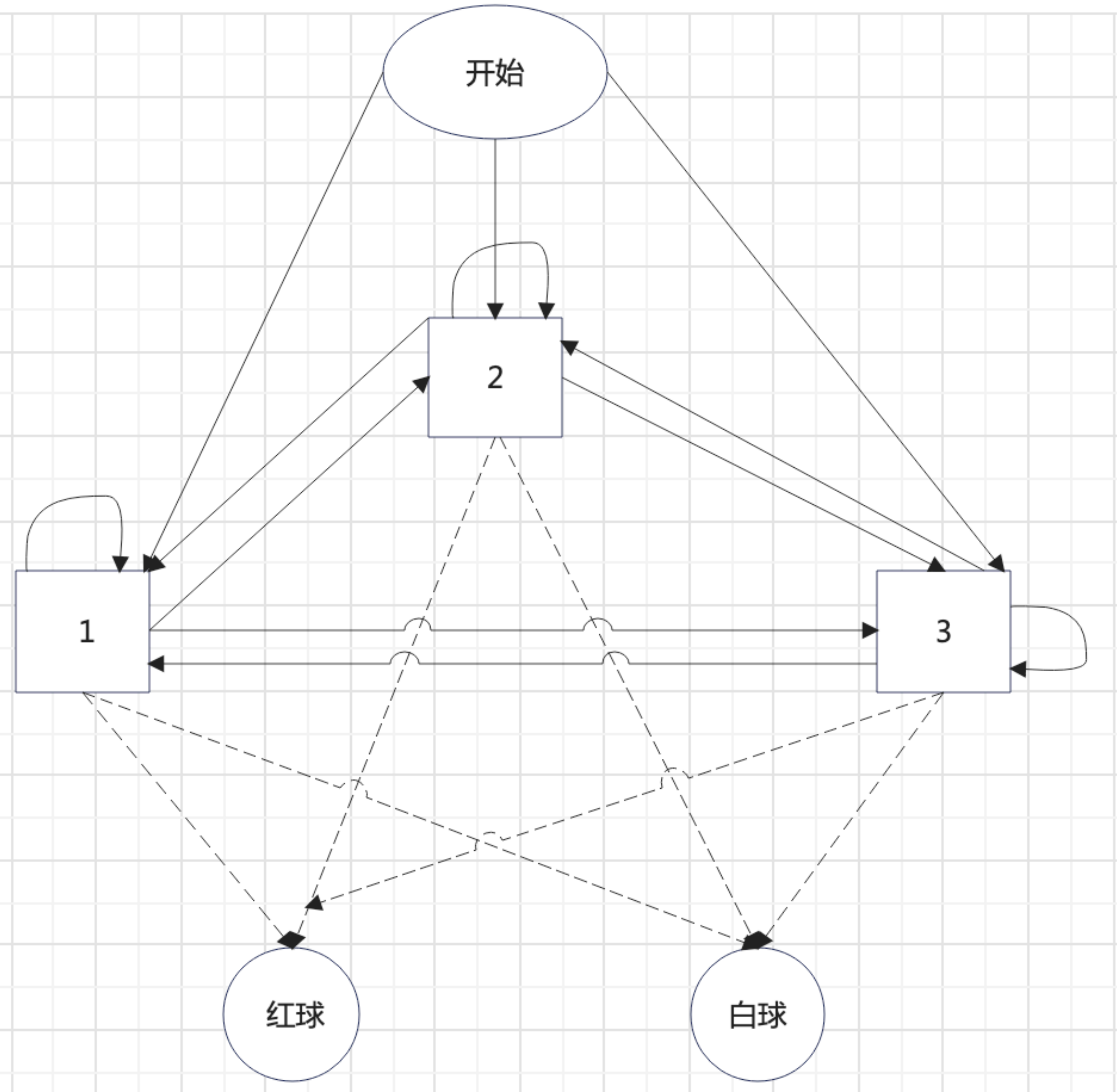


图 4-4　后向算法概率更新原理示意图

4.2.1 HMM模型评估问题

- **后向算法举例**

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



还是前面文章的那个例子，同样的条件

$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

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

$$\pi = (0.2, 0.4, 0.4)^T$$

求给定这个模型条件下，观测为O = (红，白，红) 的概率

4.2.1 HMM模型评估问题

- **后向算法举例**

$$\beta_t(j) = \sum_{i=1}^{n} \beta_{t+1}(i) e_i(o_{t+1}) a_{ji}$$

第三个观测是红球的情况：

- 第三个观测红球来自1号盒：$\beta_{red}(1) = 1$
- 第三个观测红球来自2号盒：$\beta_{red}(2) = 1$
- 第三个观测红球来自3号盒：$\beta_{red}(3) = 1$

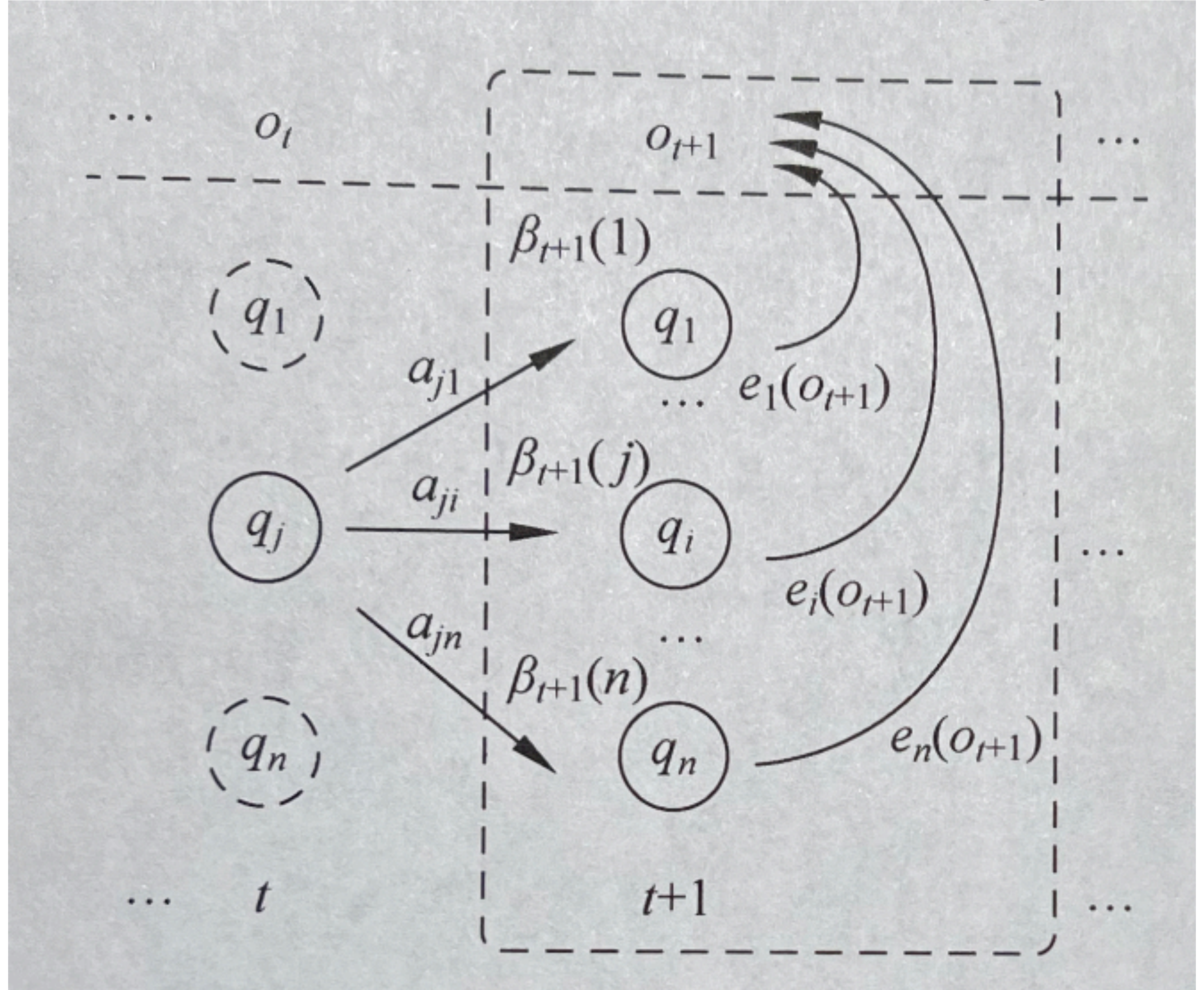


图 4-4 后向算法概率更新原理示意图

- 第二个观测白球来自1号盒：

$$\beta_{white}(1) = a_{11} b_{red}(1) \beta_{red}(1) + a_{12} b_{red}(2) \beta_{red}(2) + a_{13} b_{red}(3) \beta_{red}(3)$$

$$= 0.5 \times 0.5 \times 1 + 0.3 \times 0.4 \times 1 + 0.2 \times 0.7 \times 1 = 0.51$$

- 第二个观测白球来自2号盒：

$$\beta_{white}(2) = a_{21} b_{red}(1) \beta_{red}(1) + a_{22} b_{red}(2) \beta_{red}(2) + a_{23} b_{red}(3) \beta_{red}(3)$$

$$= 0.2 \times 0.5 \times 1 + 0.5 \times 0.4 \times 1 + 0.3 \times 0.7 \times 1 = 0.51$$

- 第二个观测白球来自3号盒：

$$\beta_{white}(3) = a_{31} b_{red}(1) \beta_{red}(1) + a_{32} b_{red}(2) \beta_{red}(2) + a_{33} b_{red}(3) \beta_{red}(3)$$

$$= 0.3 \times 0.5 \times 1 + 0.2 \times 0.4 \times 1 + 0.5 \times 0.7 \times 1 = 0.58$$

4.2.1 HMM模型评估问题

$$\boxed{\beta_t(j) = \sum_{i=1}^{n} \beta_{t+1}(i) e_i(o_{t+1}) a_{ji}}$$

- **后向算法举例**

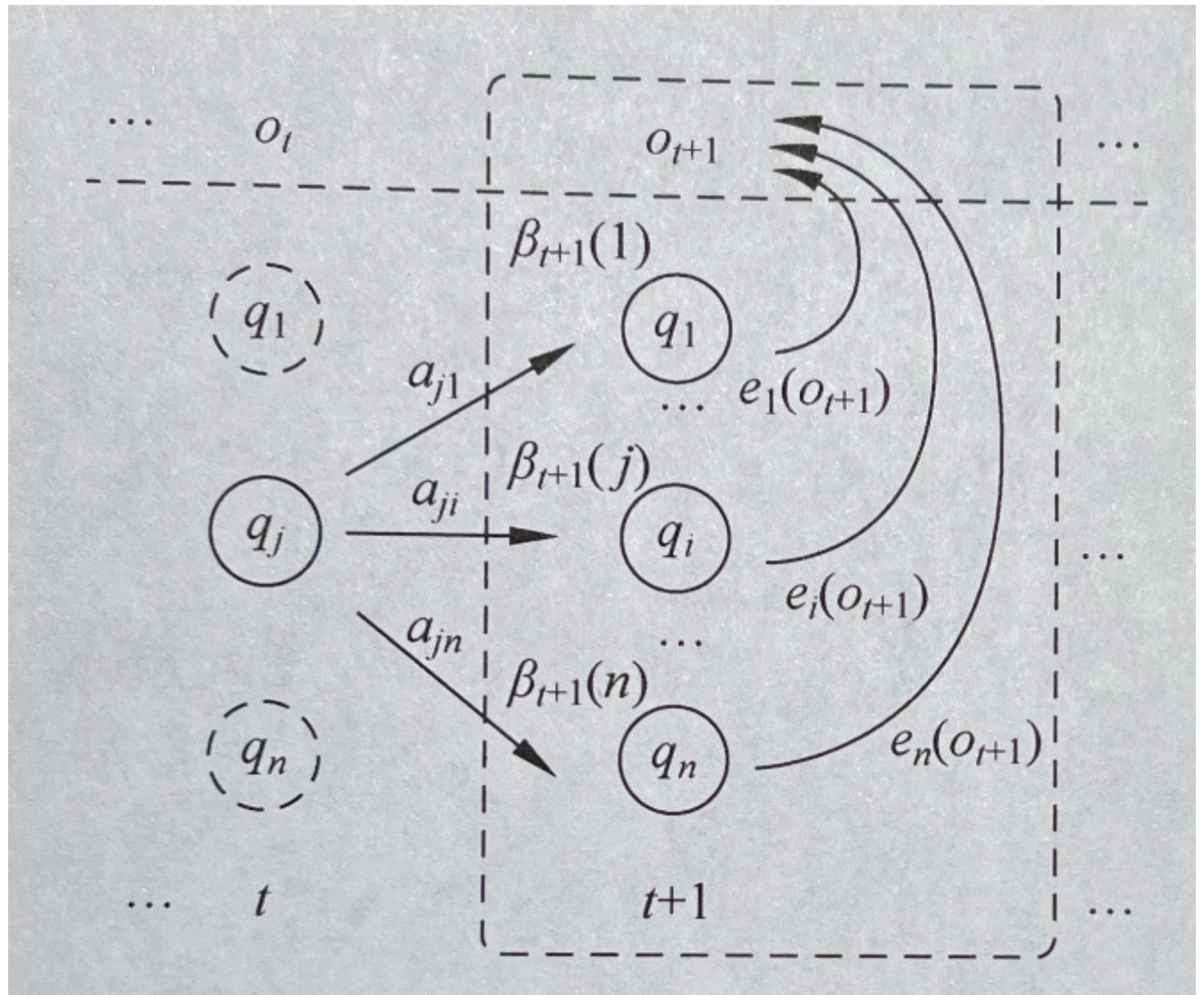


图 4-4　后向算法概率更新原理示意图

- 第一个观测红球来自1号盒:

$$\beta_{red}(1) = a_{11} b_{white}(1) \beta_{white}(1) + a_{12} b_{white}(2) \beta_{white}(2) + a_{13} b_{white}(3) \beta_{white}(3)$$

$$= 0.5 \times 0.5 \times 0.51 + 0.3 \times 0.6 \times 0.51 + 0.2 \times 0.3 \times 0.58 = 0.2541$$

- 第一个观测红球来自2号盒:

$$\beta_{red}(2) = a_{21} b_{white}(1) \beta_{white}(1) + a_{22} b_{white}(2) \beta_{white}(2) + a_{23} b_{white}(3) \beta_{white}(3)$$

$$= 0.2 \times 0.5 \times 0.51 + 0.5 \times 0.6 \times 0.51 + 0.3 \times 0.3 \times 0.58 = 0.51 = 0.2562$$

- 第一个观测红球来自3号盒:

$$\beta_{red}(3) = a_{31} b_{white}(1) \beta_{white}(1) + a_{32} b_{white}(2) \beta_{white}(2) + a_{33} b_{white}(3) \beta_{white}(3)$$

$$= 0.3 \times 0.5 \times 0.51 + 0.2 \times 0.6 \times 0.51 + 0.5 \times 0.3 \times 0.58 = 0.2247$$

4.2.1 HMM模型评估问题

- **后向算法举例**

逆向思考第一个观测是怎么来的？如何得知是从几号盒取的？答曰：初始值，所以结合初始值，这一条完整的“路径”才算走完

- 第一个观测红球来自1号盒（结合初始值）

$$\beta'_{red}(1) = \pi_1 b_{red}(1)\beta_{red}(1) = 0.2 \times 0.5 \times 0.2541 = 0.02541$$

- 第一个观测红球来自2号盒（结合初始值）

$$\beta'_{red}(2) = \pi_2 b_{red}(2)\beta_{red}(2) = 0.4 \times 0.4 \times 0.2562 = 0.0410$$

- 第一个观测红球来自3号盒（结合初始值）

$$\beta'_{red}(3) = \pi_3 b_{red}(3)\beta_{red}(3) = 0.4 \times 0.7 \times 0.2247 = 0.06292$$

终止：

$$P(O|\lambda) = \beta'_{red}(1) + \beta'_{red}(2) + \beta'_{red}(3) = 0.12932$$

以上“走法”都可以实现“红，白，红”的观测，我们是从后面状态向前推算的，所以称之为“后向算法”

4.2.1 HMM模型评估问题

- 前向算法和后向算法统一

$$P(O \mid M) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i(i) a_{ij} e_j\left(o_{t+1}\right) \beta_{t+1}(j),\ t = 1,2,\cdots,L-1$$

4.2.2 HMM隐状态推断问题

已知HMM模型M和观测序列O，求产生该序列的最有可能的**隐状态序列** $\boldsymbol{x} = \arg\max_{x} p(\boldsymbol{x}, O \mid M)$

- 维特比（Viterbi）算法
  - 一种利用动态规划**求解最优路径**的算法，一条路径对应一个状态序列
  - 递推计算在每一时刻各条部分路径的最大概率

$$v_t(j) = \max_{x_1 \cdots x_{t-1}} p(x_1 \cdots x_{t-1}, o_1 \cdots o_t, x_t = q_j \mid M)$$

$$v_t(j) = \max_{i=1:n} v_{t-1}(i) a_{ij} e_j(o_t)$$

$$pa_t(j) = \arg\max_{i=1:n} v_{t-1}(i) a_{ij} e_j(o_i)$$

  - 记录最大值对应的隐状态路径

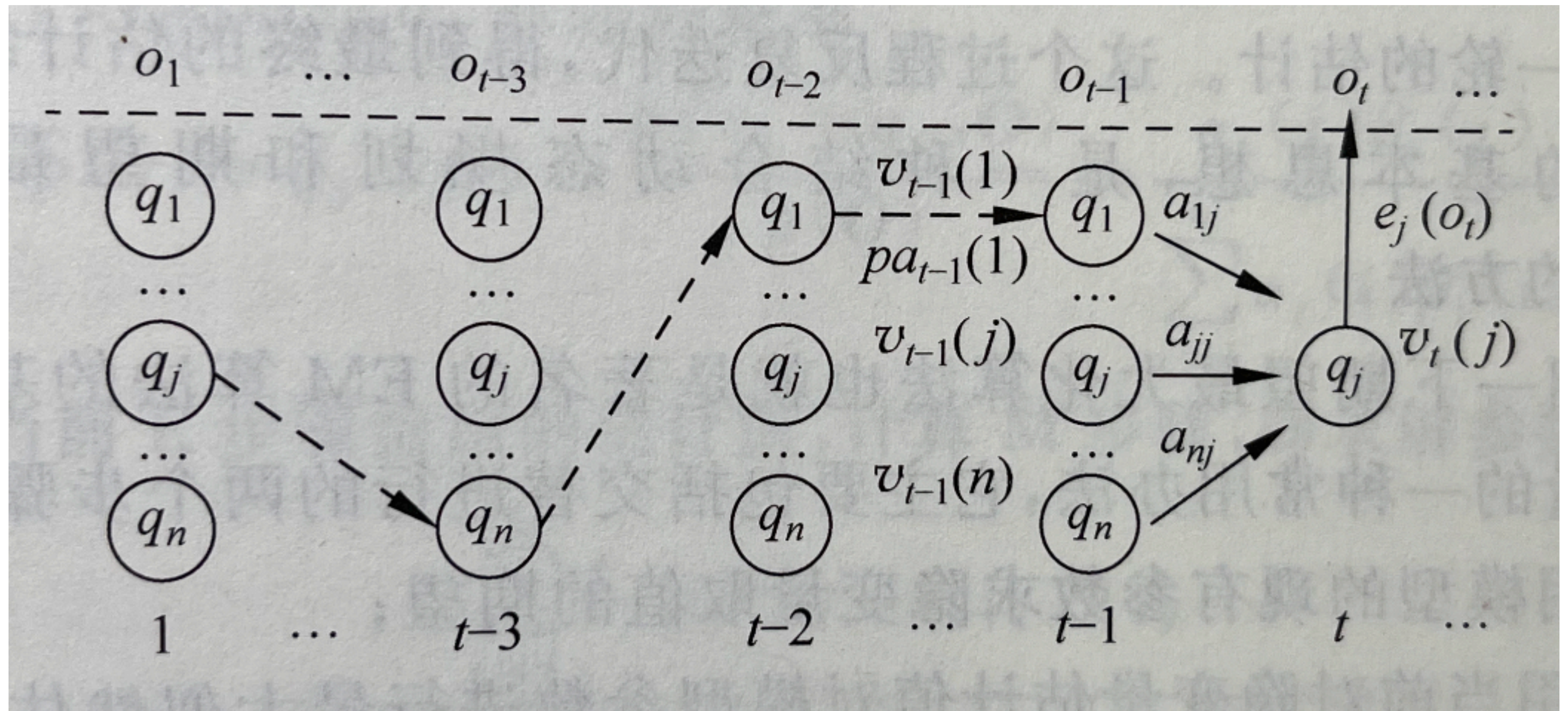
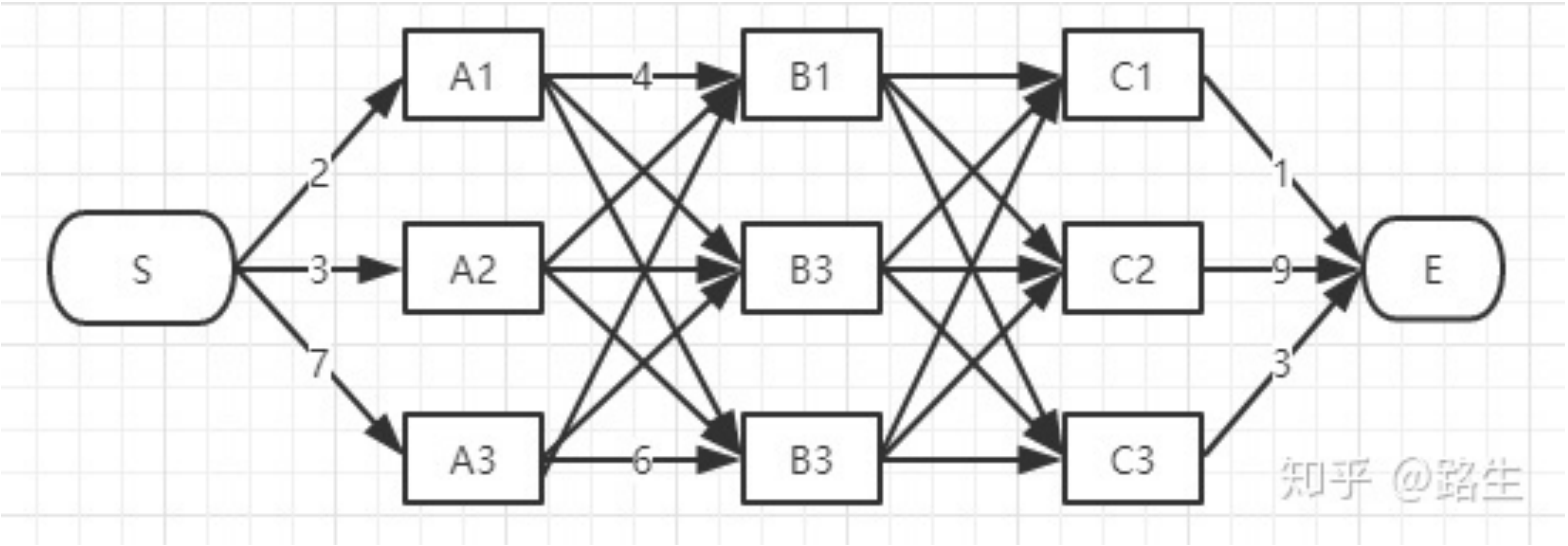
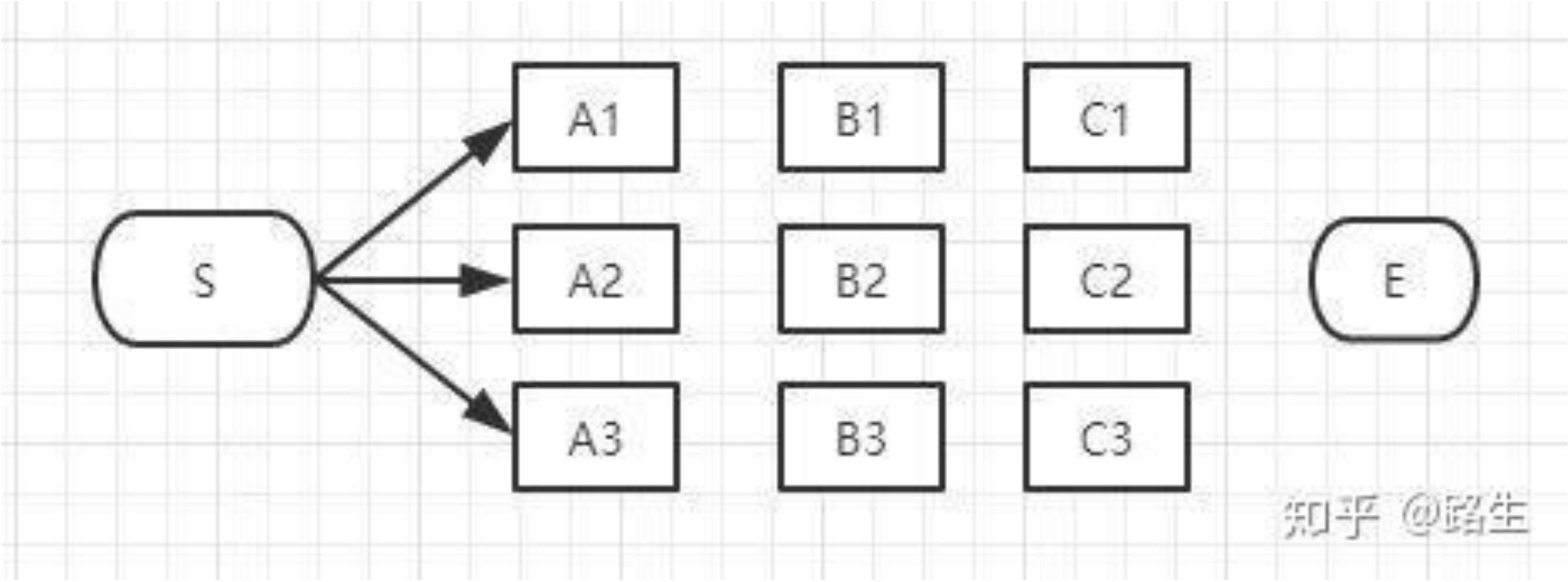


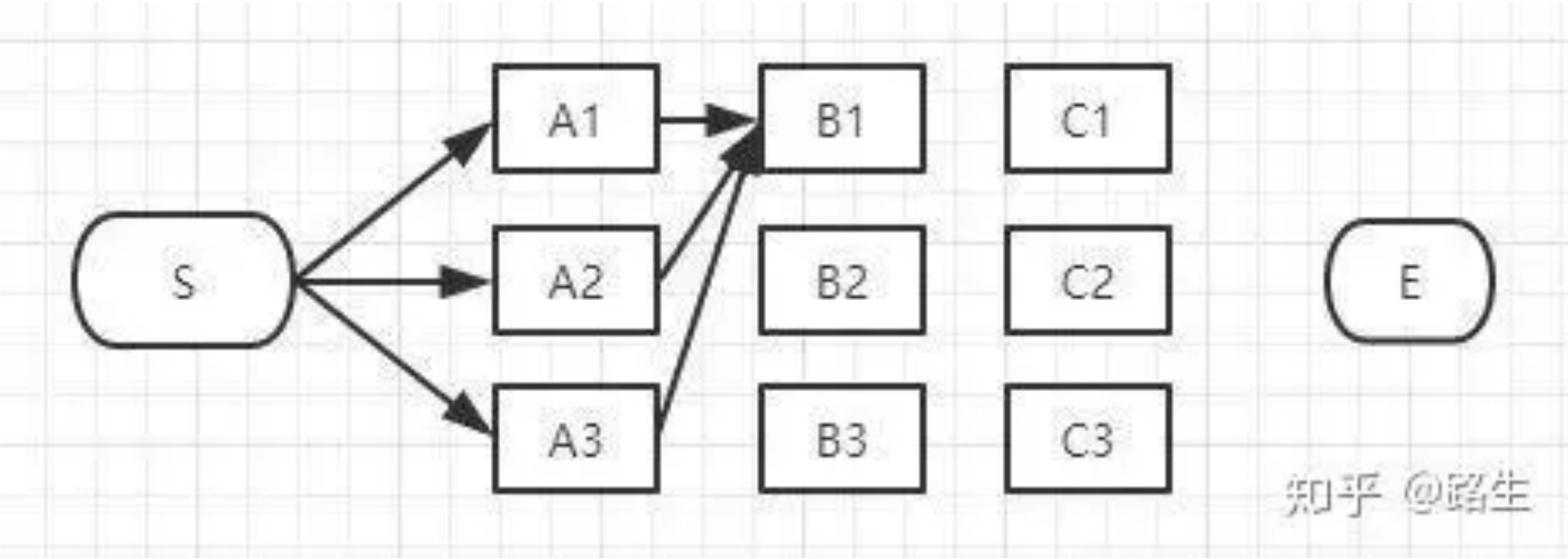
图 4-5 维特比算法的基本原理示意图

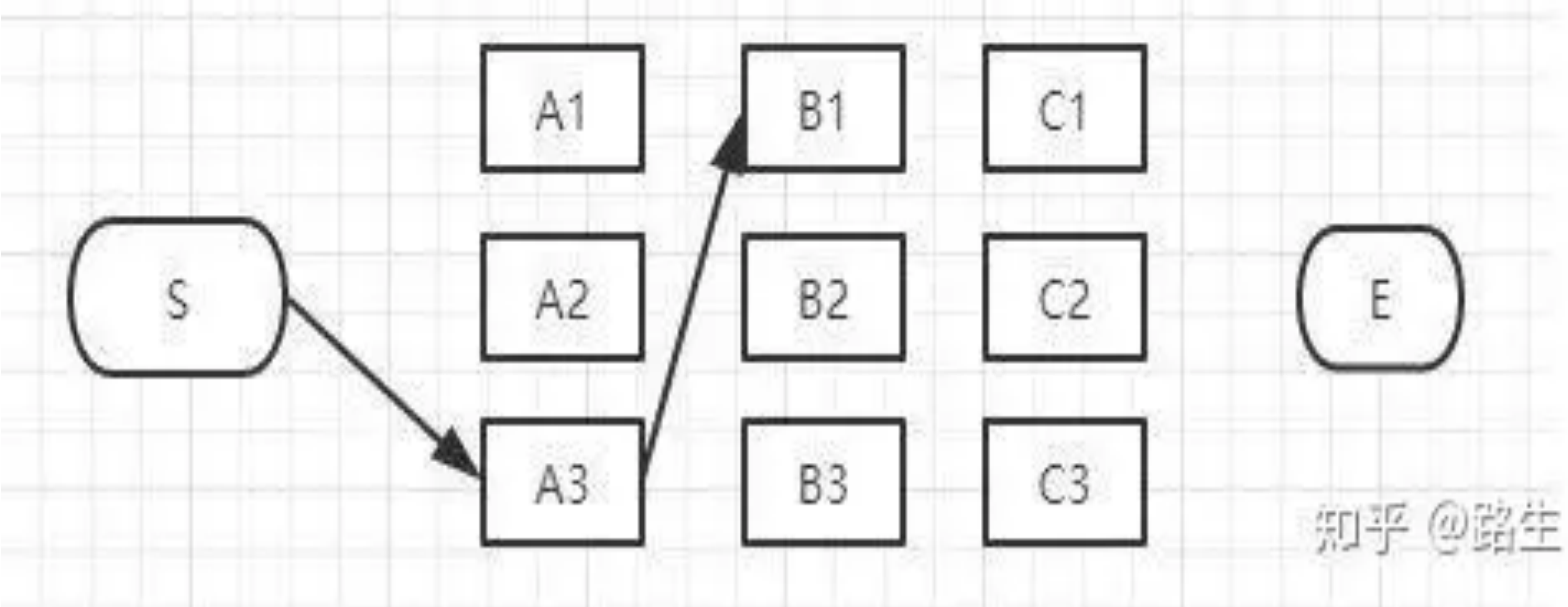
4.2.2 HMM隐状态推断问题



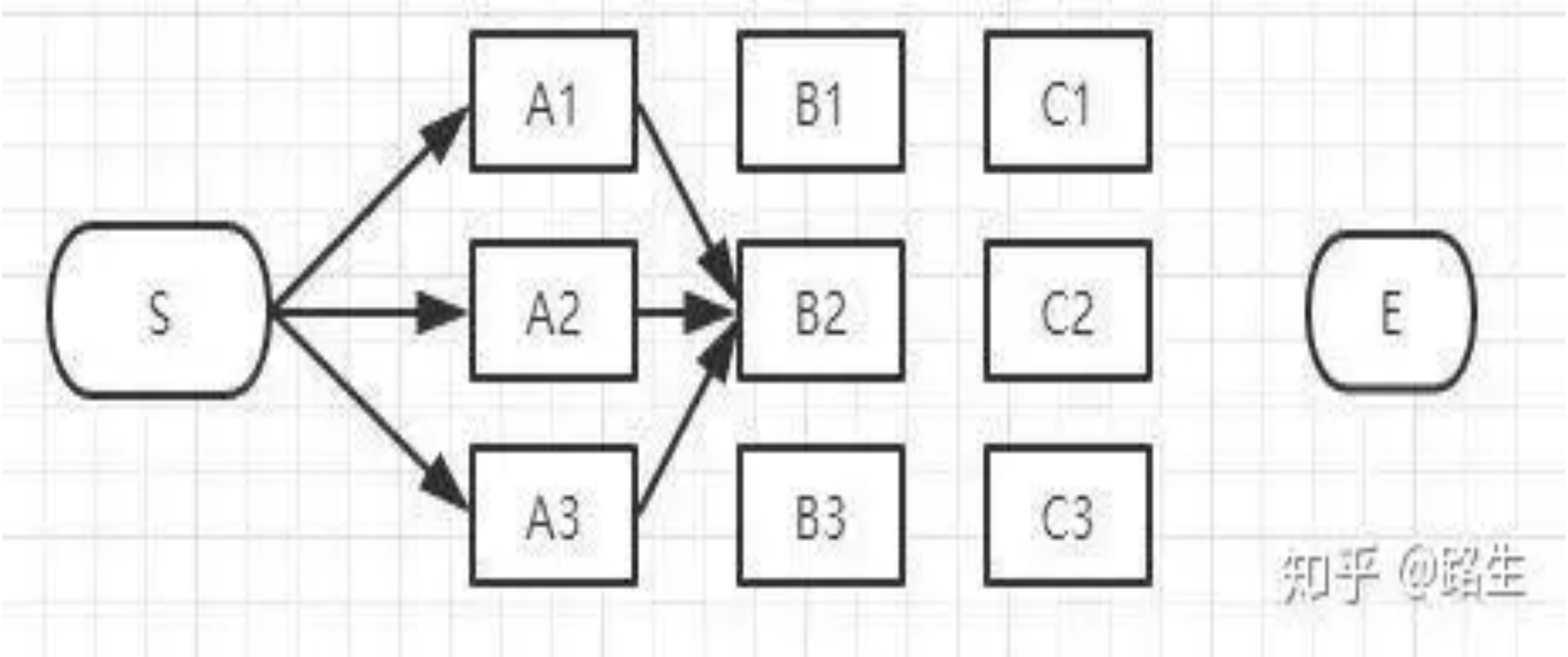
4.2.2 HMM隐状态推断问题

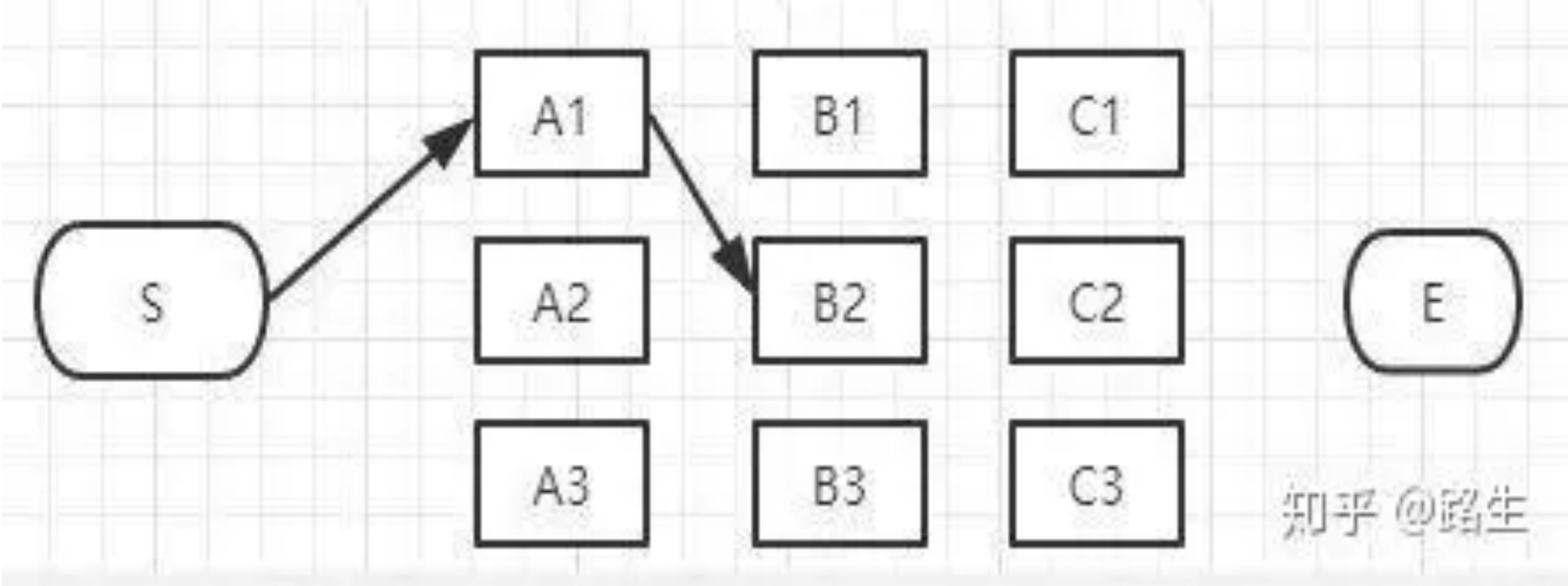


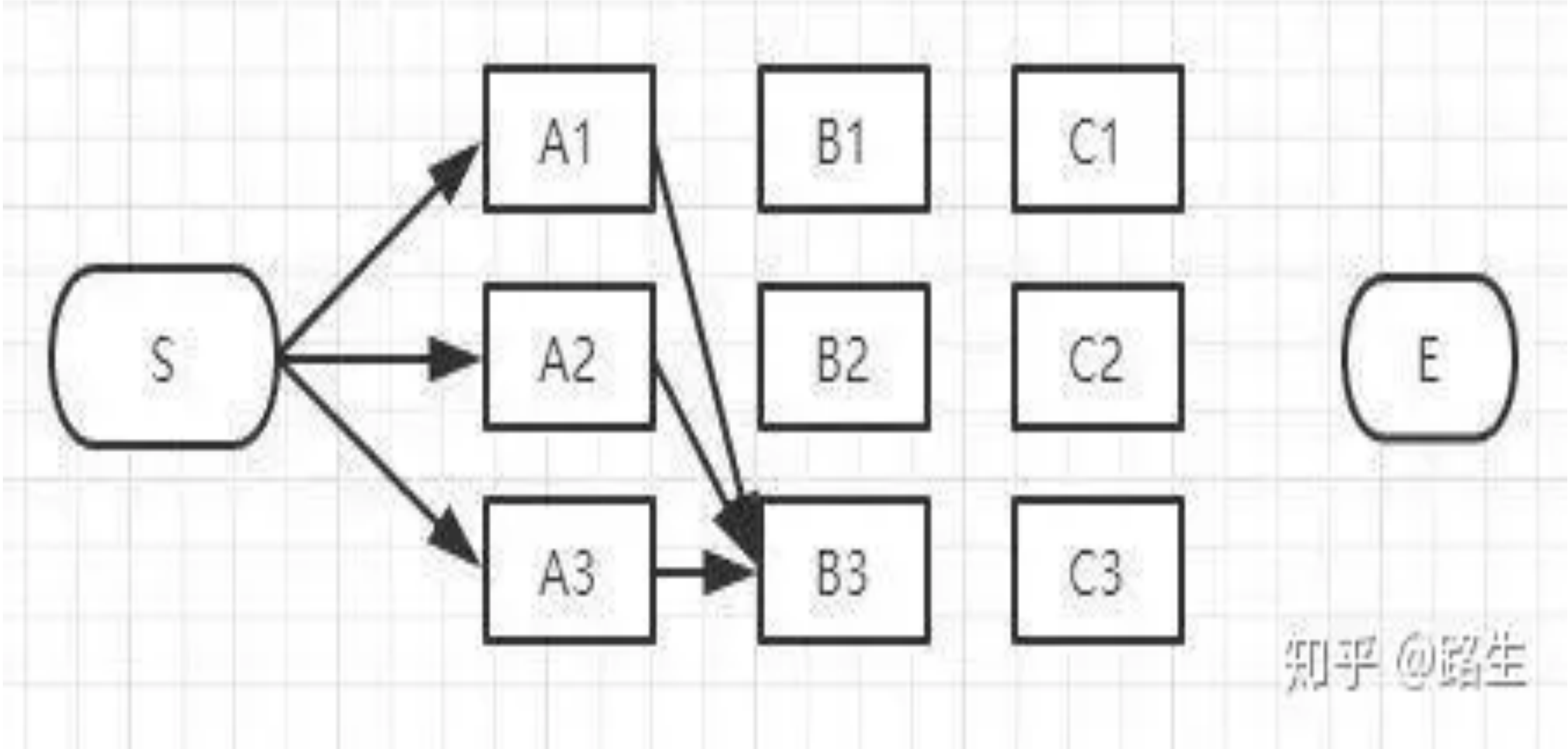


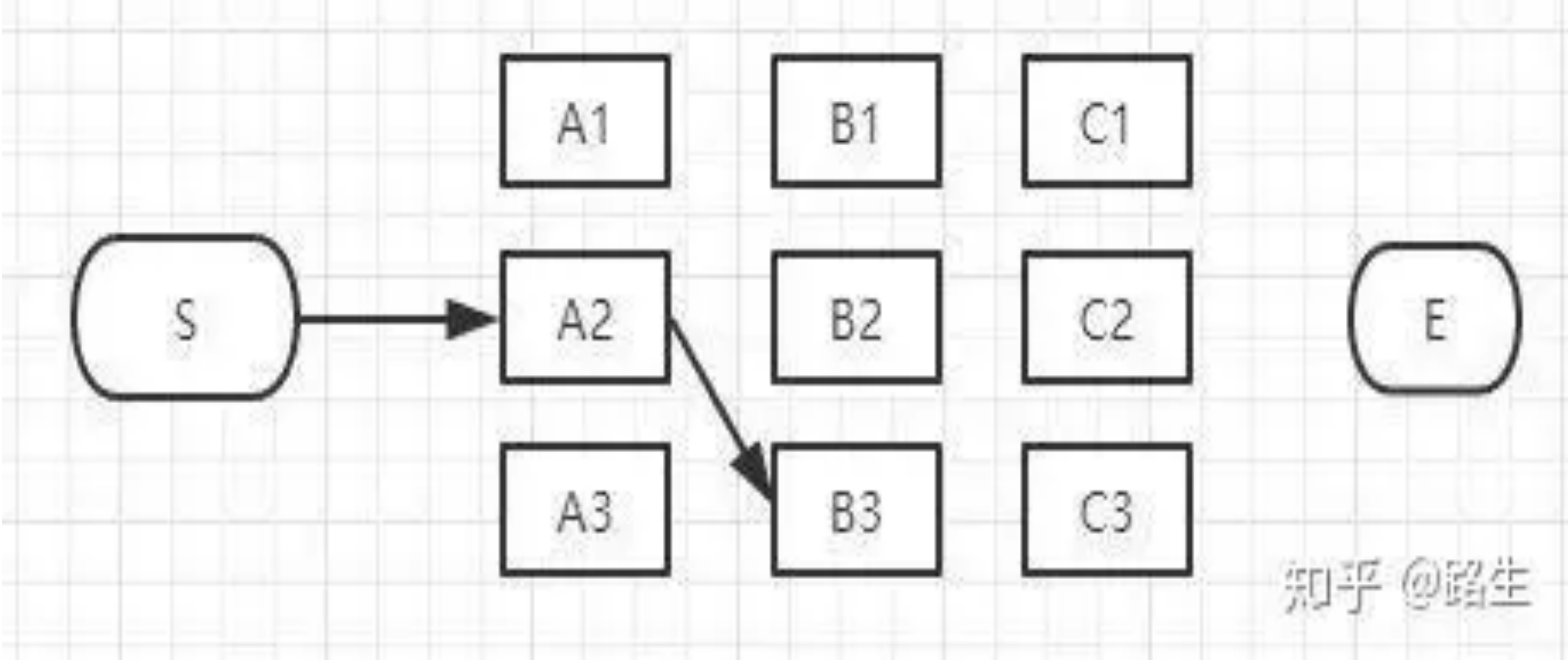


4.2.2 HMM隐状态推断问题

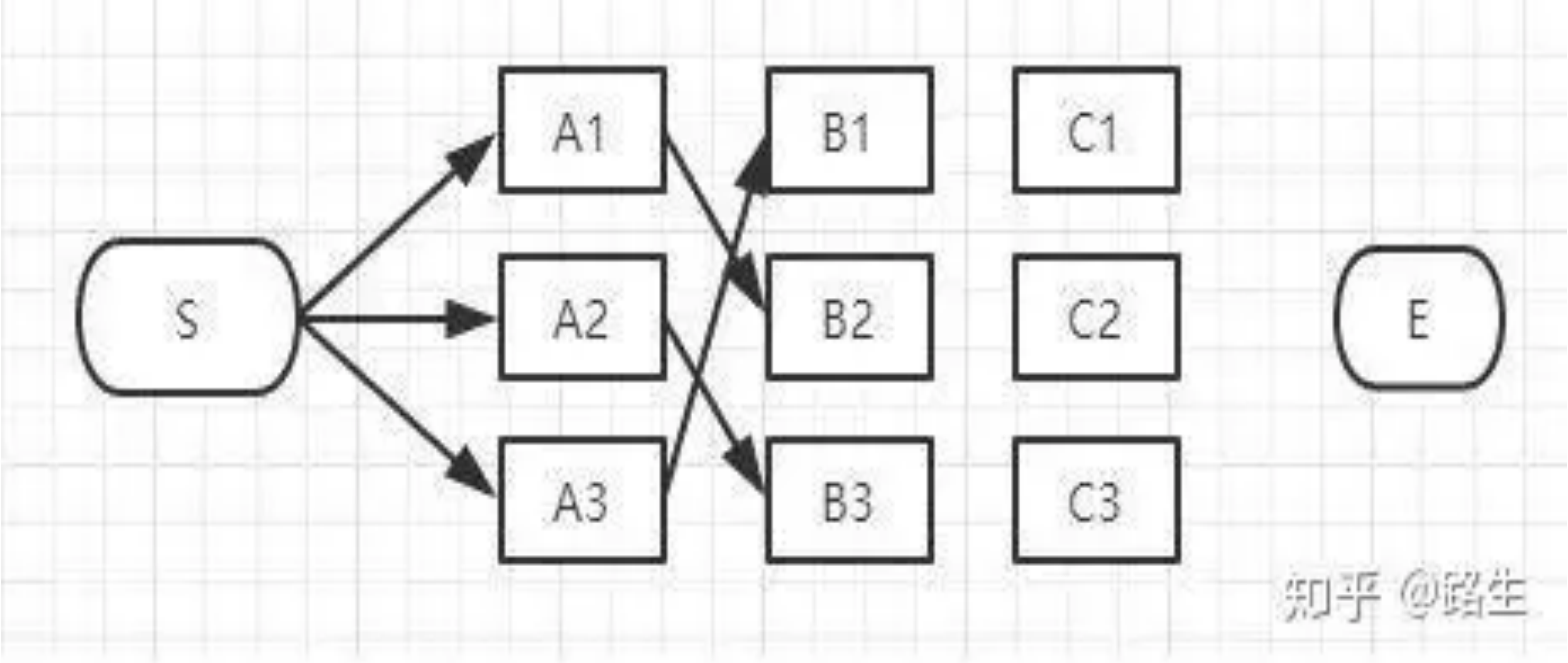


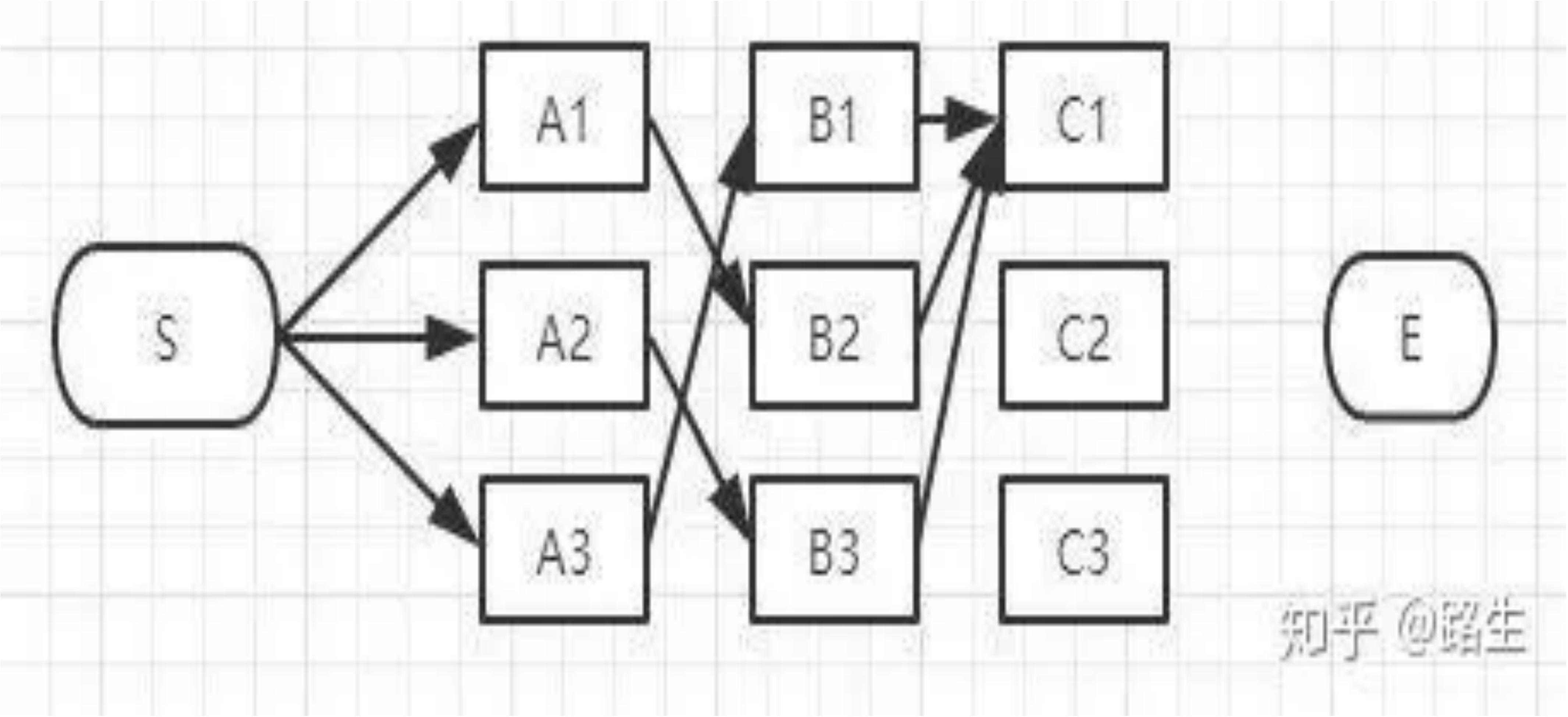


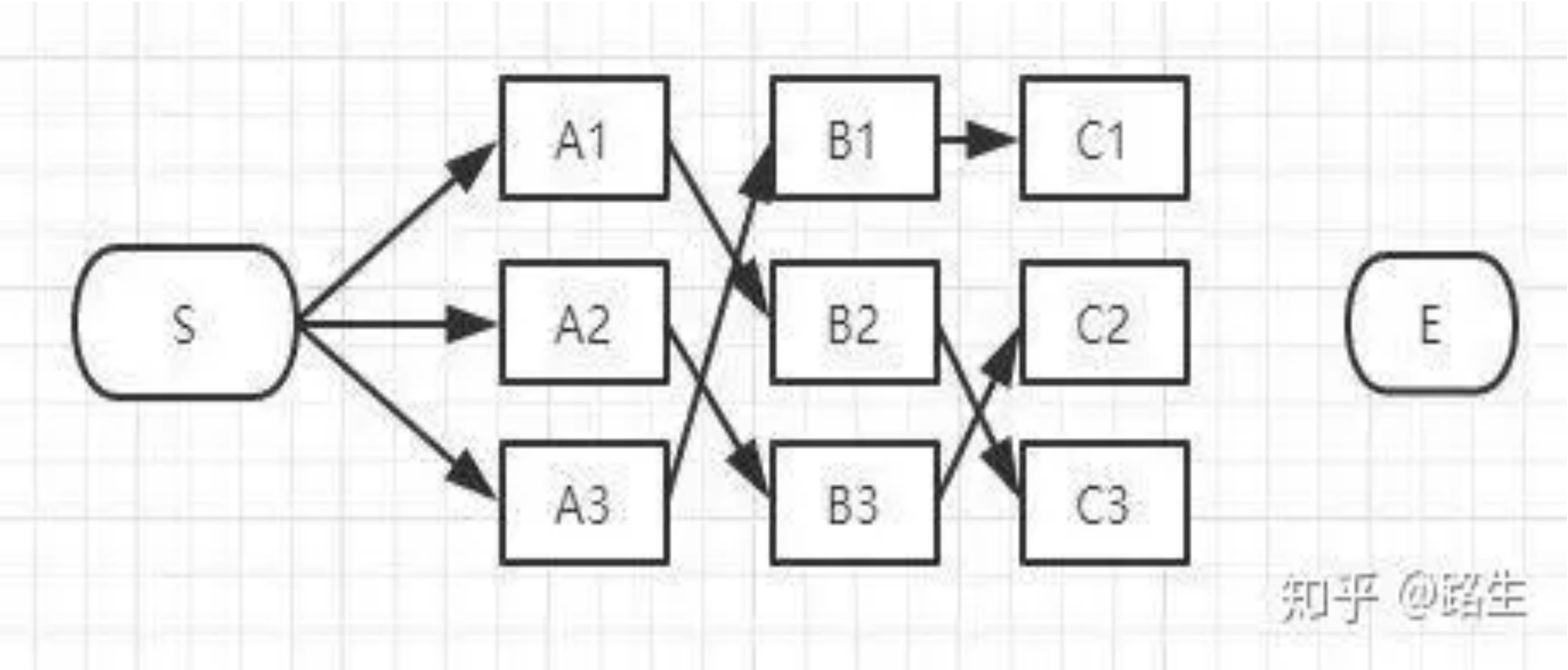




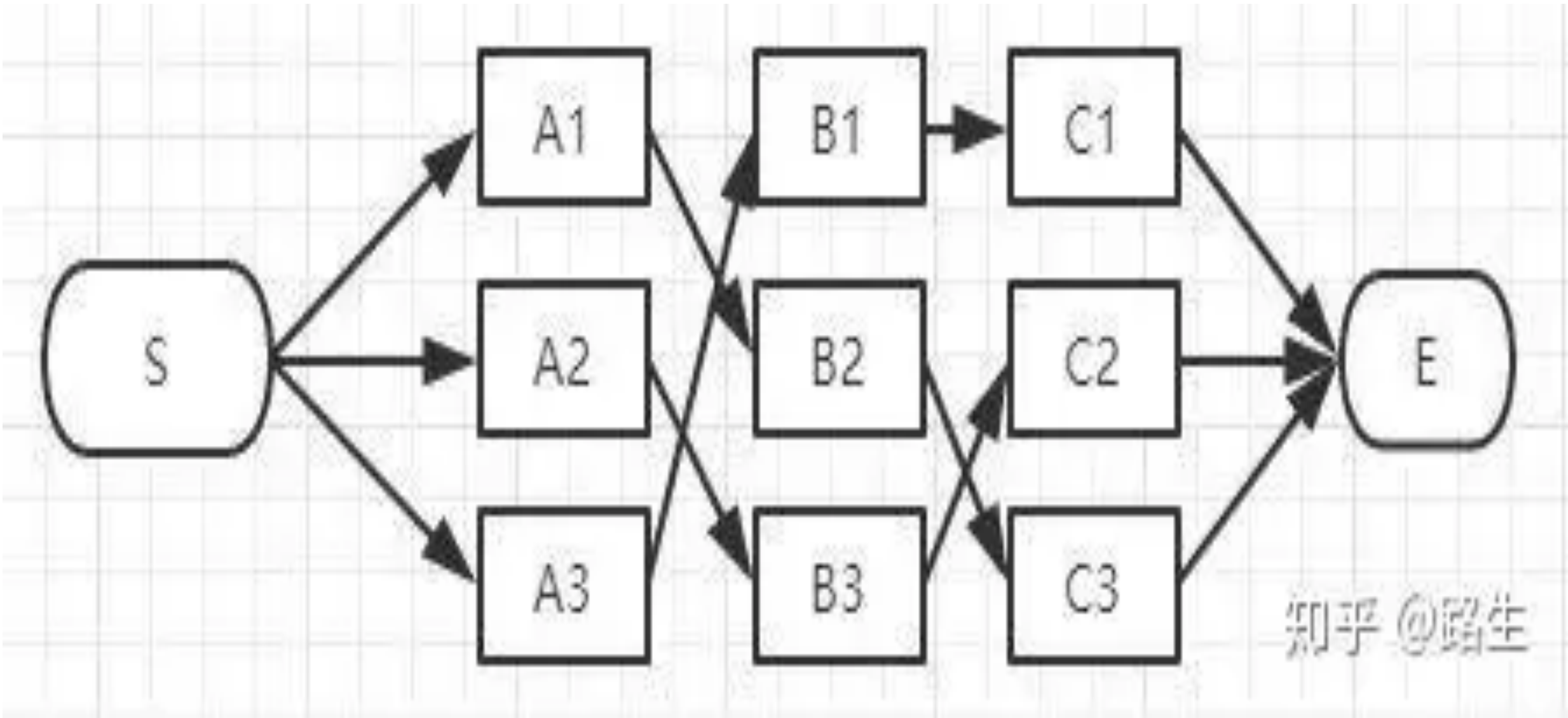
4.2.2 HMM隐状态推断问题

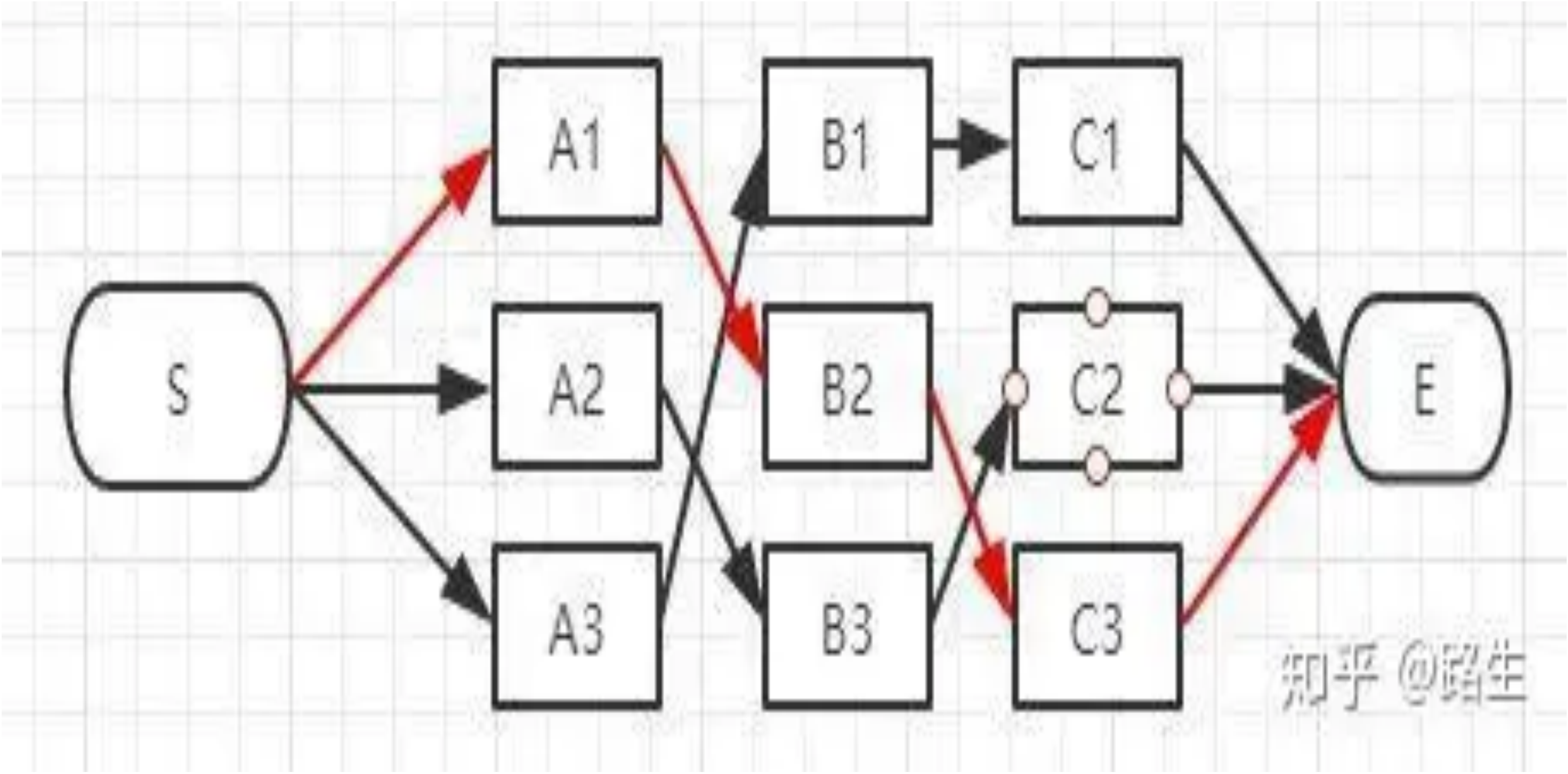






4.2.2 HMM隐状态推断问题





4.2.3 HMM模型学习问题

已知模型结构、观测值取值范围V、模型隐状态集Q，根据观测序列O的样本，学习模型参数A、E、π。

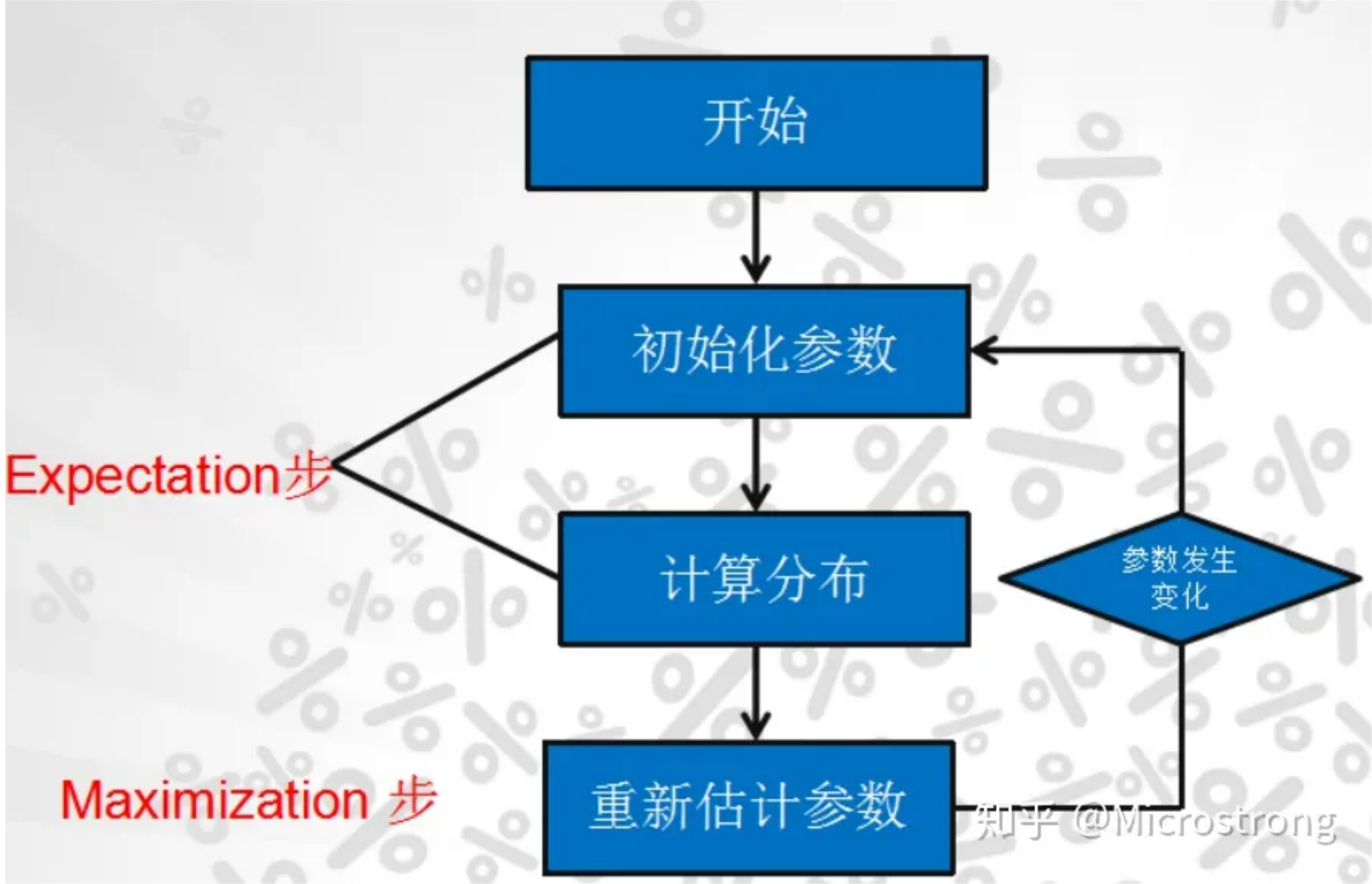
- 隐状态序列 $\boldsymbol{x}$ 已知：**最大似然估计法**

$$\hat{a}_{ij} = \frac{\#T_{ij}}{\sum_{j'}^{n} \#T_{ij'}}, \quad \hat{e}_{ik} = \frac{\#E_{ij}}{\sum_{k'}^{V} \#E_{ij'}}, \quad \hat{\pi}_i = \frac{\#S_i}{S}$$

- 隐状态序列 $\boldsymbol{x}$ 未知：**最大方差估计法**（Expectation-Maximum，EM）

4.2.3 HMM模型学习问题

- **EM算法**
  - **E步骤**：利用模型现有参数求隐变量取值的期望
  - **M步骤**：利用当前对隐变量估计值对模型参数最大似然估计，更新模型



4.2.3 HMM模型学习问题

- **EM算法**

输入：观察到的数据$x=(x_1,x_2,...x_n)$，联合分布 $p(x,z;\theta)$，条件分布 $p(z|x,\theta)$，最大迭代次数J。

算法步骤：

（1）随机初始化模型参数θ的初值 θ_0 。

（2）j=1,2,...,J 开始EM算法迭代：

- E步：计算联合分布的条件概率期望：

$$Q_i(z_i)=p(z_i|x_i,\theta_j)$$

$$l(\theta,\theta_j)=\sum_{i=1}^{n}\sum_{z_i}Q_i(z_i)log\frac{p(x_i,z_i;\theta)}{Q_i(z_i)}$$

- M步：极大化 $l(\theta,\theta_j)$,得到 θ_{j+1} ：

$$\theta_{j+1}=argmaxl(\theta,\theta_j)$$

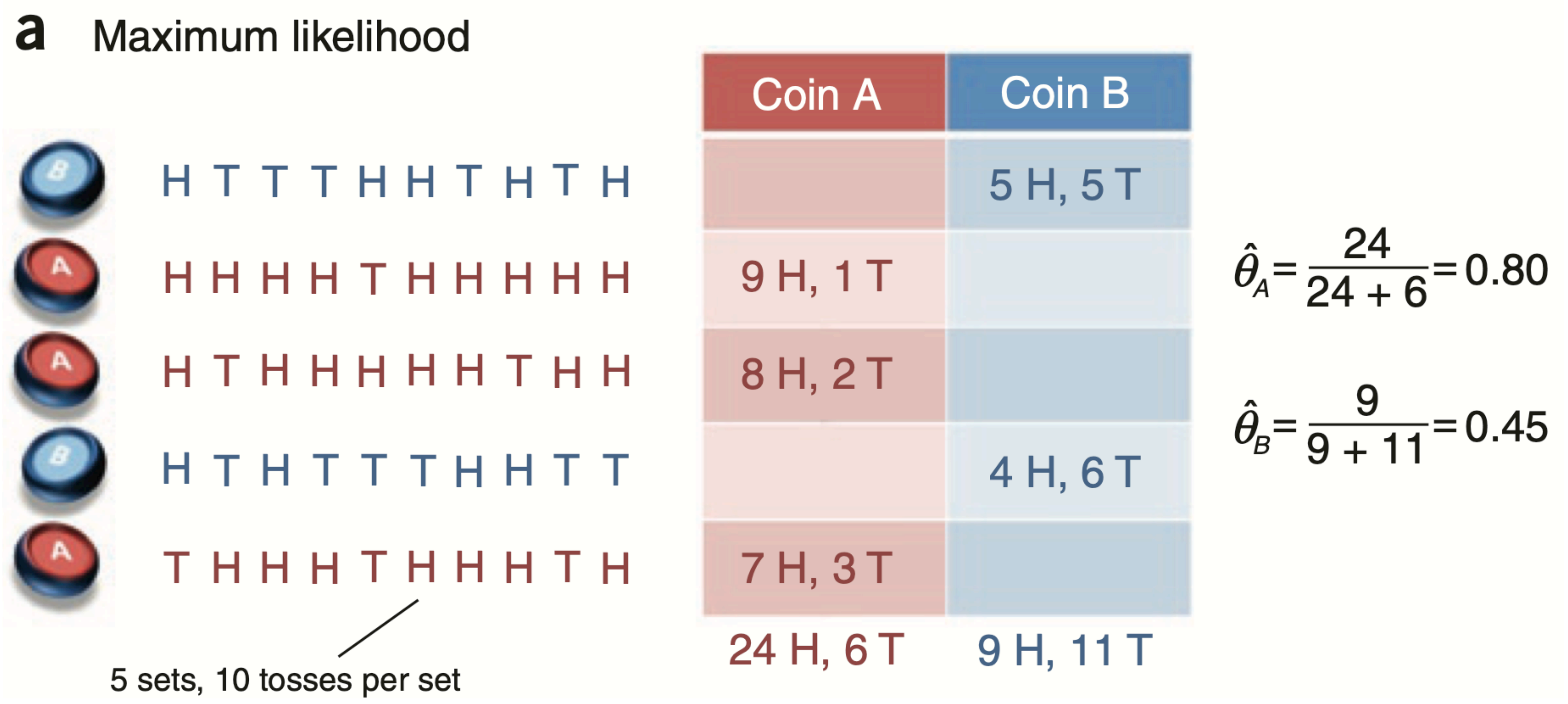
- 如果θ_{j+1} 已经收敛，则算法结束。否则继续进行E步和M步进行迭代。

输出：模型参数θ。

4.2.3 HMM模型学习问题

例：假设有两枚硬币A、B，以相同的概率随机选择一个硬币，进行如下的掷硬币实验：共做 5 次实验，每次实验独立的掷十次 (H代表正面朝上)。a 是在知道每次选择的是A还是B的情况下进行，b是在不知道选择的是A还是B的情况下进行，问如何估计两个硬币正面出现的概率?

4.2.3 HMM模型学习问题



CASE a

已知每个实验选择的是硬币A 还是硬币 B，重点是如何计算输出的概率分布，这其实也是极大似然求导所得。

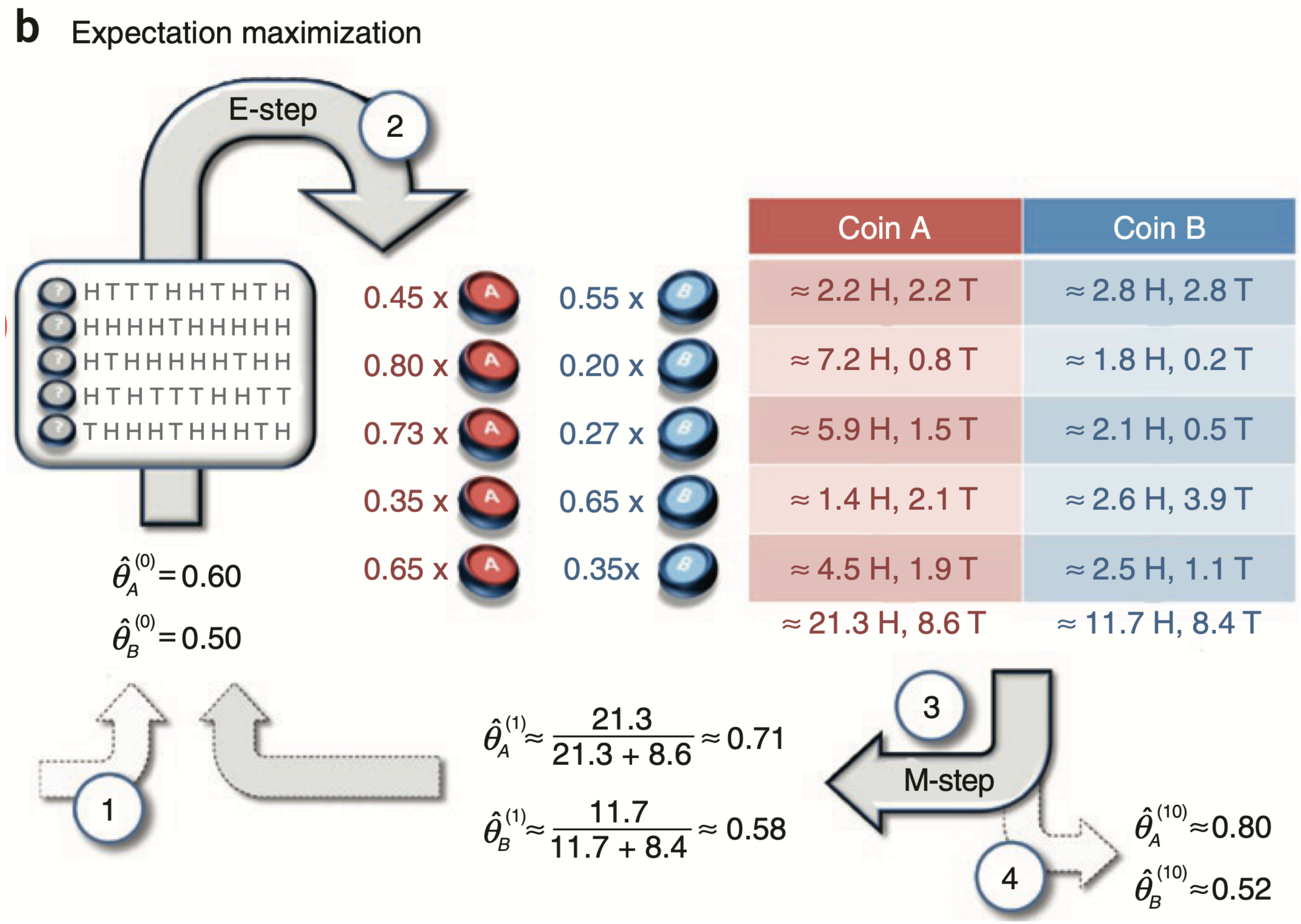
$$\begin{aligned}\underset{\theta}{argmax} \log P(Y|\theta) &= \log((\theta_B^5(1-\theta_B)^5)(\theta_A^9(1-\theta_A))(\theta_A^8(1-\theta_A)^2)(\theta_B^4(1-\theta_B)^6)(\theta_A^7(1-\theta_A)^3)) \\ &= \log[(\theta_A^{24}(1-\theta_A)^6)(\theta_B^9(1-\theta_B)^{11})]\end{aligned}$$

上面这个式子求导之后发现，5 次实验中A正面向上的次数再除以总次数作为即为 $\hat{\theta}_A$ ， 5次实验中B正面向上的次数再除以总次数作为即为 ，即:

$$\hat{\theta}_A = \frac{24}{24+6} = 0.80$$

$$\hat{\theta}_B = \frac{9}{9+11} = 0.45$$

4.2.3 HMM模型学习问题



CASE b

由于并不知道选择的是硬币 A 还是硬币 B，因此采用EM算法。

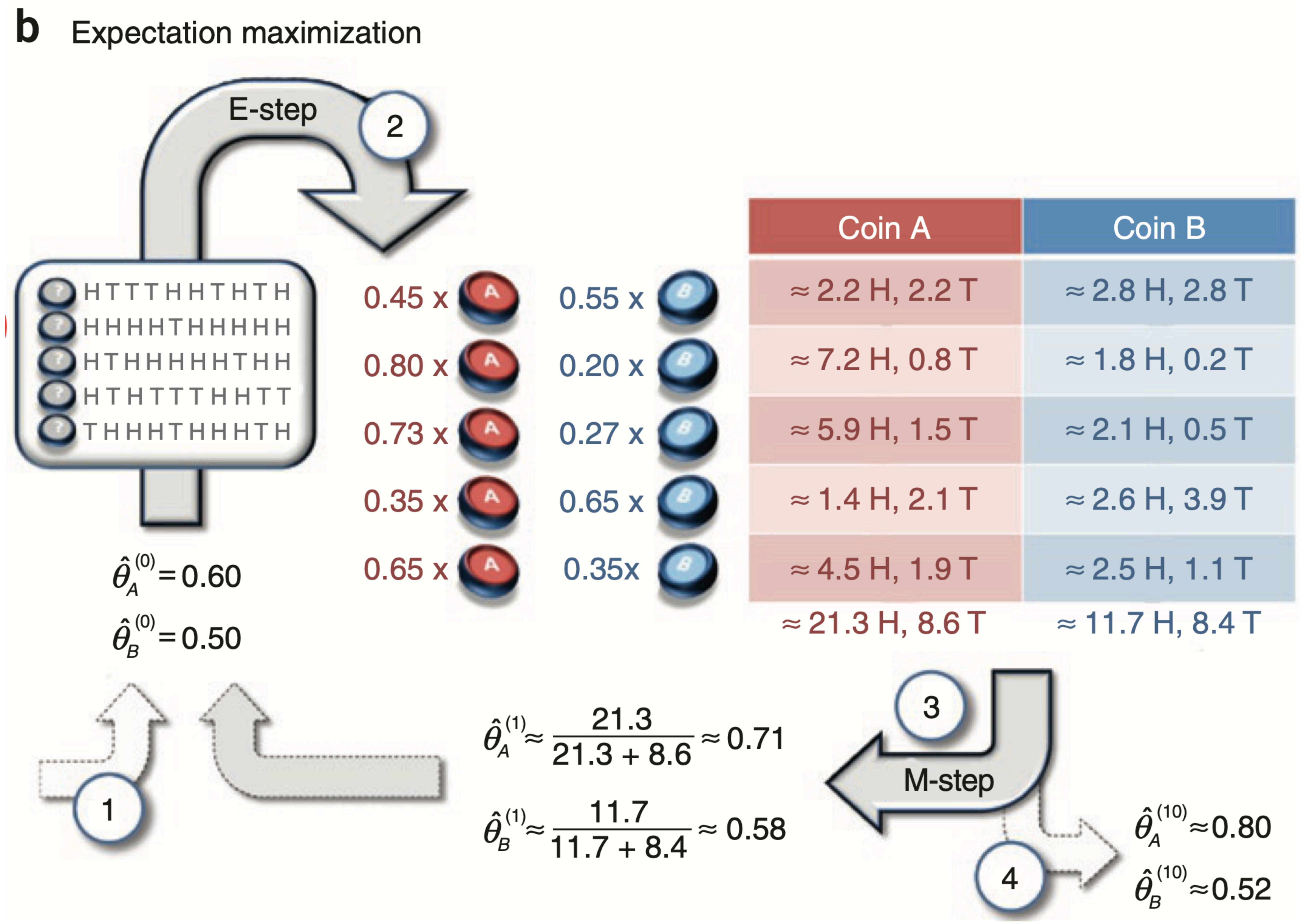
E步：初始化$\hat{\theta}_A^{(0)} = 0.60$和 $\hat{\theta}_B^{(0)} = 0.50$，计算每个实验中选择的硬币是 A 和 B 的概率，例如第一个实验中选择 A 的概率为：

$$P(z = A|y_1, \theta) = \frac{P(z = A, y_1|\theta)}{P(z = A, y_1|\theta) + P(z = B, y_1|\theta)} = \frac{(0.6)^5 * (0.4)^5}{(0.6)^5 * (0.4)^5 + (0.5)^{10}} = 0.45$$

$$P(z = B|y_1, \theta) = 1 - P(z = A|y_1, \theta) = 0.55$$

计算出每个实验为硬币 A 和硬币 B 的概率，然后进行加权求和。

4.2.3 HMM模型学习问题



M步：求出似然函数下界 $Q(\theta,\theta^i)$，y_j代表第 j 次实验正面朝上的个数，μ_j 代表第 j 次实验选择硬币 A 的概率，$1-\mu_j$ 代表第 j 次实验选择硬币B的概率 。

$$Q(\theta,\theta^i) = \sum_{j=1}^{5}\sum_{z} P(z|y_j,\theta^i)logP(y_j,z|\theta)$$

$$= \sum_{j=1}^{5} \mu_j log(\theta_A^{y_j}(1-\theta_A)^{10-y_j}) + (1-\mu_j)log(\theta_B^{y_j}(1-\theta_B)^{10-y_j})$$

针对L函数求导来对参数求导，例如对 θ_A 求导：

$$\frac{\partial Q}{\partial \theta_A} = \mu_1(\frac{y_1}{\theta_A} - \frac{10-y_1}{1-\theta_A}) + \cdots + \mu_5(\frac{y_5}{\theta_A} - \frac{10-y_5}{1-\theta_A}) = \mu_1(\frac{y_1-10\theta_A}{\theta_A(1-\theta_A)}) + \cdots$$

$$+ \mu_5(\frac{y_5-10\theta_A}{\theta_A(1-\theta_A)})$$

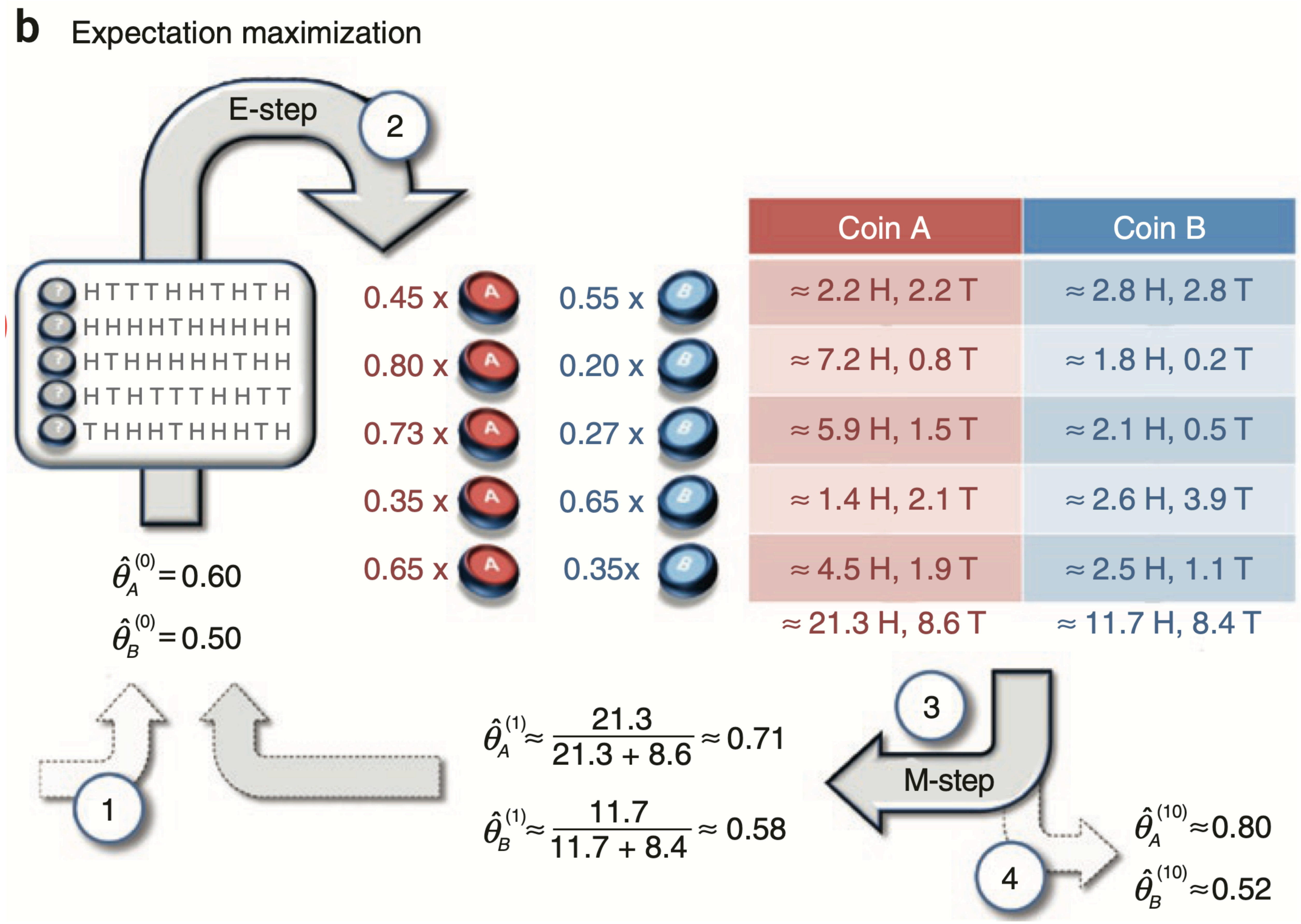
$$= \frac{\sum_{j=1}^{5}\mu_j y_j - \sum_{j=1}^{5} 10\mu_j\theta_A}{\theta_A(1-\theta_A)}$$

求导等于 0 之后就可得到图中的第一次迭代之后的参数值:

$$\hat{\theta}_A^{(1)} = 0.71$$

$$\hat{\theta}_B^{(1)} = 0.58$$

4.2.3 HMM模型学习问题



第二轮迭代：基于第一轮EM计算好的 $\hat{\theta}_A^{(1)}, \hat{\theta}_B^{(1)}$ ，进行第二轮 EM，计算每个实验中选择的硬币是 A 和 B 的概率（E步），然后在计算M步，如此继续迭代......迭代十步之后

$$\hat{\theta}_A^{(10)} = 0.8, \hat{\theta}_B^{(10)} = 0.52$$

4.3 朴素贝叶斯分类器

- **贝叶斯分类器**

$$P(B_i \mid A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A \mid B_i)}{P(A)}$$

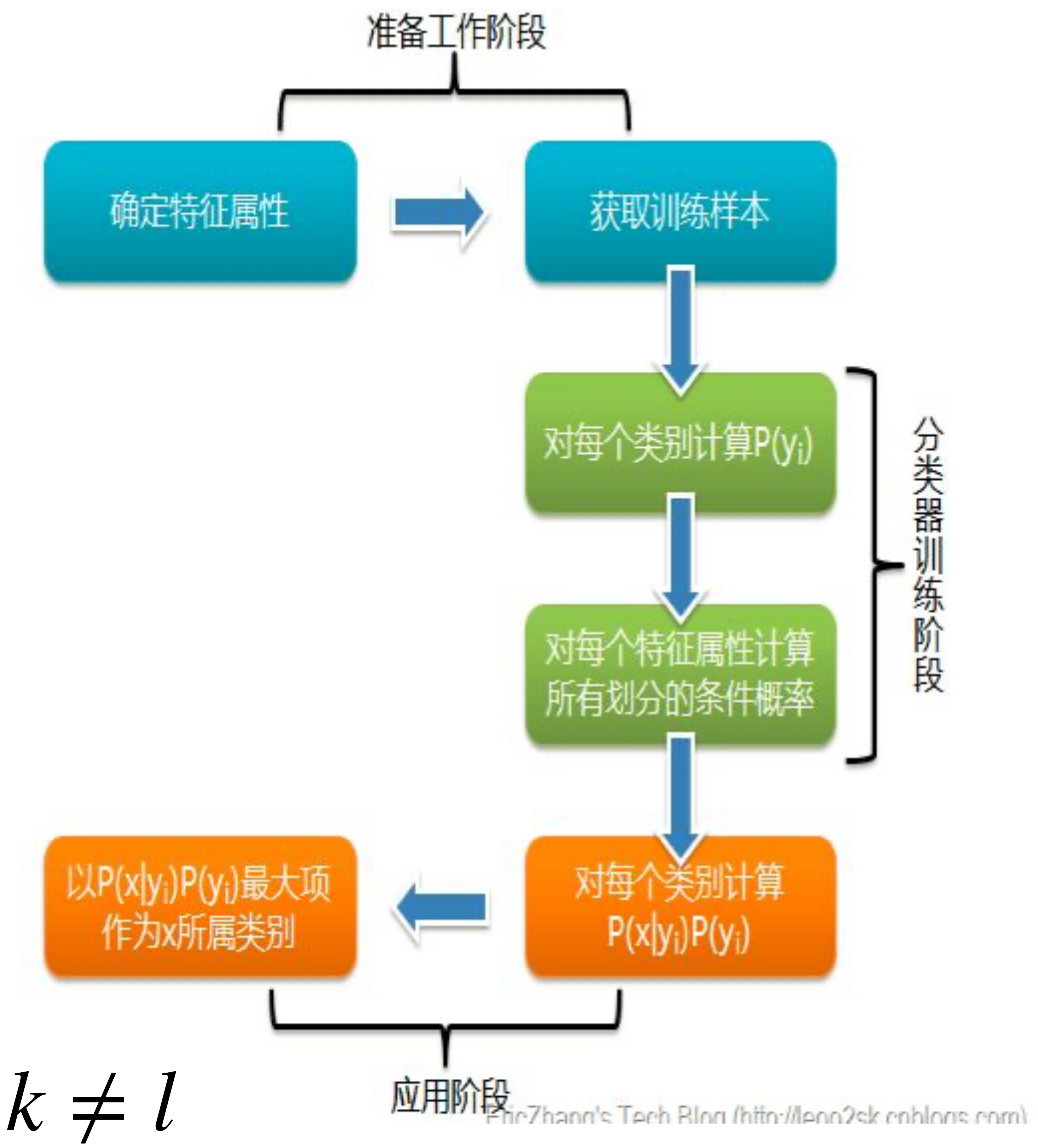
$$P(\text{类别} \mid \text{特征}) = \frac{P(\text{特征} \mid \text{类别})P(\text{类别})}{P(\text{特征})}$$

- **朴素贝叶斯分类器**
  - 假设各个特征取值只依赖类别标签，特征间互相独立：**朴素**

$$p(x_l x_k \mid \omega_i) = \mathrm{p}\left(x_l \mid \omega_i\right)\mathrm{p}\left(x_k \mid \omega_i\right), \quad l, k = 1, \cdots, d, k \neq l$$

  - 联合概率可以分解为

$$p\left(x_1, x_2, \cdots, x_d, \omega_i\right) = p\left(x_1 \mid \omega_i\right)p\left(x_2 \mid \omega_i\right)\cdots p\left(x_d \mid \omega_i\right)p(\omega_i)$$



4.3 朴素贝叶斯分类器

- **朴素贝叶斯分类器**
  - 类别先验概率估计

$$p(Y=\omega_\mathrm{i})=\frac{\sum_{i=1}^{N} I(y_i=\omega_i)}{N}$$

  - 特征的条件概率：参数的极大似然估计

$$p\left(x_j=v_i \middle| Y=\omega_i\right)=\frac{\sum_{i=1}^{N} I(x_j^{(i)}=v_i, y_i=\omega_i)}{\sum_{i=1}^{N} I(y_i=\omega_i)}$$

4.3 朴素贝叶斯分类器

- **朴素贝叶斯分类器**
  - 拉普拉斯平滑：对概率值平滑矫正，解决样本量少或者特征取值概率较低的情况
    - $$p(Y=\omega_i)=\frac{\sum_{i=1}^{N} I(y_i=\omega_i)+1}{N+C}$$
    - $$p\left(x_j=v_i \mid Y=\omega_i\right)=\frac{\sum_{i=1}^{N} I(x_j^{(i)}=v_i, y_i=\omega_i)+1}{\sum_{i=1}^{N} I(y_i=\omega_i)+S_j}$$