第 9 章 序列标注 – 隐马尔科夫模型

1/2

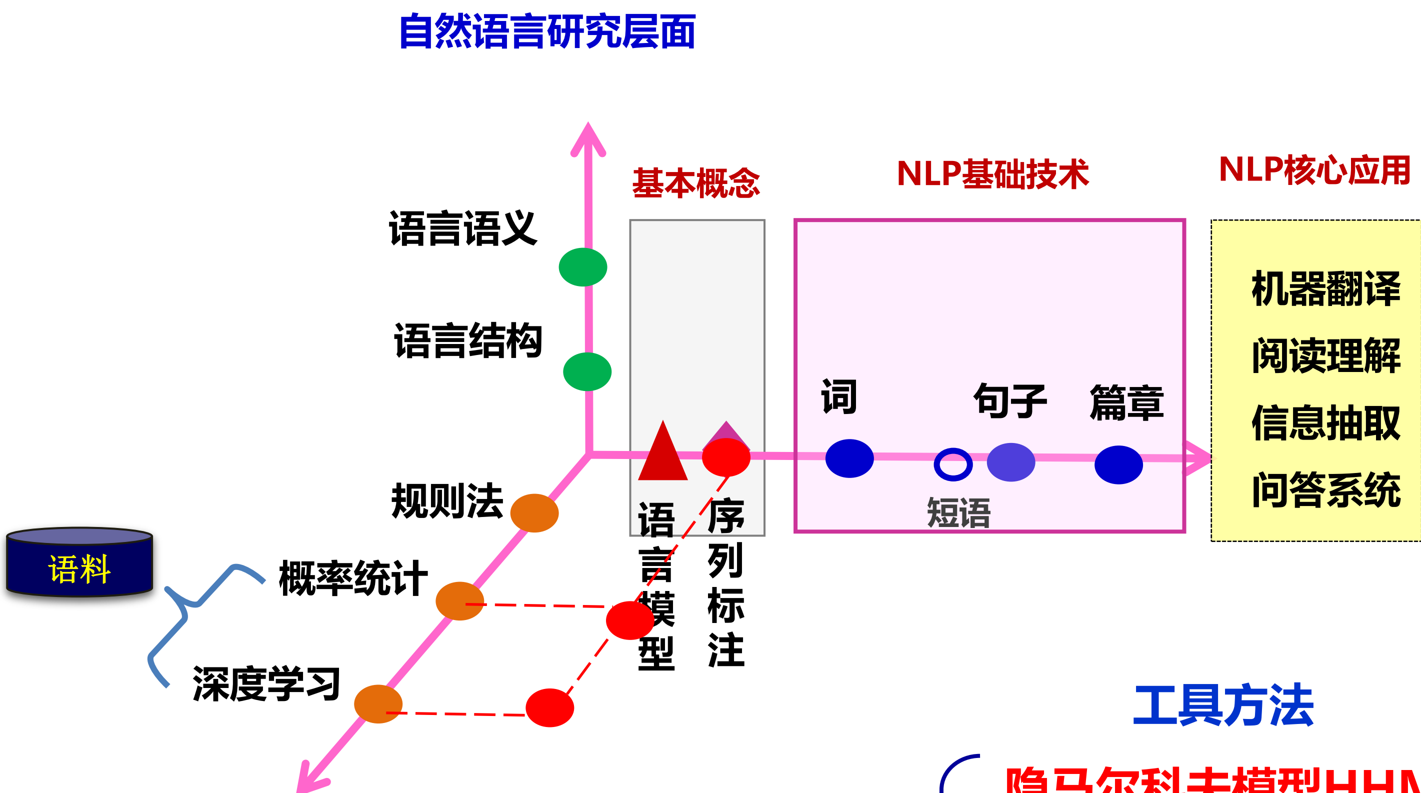
中科院信息工程研究所第二研究室

胡玥

huyue@iie.ac.cn

自然语言处理课程内容及安排

✧ 课程内容：



语言处理方法

工具方法

序列标注问题

问题

隐马尔科夫模型HHM
条件随机场 CRF
神经网络 RNN+ CRF

概率图模型

神经网络

内 容 提 要

9.1 序列标注问题

- 9.2 隐马尔科夫模型HHM
- 9.3 条件随机场 CRF
- 9.4 神经网络 RNN+ CRF

9.1 序列标注问题

问题引入

在自然语言处理任务中，有许多的任务可以转化为“将输入的语言序列转化为标注序列”来解决问题。如，命名实体识别、信息抽取、词性标注.....

- **命名实体识别（人名识别）**

目标：将给定的输入序列中的人名识别出来

如：输入序列： 新 任 总 裁 罗 建 国 宣 布 了 对 部 门 经 理 邓 奇 的 任 免 通 知

输出序列： 0 0 0 0 B I E 0 0 0 0 0 0 0 0 B E 0 0 0 0 0

{ B I E O } 或 { B I O }

B － 词首
I － 词中
E － 词尾
0 － 单个词

9.1 序列标注问题

序列标注问题：

$$f: X \rightarrow Y$$

序列　　序列

输入：	x:	x_1	x_2	x_3	……	x_L	序列
输出：	y:	y_1	y_2	y_3	……	y_L	序列

Y的符号集为类别

标注问题是分类问题的推广，是复杂结构预测的简单形式（监督学习）

许多自然语言处理问题　均可转化为序列标注问题

9.1 序列标注问题

- **命名实体识别（组织机构名识别）**

目标：将给定的输入序列中的组织机构名识别出来

输入序列：新任总裁罗建国宣布了对远大公司经理国庆的任免通知

标注序列：0 0 0 0 0 0 0 0 0 0 0 B I I E 0 0 0 0 0 0 0 0 0

- **信息抽取**

目标：将给定的输入序列中的军事术语抽取出来

输入： 鹰 式 战 斗 机 是 一 款 极 为 优 秀 的 多 用 途 战 斗 机

输出： B I I I E O O O O O O O O B I I I I E

9.1 序列标注问题

- **词性序列标注（POS）**

目标：将给定的输入序列中词的词性标出来

如：　　输入：　Flies　like　a　flower

　　　　输出：　N　V　ART　N

词性标注结果：　Flies/N　like /V　a/ART　flower/N

f：X → Y

Y的符号集　{ 单词的词性，如 N 、V 等 }

- **分词、短语识别、依存分析、语义角色标注……**

内 容 提 要

9.1 序列标注问题

9.2 隐马尔科夫模型HHM

9.2.1 马尔科夫模型

9.2.2 隐马尔科夫模型（结构）

9.2.3 隐马尔科夫模型学习

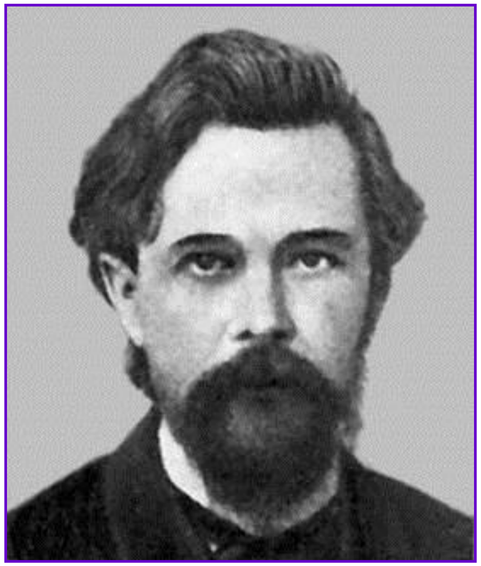
9.2.4 隐马模型应用

9.3 条件随机场 CRF

9.4 神经网络 RNN+ CRF

9.2.1 马尔科夫模型

马尔可夫（**Andrei Andreyevich Markov, 1856.6.14 ～ 1922.7.20** ）

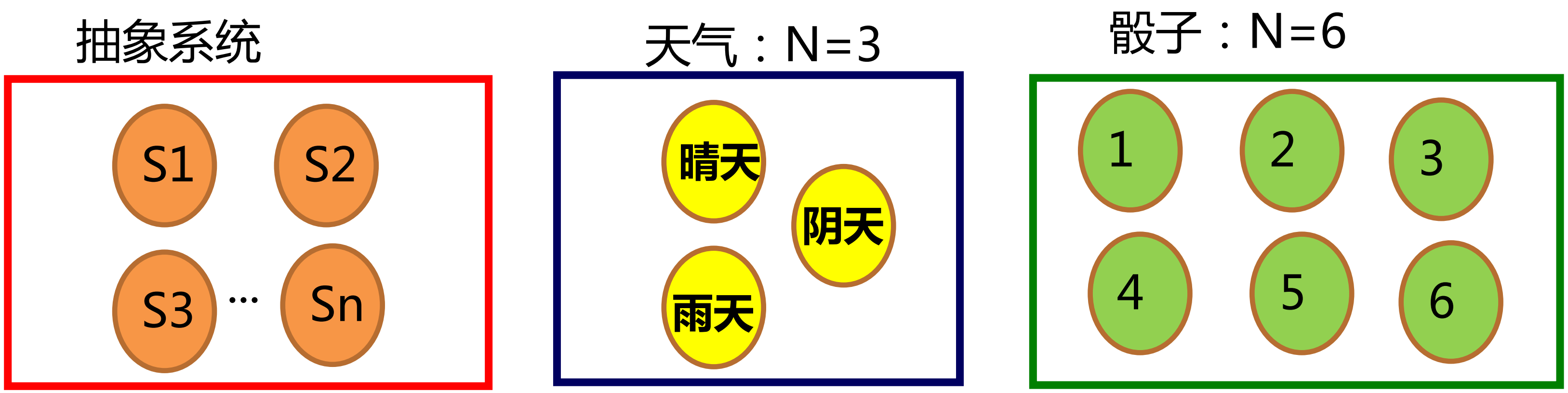


前苏联数学家。切比雪夫(1821年5月16日～ 1894年12月8日)的学生。在概率论、数论、函数逼近论和微分方程等方面卓有成就。他提出了用数学分析方法研究自然过程的一般图式—马尔可夫链，并**开创了随机过程(马尔可夫过程)**的研究。

9.2.1 马尔科夫模型

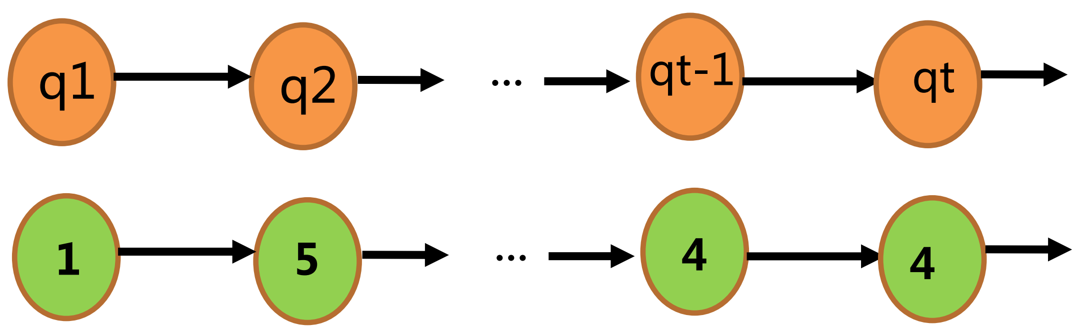
马尔可夫链

一个系统有N个状态 S1，S2，…，Sn，



随着时间（空间）推移，系统从某一状态转移到另一状态，**注意状态序列和状态集的区别**

如:骰子



如果系统在t时间的状态q t 只与其在时间 t -1的状态相关，则系统构成离散的一阶**马尔可夫链(马尔可夫过程)**

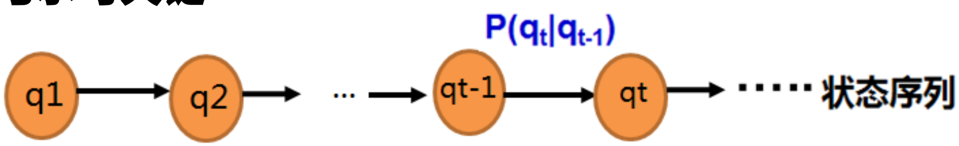
马尔可夫假设

9.2.1 马尔科夫模型

马尔可夫模型：

$$\boxed{p(S_0, S_1, \ldots, S_T) = \prod_{t=1}^{T} p(S_t|S_{t-1})\; p(S_0)}$$

马尔可夫链



独立于时间 t 的随机过程：

$$P(q_t = S_j \mid q_{t\text{-}1} = S_i) = a_{i,j}, 1 \le i, j \le N$$

其中：状态转移概率 a_{ij} 必须满足 $a_{ij}>=0$，

且 $\sum_{j=1}^{N} a_{i,j} = 1$

模型输入：状态序列

模型输出：状态序列的概率值

模型参数：$P(q_t|q_{t-1})$)

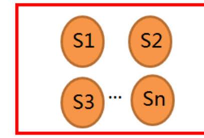
9.2.1 马尔科夫模型

■ 马尔可夫模型组成

三元组　M =（ S, π ，A ）

参数	含义
S	模型中状态的有限集合
A	与时间无关的状态转移概率矩阵
π	初始状态空间的概率分布

S={ s1 s2 s3 ….. sn }



$$
\mathbf{A}=[aij]=\begin{array}{c|ccccc|}
 & s1 & s2 & s3 & \cdots & sn \\
s1 & a11 & a12 & a13 & \cdots & a1n \\
s2 & a21 & a22 & a23 & \cdots & a2n \\
s3 & & & & & \\
\vdots & \vdots & & & & \\
sn & an1 & an2 & an3 & \cdots & ann
\end{array}
$$

其中： 状态转移概率 a_{ij}

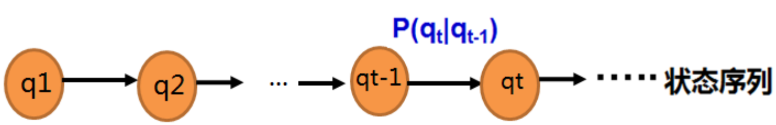
$$P(q_t = S_j \mid q_{t-1} = S_i) = a_{i,j}, 1 \le i, j \le N$$

满足 $a_{ij}>=0$，且 $\sum_{j=1}^{N} a_{i,j} = 1$

π　初始状态向量

■ 马尔可夫模型作用:

定量描述随机事件的变化过程



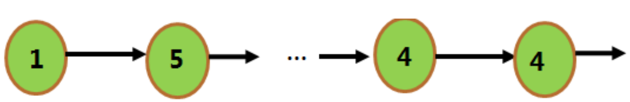
9.2.1 马尔科夫模型

马尔可夫模型作用

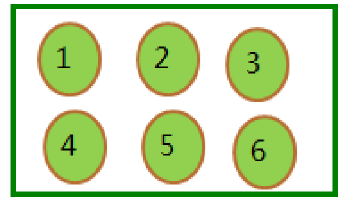
三元组　M = (S, π , A)

参数	含义
S	模型中状态的有限集合
A	与时间无关的状态转移概率矩阵
π	初始状态空间的概率分布

马尔科夫模型定量描述随机事件：



例1 ：掷骰子



S={ 1 2 3 4 5 6 }

$$
\mathbf{A} \quad \begin{array}{c|c}
 & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} &
\begin{bmatrix}
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6
\end{bmatrix}
\end{array}
$$

其中：　状态转移概率 a_{ij}

$$P(q_t = S_j \mid q_{t-1} = S_i) = a_{i,j}, 1 \le i, j \le N$$

满足 $a_{ij}>=0$，且 $\sum_{j=1}^{N} a_{i,j} = 1$

π =(1, 0, 0, 0, 0, 0)

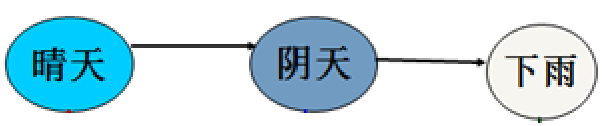
9.2.1 马尔科夫模型

马尔可夫模型作用

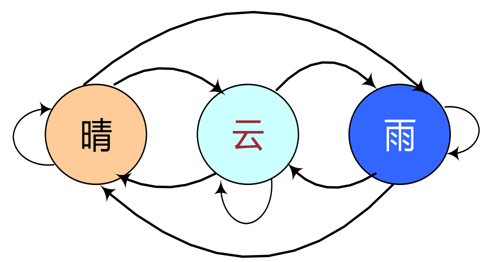
三元组 M = (S, π , A)

参数	含义
S	模型中状态的有限集合
A	与时间无关的状态转移概率矩阵
π	初始状态空间的概率分布

定量描述随机事件：天气变化



例2：预测天气变化



S={ 晴 云 雨 }

A

$$\begin{array}{c} \\ 晴天 \\ 阴天 \\ 下雨 \end{array} \begin{array}{c} \begin{array}{ccc} 晴天 & 阴天 & 下雨 \end{array} \\ \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.375 & 0.25 & 0.375 \\ 0.25 & 0.125 & 0.625 \end{pmatrix} \end{array}$$

其中： 状态转移概率 a_{ij}

$$P(q_t = S_j \mid q_{t-1} = S_i) = a_{i,j}, 1 \le i, j \le N$$

满足 $a_{ij}>=0$， 且 $\sum_{j=1}^{N} a_{i,j} = 1$

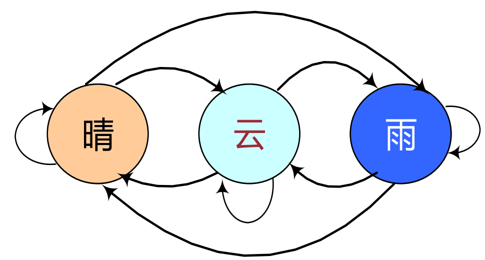
π =(1, 0, 0)

9.2.1 马尔科夫模型

例3： 假定一段时间内的气象可由一3状态马尔可夫模型 M 描述：

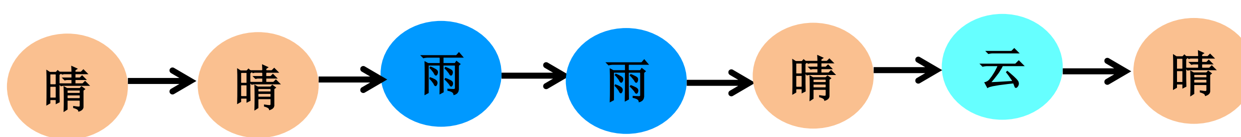
S_1 ：雨，S_2 ：多云，S_3 ：晴，转移概率矩阵为：

$$A=[a_{ij}]=\begin{array}{c}\\ S_1\\ S_2\\ S_3\end{array}\begin{array}{c}\begin{array}{ccc}S_1 & S_2 & S_3\end{array}\\ \left|\begin{array}{ccc}0.4 & 0.3 & 0.3\\ 0.2 & 0.6 & 0.2\\ 0.1 & 0.1 & 0.8\end{array}\right|\end{array}$$



如果第一天为晴天，根据这一模型，求在今后七天中天气为 S= “晴晴雨雨晴云晴”的概率

即，求



的概率

9.2.1 马尔科夫模型

解：　　马尔可夫模型状态序列概率：

$$\boxed{p(S_0, S_1, \ldots, S_T) = \prod_{t=1}^{T} p(S_t|S_{t-1})\ p(S_0)}$$

S = 晴 晴 雨 雨 晴 云 晴

$$
\begin{aligned}
&P(O \mid M) \\
&= P(S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 \mid M) \\
&= P(S_3) \cdot P(S_3 \mid S_3) \cdot P(S_3 \mid S_3) \cdot P(S_1 \mid S_3) \cdot P(S_1 \mid S_1) \cdot \\
&P(S_3 \mid S_1) \cdot P(S_2 \mid S_3) \cdot P(S_3 \mid S_2) \\
&= 1 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\
&= (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\
&= 1.536 \times 10^{-4}
\end{aligned}
$$

$$
A = [a_{ij}] = \begin{array}{c|ccc|}
 & 雨 & 云 & 晴 \\
雨 & 0.4 & 0.3 & 0.3 \\
云 & 0.2 & 0.6 & 0.2 \\
晴 & 0.1 & 0.1 & 0.8
\end{array}
$$

内 容 提 要

9.1 序列标注问题

- 9.2 隐马尔科夫模型HHM
  - 9.2.1 马尔科夫模型
  - 9.2.2 隐马尔科夫模型（结构）
  - 9.2.3 隐马尔科夫模型学习
  - 9.2.4 隐马模型应用
- 9.3 条件随机场 CRF
- 9.4 神经网络 RNN+ CRF

9.2.2 隐马尔科夫模型（结构）

隐马尔可夫模型 （Hidden Markov Model, HMM）

--- 创建于20世纪70年代 ---

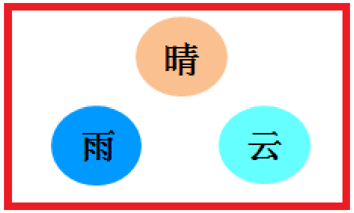
描写：该模型是一个双重随机过程，我们不知道具体的状态序列，只知道状态转移的概率，即模型的状态转换过程是不可观察的（隐蔽的），而可观察事件的随机过程是隐蔽状态转换过程的随机函数。

通过可见的事物的变化揭示深藏其后的内在的本质规律

9.2.2 隐马尔科夫模型（结构）

马尔可夫模型：

S:



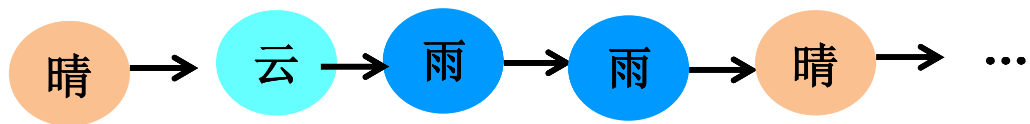
A:

$$A = [a_{ij}] = \begin{array}{c} \\ 雨 \\ 云 \\ 晴 \end{array}\begin{array}{c} \begin{array}{ccc} 雨 & 云 & 晴 \end{array} \\ \left|\begin{array}{ccc} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{array}\right| \end{array}$$

π：晴 云 雨

（1，0，0 ）

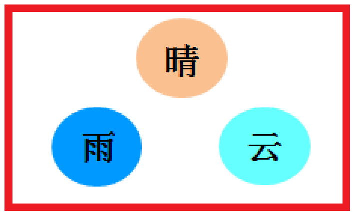
天气变化



9.2.2 隐马尔科夫模型（结构）

马尔可夫模型： → 隐马尔可夫模型 HMM

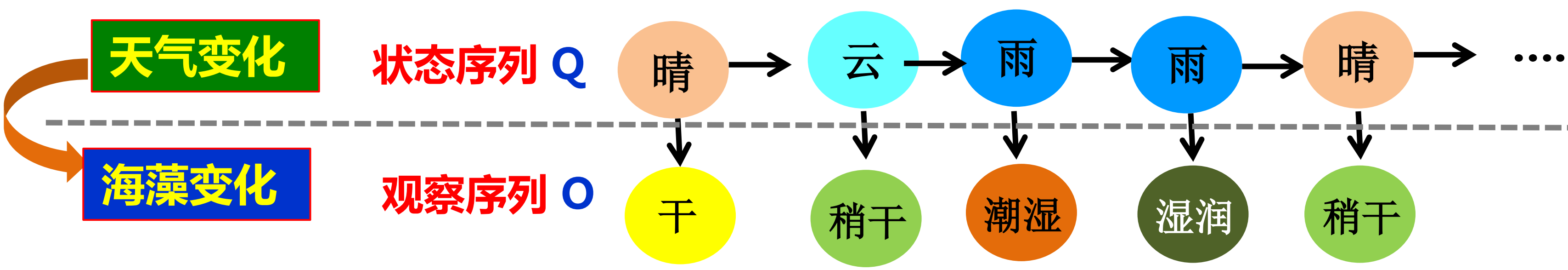
S:



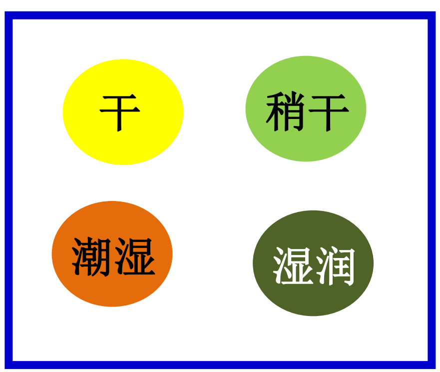
A:

A = [aij]=	雨	云	晴
雨	0.4	0.3	0.3
云	0.2	0.6	0.2
晴	0.1	0.1	0.8

π ： 晴 云 雨
（1，0，0 ）



O:



B:

天气	海藻：干	稍干	潮湿	湿润
晴天	0.60	0.20	0.15	0.05
阴天	0.25	0.25	0.25	0.25
下雨	0.05	0.10	0.35	0.50

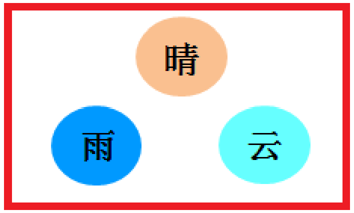
观察序列变化由状态序列变化引起

（两者相关联）

9.2.2 隐马尔科夫模型（结构）

隐马尔可夫模型 HMM作用

S:



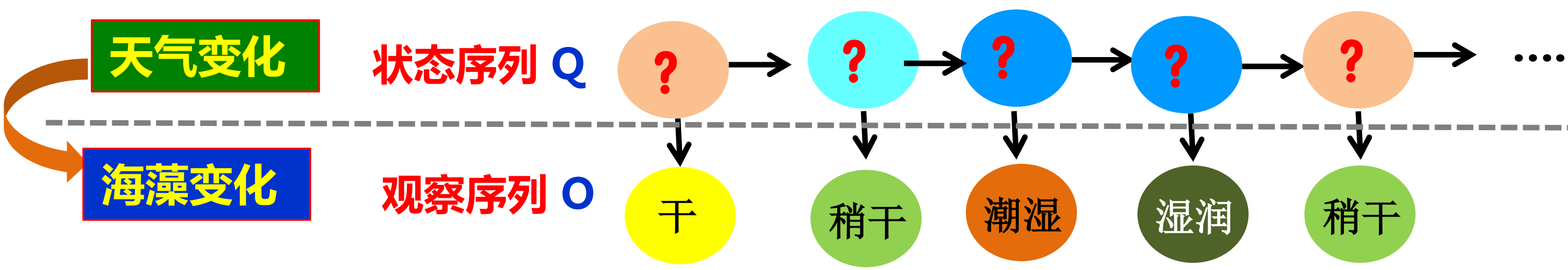
A:

$A=[a_{ij}]=$

	雨	云	晴
雨	0.4	0.3	0.3
云	0.2	0.6	0.2
晴	0.1	0.1	0.8

π： 晴 云 雨

（1，0，0 ）



O:

B:

		海藻			
		干	稍干	潮湿	湿润
天气	晴天	0.60	0.20	0.15	0.05
	阴天	0.25	0.25	0.25	0.25
	下雨	0.05	0.10	0.35	0.50

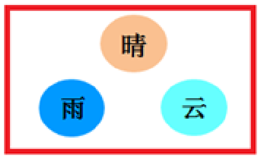
当状态序列不可见时

可以通过观察序列推出状态序列

9.2.2 隐马尔科夫模型（结构）

隐马尔可夫模型(HMM)组成：

S:

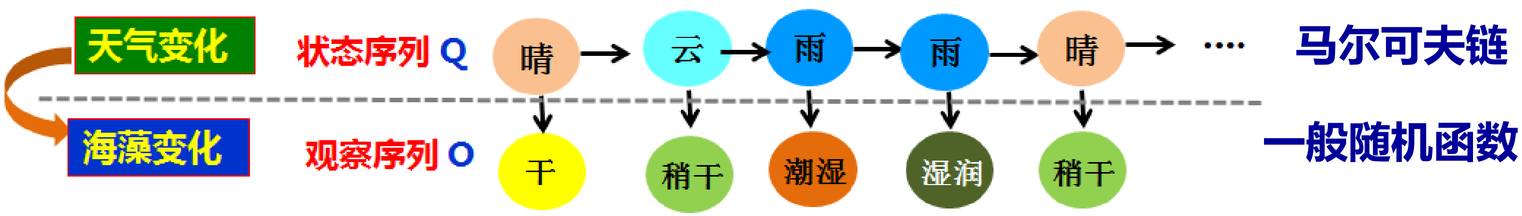


A:

$$A = [a_{ij}] = \begin{array}{c|ccc} & 雨 & 云 & 晴 \\ \hline 雨 & 0.4 & 0.3 & 0.3 \\ 云 & 0.2 & 0.6 & 0.2 \\ 晴 & 0.1 & 0.1 & 0.8 \end{array}$$

π： 晴 云 雨

(1, 0, 0)



O:

B:

$$\begin{array}{cc} & \text{海藻} \\ & \begin{array}{cccc} 干 & 稍干 & 潮湿 & 湿润 \end{array} \\ \text{天气} \begin{array}{c} 晴天 \\ 阴天 \\ 下雨 \end{array} & \begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{pmatrix} \end{array}$$

五元组 λ = (S, O, π , A , B)

或简写为 λ = (π , A , B)

要素	含义	实例
S	模型中状态的有限集合	天气
O	每个状态可能的观察值	海藻
A	与时间无关的状态转移概率矩阵	天气转移概率矩阵
B	给定状态下，观察值概率分布	每个天气状态的海藻观测概率
π	初始状态空间的概率分布	初始时选择某天气概率

9.2.2 隐马尔科夫模型（结构）

HMM五元组 说明 ：

1. **隐藏状态s**：一个系统的（真实）状态，可以由一个马尔 科夫过程进行描述（如，天气）
2. **观察状态 o**：在这个过程中‘可视’的状态（例如，海藻的湿度）

3. **状态转移概率矩阵 A = a_{ij}** ：包含了一个隐藏状态到另一个隐藏状态的概率 。其中,

$$
\begin{cases}
a_{ij} = p(q_{t+1} = S_j \mid q_t = S_i), \quad 1 \le i, j \le N \\
a_{ij} \ge 0 \\
\sum_{j=1}^{N} a_{ij} = 1
\end{cases}
$$

9.2.2 隐马尔科夫模型（结构）

4. 观察概率矩阵 B=b$_j$(k)：从隐藏状态 S_j观察到某一特定符号 v_k的概率分布矩阵。 其中，

$$
\begin{cases}
b_j(k) = p(O_t = v_k \mid q_t = S_j), \quad 1 \le j \le N, \quad 1 \le k \le M \\
b_j(k) \ge 0 \\
\sum_{k=1}^{M} b_j(k) = 1
\end{cases}
$$

5. 初始状态的概率分布为：$\pi = \pi_i$，其中，

$$
\begin{cases}
\pi_i = p(q_1 = S_i), \quad 1 \le i \le N \\
\pi_i \ge 0 \\
\sum_{i=1}^{N} \pi_i = 1
\end{cases}
$$

9.2.2 隐马尔科夫模型（结构）

HMM的三个假设：

对于一个随机事件，有一观察值序列： $O=O_1,O_2,\ldots O_T$

该事件隐含着一个状态序列： $Q = q_1,q_2,\ldots q_T$。

假设1：马尔可夫性假设（状态构成一阶马尔可夫链）

$$P(q_i|q_{i-1}\ldots q_1) = P(q_i|q_{i-1})$$

假设2：不动性假设（状态与具体时间无关）

$$P(q_{i+1}|q_i) = P(q_{j+1}|q_j)$$，对任意i，j成立

假设3：输出独立性假设（输出仅与当前状态有关）

$$p(O_1,\ldots,O_T \mid q_1,\ldots,q_T) = \Pi p(O_t \mid q_t)$$

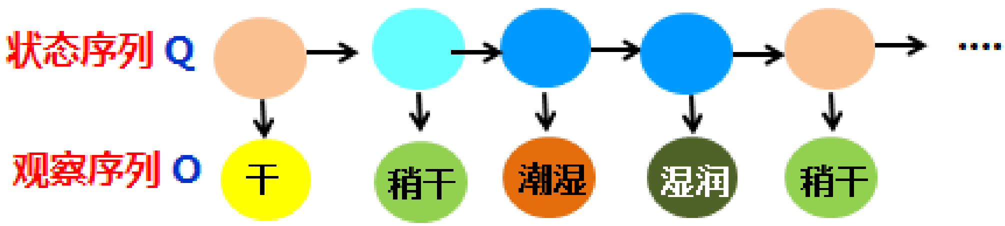
9.2.2 隐马尔科夫模型（结构）

HMM的特点：

- HMM的状态是不确定或不可见的，只有通过观测序列的随机过程才能表现出来
- 观察到的事件与状态并不是一一对应，而是通过一组概率分布相联系
- HMM是一个双重随机过程，两个组成部分：
  - 马尔可夫链：描述状态的转移，用转移概率描述。
  - 一般随机函数：描述状态与观察序列间的关系， 用观察值概率描述。

9.2.2 隐马尔科夫模型（结构）

隐马尔可夫模型(HMM)：



要素	含义	实例
S	模型中状态的有限集合	天气
O	每个状态可能的观察值	海藻
A	与时间无关的状态转移概率矩阵	天气转移概率矩阵
B	给定状态下，观察值概率分布	每个天气状态的海藻观测概率
π	初始状态空间的概率分布	初始时选择某天气概率

输入： **观察序列**

输出： **1. 观察序列的概率值 2. 隐状态序列**

参数： $P(q_t|q_{t-1})$，$P(O_t|q_t)$

A矩阵　　B矩阵

用隐马尔可夫模型可求：

Ⅰ. 在给定模型中出现观察序列的可能性（**概率值**） **HMM评估问题**

Ⅱ. 通过观察序列找出最大可能的**隐状态序列** **HMM解码问题**

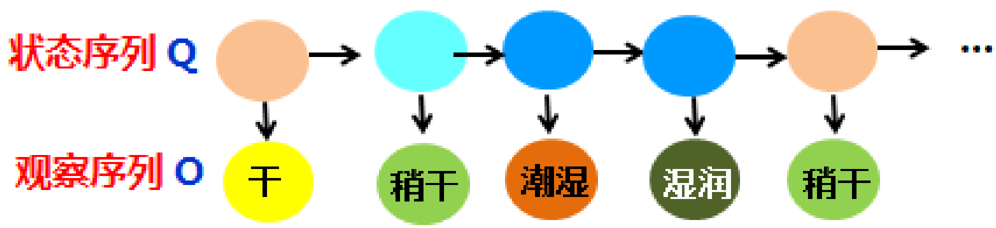
9.2.2 隐马尔科夫模型（结构）—Ⅰ. HMM评估问题

HMM评估问题：

对于给定观察序列 $O=O_1,O_2,\ldots O_T$，以及模型 $\lambda=(A,B,\pi)$

求观察序列的概率P(O|λ)

如：



求：观察序列概率P（O | λ）= ?

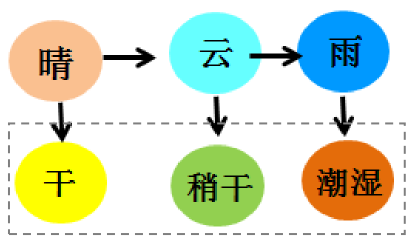
问题：

1. 观察序列概率P(O|λ)定义
2. 如何计算 P(O|λ)

9.2.2 隐马尔科夫模型（结构）— I．HMM评估问题

1. 观察序列概率P(O|λ)定义

- 对于给定的一个状态序列 $Q = q_1q_2\ldots q_T$，



观察序列

已知

S = {晴天，阴天 ，下雨 }

O = {湿润， 潮湿， 稍干，干燥 }

$$A \quad \begin{array}{c|ccc} & 晴天 & 阴天 & 下雨 \\ \hline 晴天 & 0.50 & 0.25 & 0.25 \\ 阴天 & 0.375 & 0.25 & 0.375 \\ 下雨 & 0.25 & 0.125 & 0.625 \end{array}$$

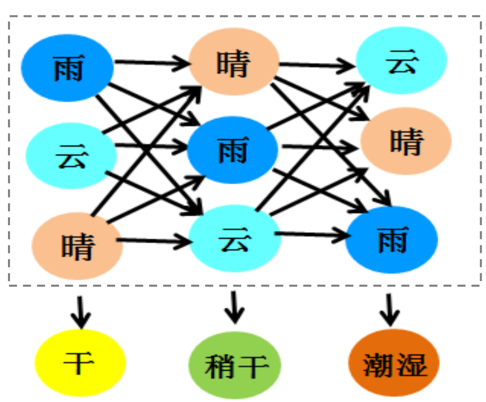
$$B \quad 天气 \begin{array}{c|cccc} & 干 & 稍干 & 潮湿 & 湿润 \\ \hline 晴天 & 0.60 & 0.20 & 0.15 & 0.05 \\ 阴天 & 0.25 & 0.25 & 0.25 & 0.25 \\ 下雨 & 0.05 & 0.10 & 0.35 & 0.50 \end{array}$$

$\pi = (1,\ 0,\ 0)$

$$P(O,Q\,|\,\lambda)=\underbrace{\pi_{q1}a_{q1q2}a_{q2q3}\cdots a_{qT\text{-}1qT}}_{P(Q\,|\,\lambda)}\ \underbrace{b_{q1}(O_1)\ b_{q2}(O_2)\ ,\ldots b_{qT}(O_T)}_{P(O\,|Q\,,\,\lambda)}$$

$$P(O,Q\,|\,\lambda) = P(Q\,|\,\lambda)\ P(O\,|Q\,,\,\lambda)$$

- 对于全部状态序列



观察序列来源

观察序列概率：

$$P(O\,|\,\lambda) = \sum_{Q} P(O,Q\,|\,\lambda) = \sum_{Q} P(Q\,|\,\lambda)P(O\,|\,Q,\lambda)$$

9.2.2 隐马尔科夫模型（结构）—Ⅰ. HMM评估问题

2. 如何计算 P(O|λ)

已知

S = {晴天，阴天 ，下雨 }

O = {湿润， 潮湿， 稍干，干燥 }

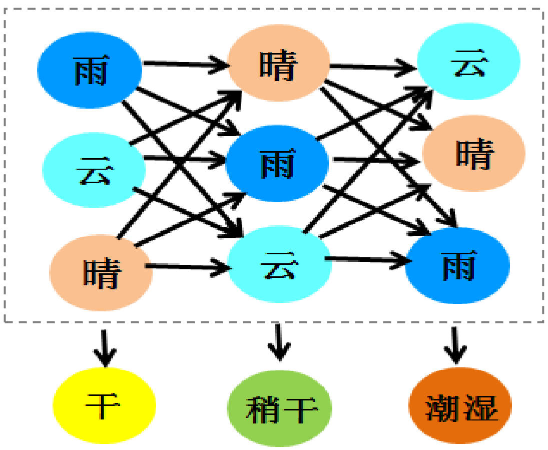
A
$$\begin{array}{c} \\ 晴天 \\ 阴天 \\ 下雨 \end{array} \begin{array}{c} \begin{array}{ccc} 晴天 & 阴天 & 下雨 \end{array} \\ \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.375 & 0.25 & 0.375 \\ 0.25 & 0.125 & 0.625 \end{pmatrix} \end{array}$$

B
$$天气 \begin{array}{c} \\ 晴天 \\ 阴天 \\ 下雨 \end{array} \begin{array}{c} \begin{array}{cccc} 干 & 稍干 & 潮湿 & 湿润 \end{array} \\ \begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{pmatrix} \end{array}$$

$$\pi = (1,\ 0,\ 0)$$

观察序列概率：

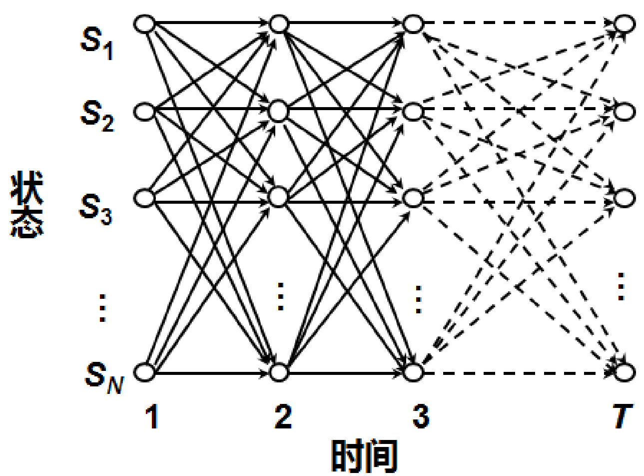
$$P(O \mid \lambda) = \sum_Q P(O, Q \mid \lambda) = \sum_Q P(Q \mid \lambda) P(O \mid Q, \lambda)$$



（1）穷举法： 找到每一个可能的隐藏状态的序列，这里有3^3= 27种，可观察序列的概率就是这27种可能的和。 很显然，这种计算的效率非常低，尤其是当模型中的状态非常多或者序列很长的时候。

9.2.2 隐马尔科夫模型（结构）—Ⅰ. HMM评估问题

穷举法的问题：



◆困难：

如果模型μ有 N个不同的状态，时间长度为T，那么有 N^T个可能的状态序列，搜索路径成指数级组合爆炸。

解决方法：

- **（2）前向算法（后向算法）**：利用动态规划使用递归来降低计算复杂度

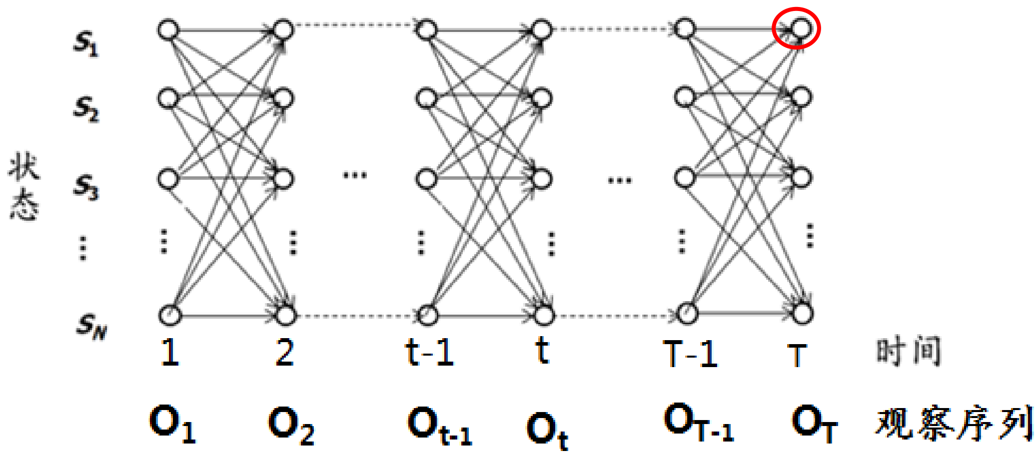
向前算法

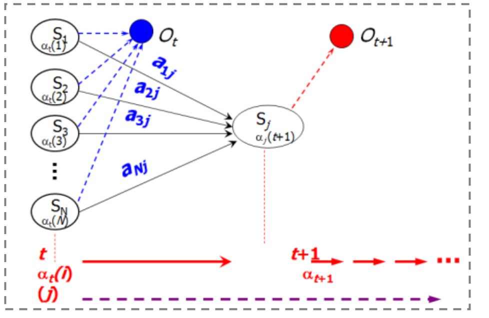
9.2.2 隐马尔科夫模型（结构）— Ⅰ. HMM评估问题

前向算法基本思想：

使用递归来降低计算复杂度

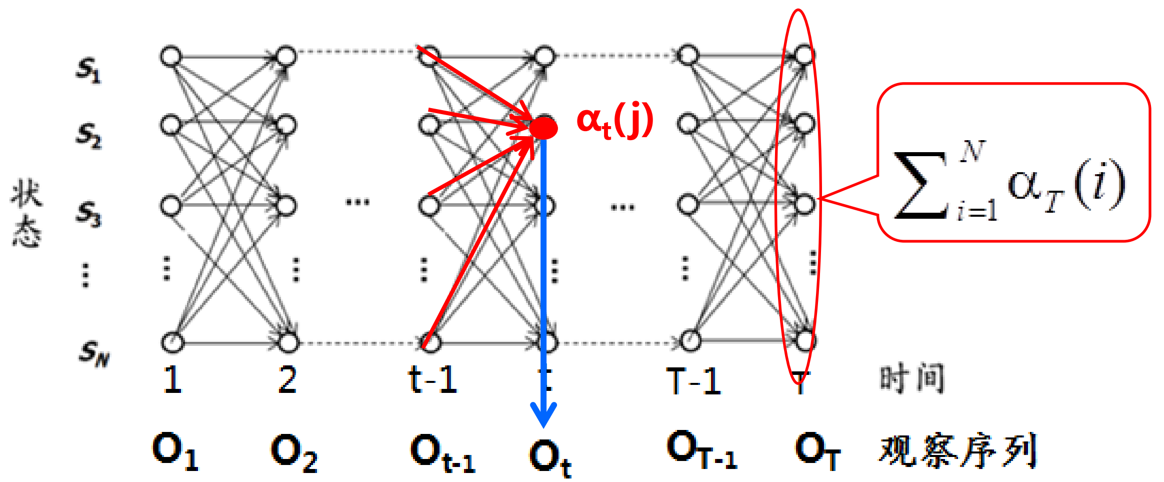
$$P(O\mid\lambda)=\sum_{Q}P(O,Q\mid\lambda)=\sum_{Q}P(Q\mid\lambda)P(O\mid Q,\lambda)$$





9.2.2 隐马尔科夫模型（结构）— I．HMM评估问题

前向算法实现：



定义 前向变量 $\alpha_t(j)$（部分概率），表示达到某个中间状态的概率

- **当 t=1 时**，是初始概率，Pr(状态 j | t=1) = π(状态 j)

$$\alpha_1(i) = \pi_i b_i(O_1), \quad t = 1$$

- **当1 ≤ t ≤ T-1 时**，

$$\alpha_t(j) = [\sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij}] b_j(O_t), \quad 1 \le t \le T-1$$

- **最终结果** $p(O|\mu) = \sum_{i=1}^{N} \alpha_T(i)$

9.2.2 隐马尔科夫模型（结构）— I. HMM评估问题

前向算法 **(The forward procedure)**

(1) 初始化： $\alpha_1(i) = \pi_i b_i(O_1)$， $1 \le i \le N$

(2) 循环计算：

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(O_{t+1}), \qquad 1 \le t \le T-1$$

(3) 结束，输出：

$$p(O \mid \mu) = \sum_{i=1}^{N} \alpha_T(i)$$

前向计算过程见附录

9.2.2 隐马尔科夫模型（结构）—Ⅰ. HMM评估问题

算法的时间复杂性：

每计算一个 $\alpha_t(i)$ 必须考虑从 t-1 时的所有 N 个状态转移到状态 S_i 的可能性，时间复杂性为 O(N)，对应每个时刻 t，要计算 N 个前向变量：$\alpha_t(1), \alpha_t(2), \ldots, \alpha_t(N)$，所以，时间复杂性为：$O(N)\times N = O(N^2)$。 又因 $t = 1, 2, \ldots, T$，所以前向算法总的复杂性为：**$O(N^2T)$**

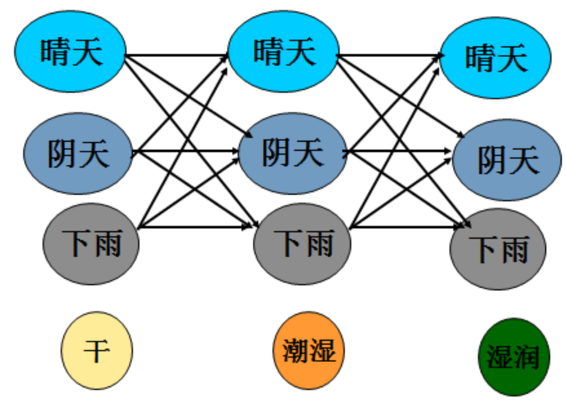
穷举算法的时间开销是和 T 指数相关　即 ***O*(N^T)**

9.2.2 隐马尔科夫模型（结构）— I. HMM评估问题

例1： 已有天气和海藻关系的HMM模型 λ ；求连续3 天海藻湿度的观察结果是（干燥、潮湿、湿润）的概率。

S = {晴天，阴天 ，下雨 }

O = {湿润， 潮湿， 稍干，干燥 }



A

$$
\begin{array}{c|ccc}
 & 晴天 & 阴天 & 下雨 \\
\hline
晴天 & 0.50 & 0.25 & 0.25 \\
阴天 & 0.375 & 0.25 & 0.375 \\
下雨 & 0.25 & 0.125 & 0.625
\end{array}
$$

B

$$
\begin{array}{cc|cccc}
 & & \multicolumn{4}{c}{海藻} \\
 & & 干 & 稍干 & 潮湿 & 湿润 \\
\hline
 & 晴天 & 0.60 & 0.20 & 0.15 & 0.05 \\
天气 & 阴天 & 0.25 & 0.25 & 0.25 & 0.25 \\
 & 下雨 & 0.05 & 0.10 & 0.35 & 0.50
\end{array}
$$

π $\pi =（1，0，0）$

9.2.2 隐马尔科夫模型（结构）—Ⅰ. HMM评估问题

解：用向前算法

$$\begin{array}{c} \\ 晴天 \\ 阴天 \\ 下雨 \end{array} \begin{array}{c} \begin{array}{ccc} 晴天 & 阴天 & 下雨 \end{array} \\ \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.375 & 0.25 & 0.375 \\ 0.25 & 0.125 & 0.625 \end{pmatrix} \end{array}$$

海藻

$$天气 \begin{array}{c} \\ 晴天 \\ 阴天 \\ 下雨 \end{array} \begin{array}{c} \begin{array}{cccc} 干 & 稍干 & 潮湿 & 湿润 \end{array} \\ \begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{pmatrix} \end{array}$$

$\pi =（1，0，0）$

1. 前向算法

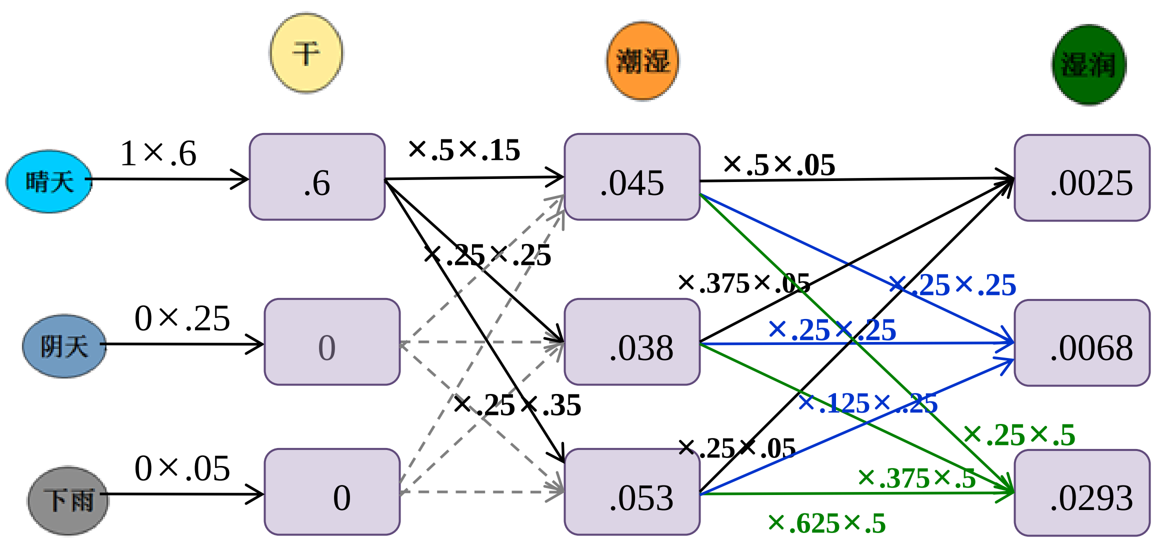
(1) 初始化： $\alpha_1(i)=\pi_i b_i(O_1),\quad 1\le i\le N$

(2) 循环计算：

$$\alpha_{t+1}(j)=[\sum_{i=1}^{N}\alpha_t(i)a_{ij}]b_j(O_{t+1}),\quad 1\le t\le T-1$$

(3) 结束，输出：

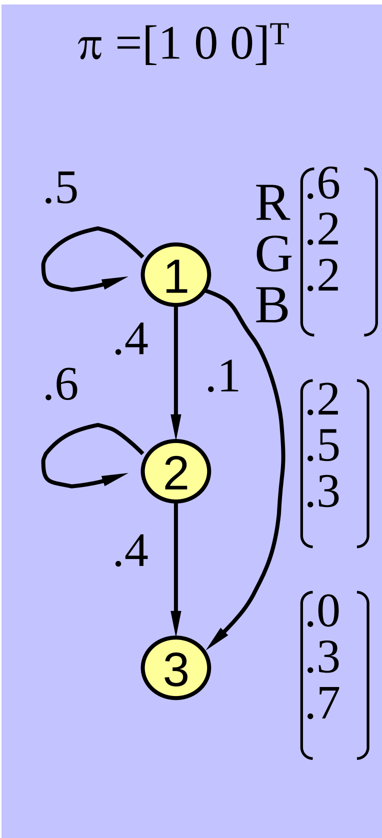
$$p(O\,|\,\mu)=\sum_{i=1}^{N}\alpha_T(i)$$

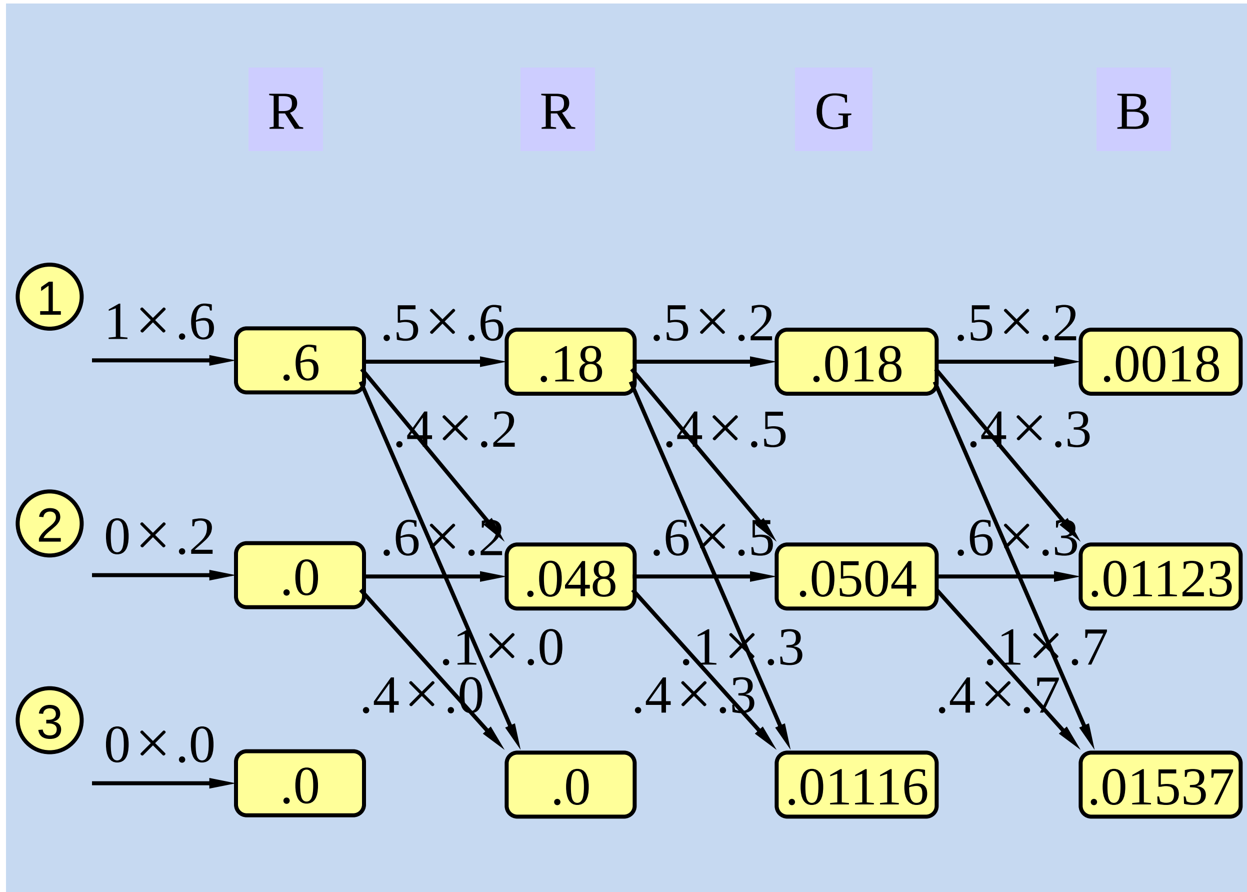


$$p(O\,|\,\mu)=\sum_{i=1}^{N}\alpha_T(i) = 0.0025 + 0.0068 + 0.0293 = 0.0386$$

9.2.2 隐马尔科夫模型（结构）— I. HMM评估问题

例 2：





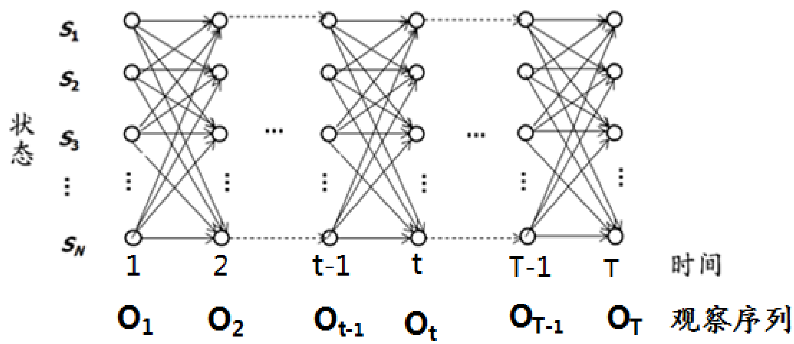
$$p(O\,|\,\mu) = \sum_{i=1}^{N} \alpha_T(i) \quad = .0018 + .01123 + .01537 = .0284$$

9.2.2 隐马尔科夫模型（结构）— Ⅰ. HMM评估问题

向后算法

9.2.2 隐马尔科夫模型（结构）— I. HMM评估问题

后向算法：与前向算法一样，运用动态规划计算**后向量**



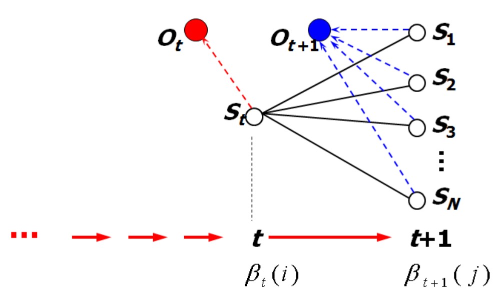
归纳顺序：$\beta_T(x), \beta_{T-1}(x), \cdots, \beta_1(x)$

- **当 t=T 时，** $\beta_T(i) = 1, \quad 1 \le i \le N$

定义**后向变量** $\beta_t(i)$

- **当T-1 ≥ t≥1 时，**

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad T-1 \ge t \ge 1, \quad 1 \le i \le N$$



- **结果：**

$$\boxed{P(O \mid \mu) = \sum_{i=1}^{N} \pi_i b_i(O_1) \beta_1(i)}$$

9.2.2 隐马尔科夫模型（结构）— Ⅰ. HMM评估问题

后向算法 **(The backward procedure)**

（1）初始化： $\beta_T(i)=1, \quad 1 \le i \le N$

（2）循环计算：

$$\beta_t(i)=\sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \qquad T-1 \ge t \ge 1, \quad 1 \le i \le N$$

（3）输出结果：

$$\mathrm{P}(\mathrm{O} \mid \mu)=\sum_{i=1}^{N} \pi_i b_i(O_1) \beta_1(i)$$

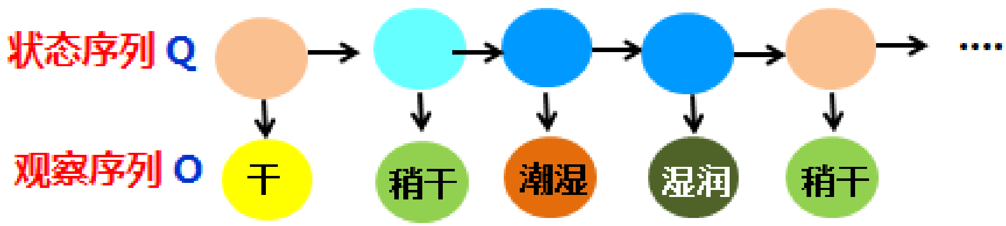
算法的时间复杂性： $O(N^2T)$

9.2.2 隐马尔科夫模型（结构）—Ⅱ. HMM解码问题

HMM解码问题：

对于给定观察序列 $O=O_1,O_2,\ldots O_T$，以及模型 $\lambda=(A,B,\pi)$ 如何选择一个对应的状态序列 $S = S_1,S_2,\ldots S_T$，使得S能够最为合理的解释观察序列 O

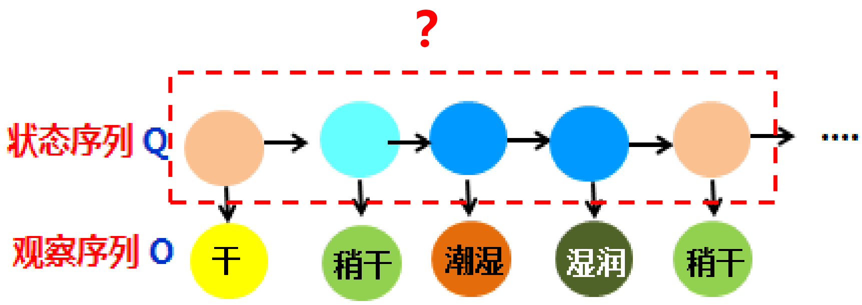
如：

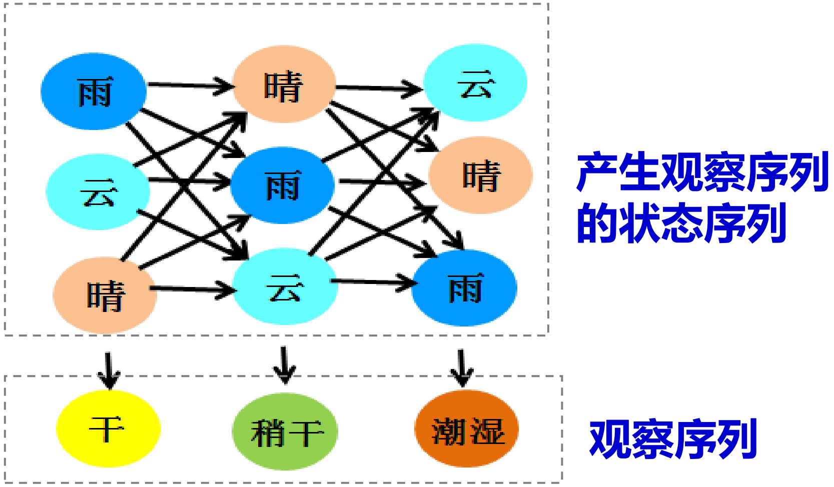


求：状态序列序列 $S = S_1,S_2,\ldots S_T$

9.2.2 隐马尔科夫模型（结构）—Ⅱ. HMM解码问题

求：状态序列序列 $S = S_1,S_2,\ldots S_T$





（1）穷举法：找到每一个可能产生观察序列的状态序列，这里有3^3= 27种，计算每种可能情况下观察序列的概率，概率最大的状态序列就是要找的状态序列。 很显然，这种计算的效率非常低，尤其是当模型中的状态非常多或者序列很长的时候。

解决方法：

（2）Viterbi 搜索算法：利用动态规划使用递归来降低计算复杂度

9.2.2 隐马尔科夫模型（结构）—Ⅱ. HMM解码问题

Viterbi 搜索算法

Andrew Viterbi

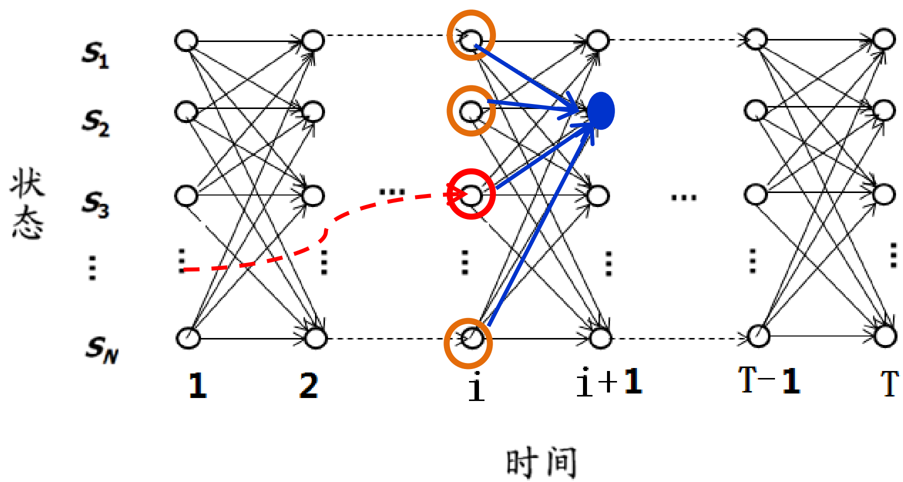
安德鲁.维特比（Andrew Viterbi ）

1967年发明了维特比算法。

维特比算法： 利用动态规划方法解决特殊的篱笆网络有向图的最短路径问题。 是现代数字通信中使用最频繁的算法；同时也是很多自然语言处理的解码算法。

9.2.2 隐马尔科夫模型（结构）—II. HMM解码问题

Viterbi 算法基本思想： 使用递归来降低复杂度

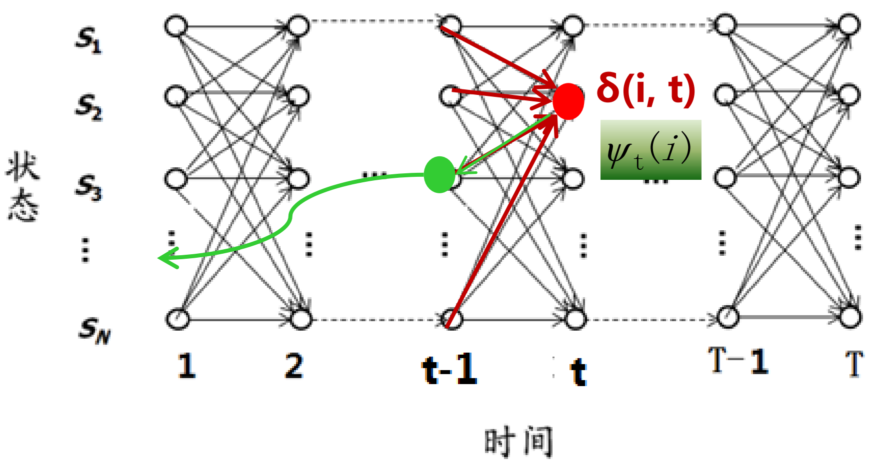


1. 如果概率最大路径（或说最短路径）经 i 时刻某个点，一定可以找到S到该点的最短路径（可将i时刻点的最短路径记录）
2. 从S到E 的路径必定经过 i时刻的某个点
3. 当从状态 i 进入到i+1状态时计算S到i+1 状态时，只考虑 i状态所有节点最短路径和和它们到 i+1状态的距离即可。

Viterbi时间复杂度： $O(N^2T)$ **穷举法** :$O(N^T)$

9.2.2 隐马尔科夫模型（结构）—II. HMM解码问题

Viterbi 算法实现：



(1) 部分最优路径概率

定义一个部分概率δ ;用 δ(i, t) 来表示在t时刻，到状态i的所有可能的序列（路径）中概率最大的序列的概率

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \le i \le N$$

$$\delta_t(j) = \max_{1\le i\le N}[\delta_{t-1}(i)\cdot a_{ij}]\cdot b_j(O_t), \quad 2\le t\le T, \quad 1\le j\le N$$

(2)向后指针记录最优路径

利用一个后向指针 φ 来记录导致某个状态最大局部概率的前一个状态

$$\phi_t(i) = \arg\max_j(\delta_{t-1}(j)a_{ji})$$

(3) 结果

$$\boxed{q_t^* = \varphi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \ldots, 1}$$

9.2.2 隐马尔科夫模型（结构）—II. HMM解码问题

Viterbi 搜索算法

1. 初始化： $\delta_1(i) = \pi_i b_i(O_1)$， $\varphi_1(i) = 0$， $1 \le i \le N$

2. 递归： $\delta_t(j) = [\max\limits_{1\le i\le N} \delta_{t-1}(i) a_{ij}] b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$

   $\varphi_t(j) = \arg\max\limits_{1\le i\le N} [\delta_{t-1}(i) a_{ij}] b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$

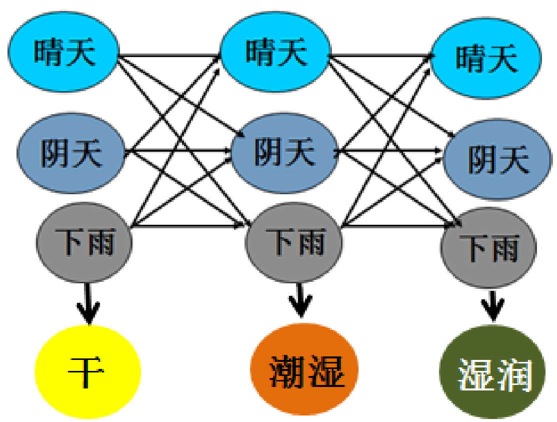
3. 终结： $p^* = \max\limits_{1\le i\le N} [\delta_T(i)])$ ， $q_T^* = \arg\max\limits_{1\le i\le N} [\delta_T(i)]$

4. 路径回溯： $q_t^* = \varphi_{t+1}(q_{t+1}^*)$， $t = T-1, T-2, \ldots, 1$

算法的时间复杂度： $O(N^2 T)$

9.2.2 隐马尔科夫模型（结构）—Ⅱ. HMM解码问题

例1： 已有天气和海藻关系的HMM模型 λ 和连续3 天海藻湿度的观察结果（干燥、潮湿、湿润），求最可能的天气序列。



A:

$$\begin{array}{c} \\ \text{晴天} \\ \text{阴天} \\ \text{下雨} \end{array} \begin{array}{c} \begin{array}{ccc} \text{晴天} & \text{阴天} & \text{下雨} \end{array} \\ \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.375 & 0.25 & 0.375 \\ 0.25 & 0.125 & 0.625 \end{pmatrix} \end{array}$$

B:

海藻

$$\text{天气} \begin{array}{c} \\ \text{晴天} \\ \text{阴天} \\ \text{下雨} \end{array} \begin{array}{c} \begin{array}{cccc} \text{干} & \text{稍干} & \text{潮湿} & \text{湿润} \end{array} \\ \begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{pmatrix} \end{array}$$

设： $\pi = (1, 0, 0)$

9.2.2 隐马尔科夫模型（结构）—II. HMM解码问题

解：用Viterbi 搜索算法

$$
\begin{array}{c} \\ 晴天 \\ 阴天 \\ 下雨 \end{array}
\begin{array}{c} \begin{array}{ccc} 晴天 & 阴天 & 下雨 \end{array} \\ \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.375 & 0.25 & 0.375 \\ 0.25 & 0.125 & 0.625 \end{pmatrix} \end{array}
$$

海澡

$$
天气\quad
\begin{array}{c} \\ 晴天 \\ 阴天 \\ 下雨 \end{array}
\begin{array}{c} \begin{array}{cccc} 干 & 稍干 & 潮湿 & 湿润 \end{array} \\ \begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{pmatrix} \end{array}
$$

$\pi =（1，0，0）$

Viterbi 搜索算法

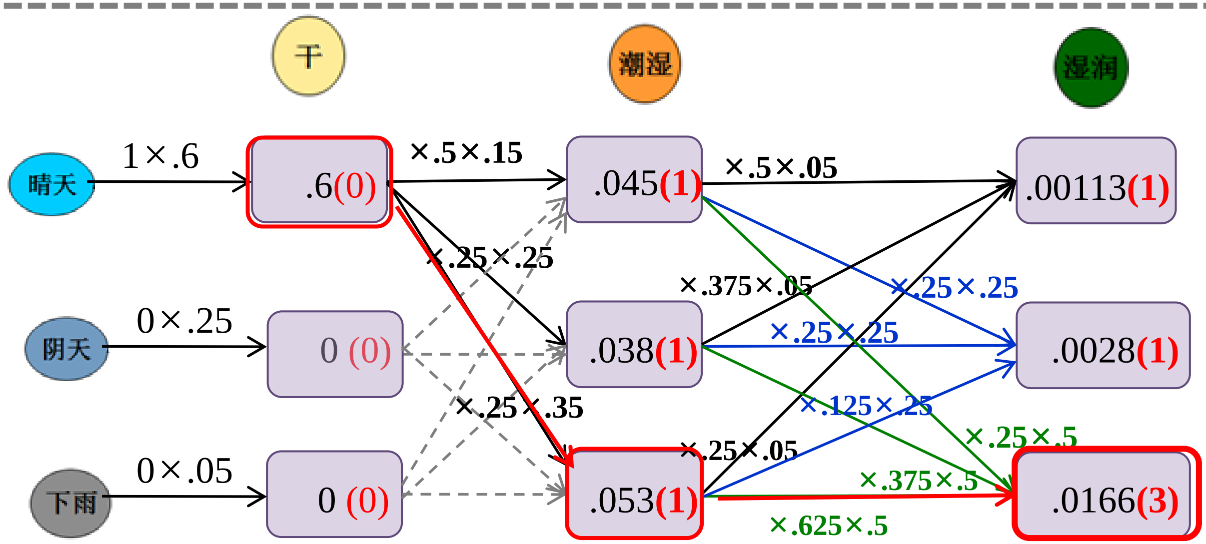
1. 初始化： $\delta_1(i) = \pi_i b_i(O_1)$， $\varphi_1(i) = 0$， $1 \le i \le N$

2. 递归： $\delta_t(j) = [\max\limits_{1\le i\le N} \delta_{t-1}(i) a_{ij}] b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$

   $\varphi_t(j) = \arg\max\limits_{1\le i\le N} [\delta_{t-1}(i) a_{ij}] b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$

3. 终结： $p^* = \max\limits_{1\le i\le N} [\delta_T(i)]$ ， $q_T^* = \arg\max\limits_{1\le i\le N} [\delta_T(i)]$

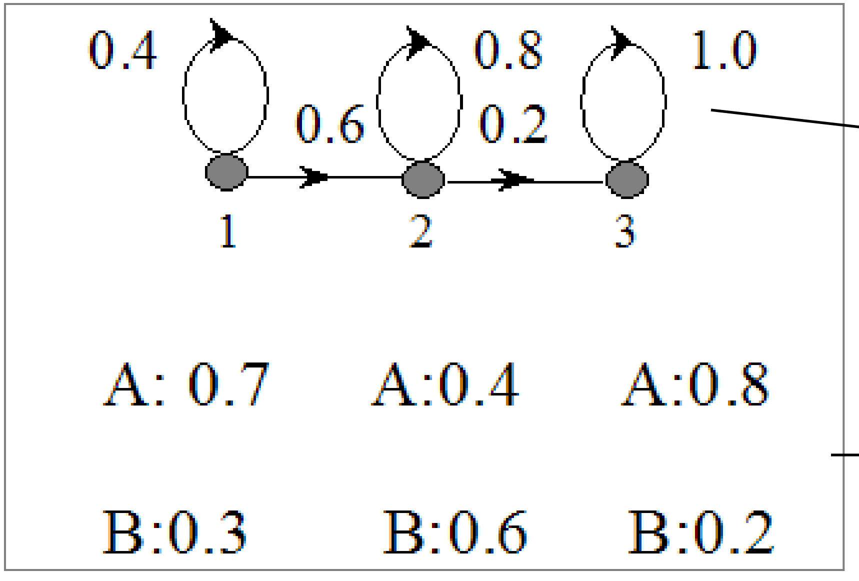
4. 路径回溯： $q_t^* = \varphi_{t+1}(q_{t+1}^*)$， $t = T-1, T-2, \ldots, 1$



（干燥、潮湿、湿润），最可能的天气序列:（晴、雨、雨）

9.2.2 隐马尔科夫模型（结构）—Ⅱ. HMM解码问题

例2： **HMM**模型如下，试根据Viterbi算法计算产生观察符号序列 O= {A B A B} 的最优状态序列Q



状态集Q { S1,S2,S3 }

观察序列集O={A,B}

解： 初始概率矩阵π=(1,0,0)

第一次观察时：

t=1

$$\delta_1(1)=\pi_1 * b_1(A)=1*0.7=0.7,\ \psi_1(1)=0$$

第二次观察时：

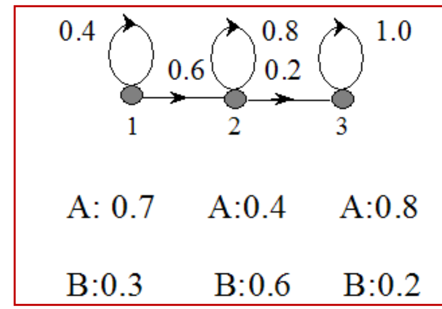
t=2

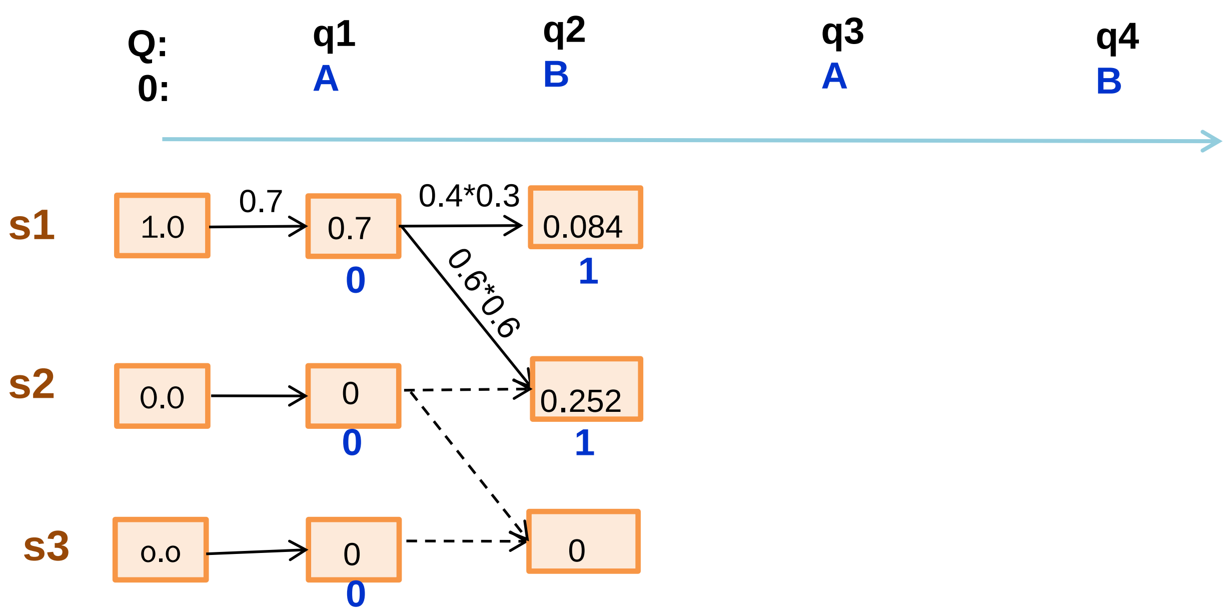
$$\delta_2(1)=\delta_1(1)*a_{11}*b_1(B)=0.7*0.4*0.3=0.084,\ \psi_2(1)=1$$

$$\delta_2(2)=\delta_1(1)*a_{12}*b_2(B)=0.7*0.6*0.6=0.252,\ \psi_2(2)=1$$

Viterbi 搜索算法

- 初始化：$\delta_1(i)=\pi_i b_i(O_1)$，$\varphi_1(i)=0$，$1\le i\le N$
- 递归：$\delta_t(j)=[\max\limits_{1\le i\le N}\delta_{t-1}(i)a_{ij}]b_j(O_t)$，$2\le t\le T, 1\le j\le N$
  $\varphi_t(j)=\arg\max\limits_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t)$，$2\le t\le T, 1\le j\le N$
- 终结：$p^*=\max\limits_{1\le i\le N}[\delta_T(i)]$ ， $q_T^*=\arg\max\limits_{1\le i\le N}[\delta_T(i)]$
- 路径回溯：$q_t^*=\varphi_{t+1}(q_{t+1}^*)$，$t=T-1,T-2,\ldots,1$





解:

第三次观察时:

t=3

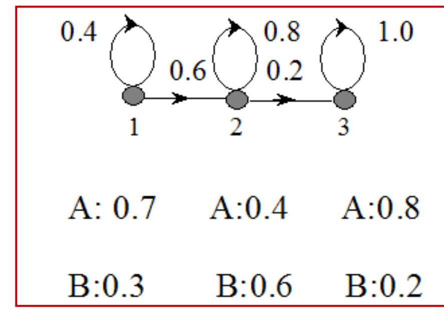
$$\delta_3(1)=\delta_2(1)*a_{11}*b_1(A)=0.084*0.4*0.7=0.02352,\ \psi_3(1)=1$$

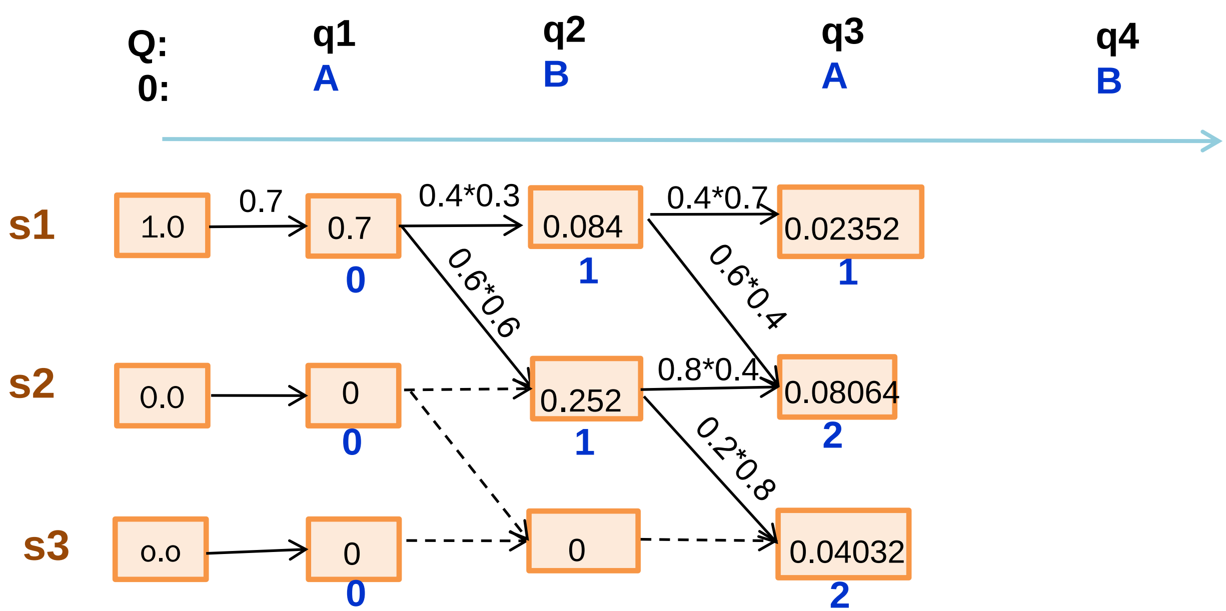
$$\delta_3(2)=\max\{\delta_2(1)*a_{12},\delta_2(2)*a_{22}\}*b_2(A)$$
$$=\max\{0.084*0.6,0.252*0.8\}*0.4\ ,\ \psi_3(2)=2$$
$$=0.08064$$

Viterbi 搜索算法

- 初始化: $\delta_1(i)=\pi_i b_i(O_1)$, $\varphi_1(i)=0$, $1\le i\le N$
- 递归: $\delta_t(j)=[\max\limits_{1\le i\le N}\delta_{t-1}(i)a_{ij}]b_j(O_t)$, $2\le t\le T, 1\le j\le N$

  $\varphi_t(j)=\arg\max\limits_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t)$, $2\le t\le T, 1\le j\le N$
- 终结: $p^*=\max\limits_{1\le i\le N}[\delta_T(i)]$, $q_T^*=\arg\max\limits_{1\le i\le N}[\delta_T(i)]$
- 路径回溯: $q_t^*=\varphi_{t+1}(q_{t+1}^*)$, $t=T-1,T-2,\dots,1$





解:

第四次观察时:

t=4

$$\delta_4(1)=\delta_3(1)*a_{11}*b_1(B)=0.0028224,\ \psi_4(1)=1$$

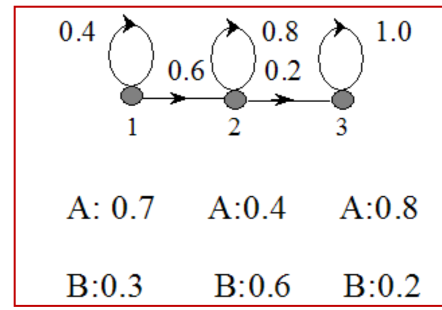
$$\delta_4(2)=\max\{\delta_3(1)*a_{12},\delta_3(2)*a_{22}\}*b_2(B)$$
$$=\max\{0.014112,0.064512\}*0.6\quad,\ \psi_4(2)=2$$
$$=0.0387072$$

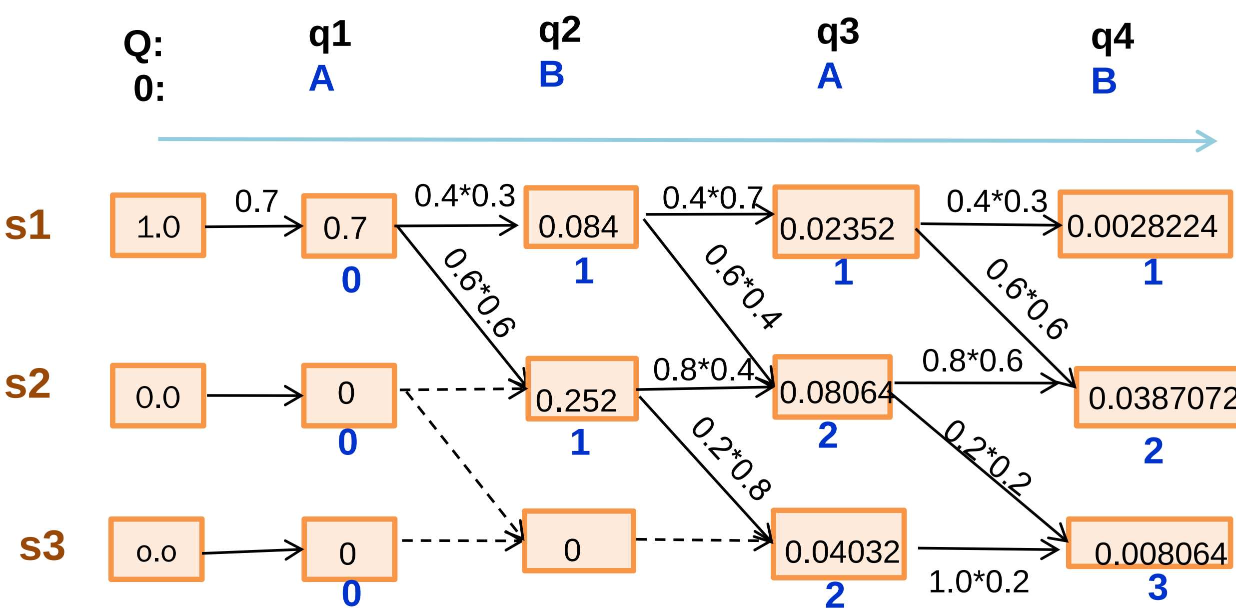
$$\delta_4(3)=\max\{\delta_3(2)*a_{23},\delta_3(3)*a_{33}\}*b_3(B)$$
$$=\max\{0.08064*0.2,0.04032*1\}*0.2,\ \psi_4(3)=3$$
$$=0.008064$$

Viterbi 搜索算法

- 初始化: $\delta_1(i)=\pi_i b_i(O_1)$, $\varphi_1(i)=0$, $1\le i\le N$
- 递归: $\delta_t(j)=[\max\limits_{1\le i\le N}\delta_{t-1}(i)a_{ij}]b_j(O_t)$, $2\le t\le T, 1\le j\le N$

  $\varphi_t(j)=\arg\max\limits_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t)$, $2\le t\le T, 1\le j\le N$
- 终结: $p^*=\max\limits_{1\le i\le N}[\delta_T(i)]$, $q_T^*=\arg\max\limits_{1\le i\le N}[\delta_T(i)]$
- 路径回溯: $q_t^*=\varphi_{t+1}(q_{t+1}^*)$, $t=T-1,T-2,\ldots,1$





解：

其递推结果为：

$$p^* = \max_{1\le i\le 3}[\delta_4(i)] = \max[0.0028224, 0.0387072, 0.008064] = 0.0387072$$

$$q_4^* = \arg\max_{1\le i\le 3}[\delta_4(i)] = 2$$

$$q_3^* = \psi_4(q_4^*) = \psi_4(2) = 2$$

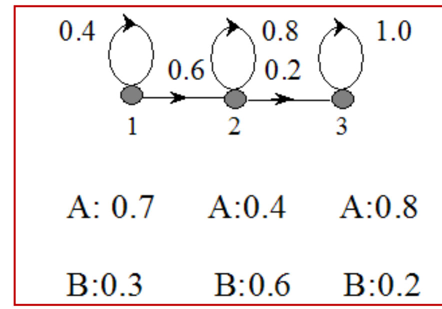
$$q_2^* = \psi_3(q_3^*) = \psi_3(2) = 2$$

$$q_1^* = \psi_2(q_2^*) = \psi_2(2) = 1$$

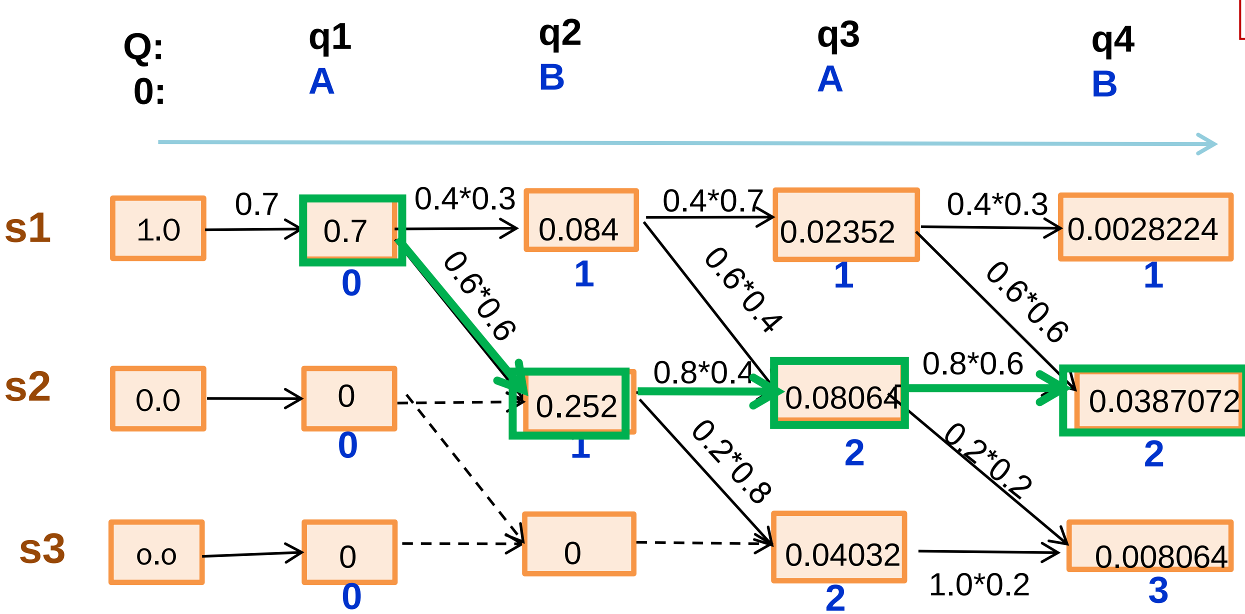
Viterbi 搜索算法

- 初始化： $\delta_1(i) = \pi_i b_i(O_1)$， $\varphi_1(i) = 0$， $1 \le i \le N$
- 递归： $\delta_t(j) = [\max_{1\le i\le N} \delta_{t-1}(i) a_{ij}] b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$

  $\varphi_t(j) = \arg\max_{1\le i\le N}[\delta_{t-1}(i) a_{ij}] b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$
- 终结： $p^* = \max_{1\le i\le N}[\delta_T(i)]$ ， $q_T^* = \arg\max_{1\le i\le N}[\delta_T(i)]$
- 路径回溯： $q_t^* = \varphi_{t+1}(q_{t+1}^*)$， $t = T-1, T-2, \ldots, 1$



最后的结果状态序列为：S1、S2、S2、S2



内 容 提 要

9.1 序列标注问题

9.2 隐马尔科夫模型HHM

9.2.1 马尔科夫模型

9.2.2 隐马尔科夫模型（结构）

9.2.3 隐马尔科夫模型学习

9.2.4 隐马模型应用

9.3 条件随机场 CRF

9.4 神经网络 RNN+ CRF

9.2.3 隐马尔科夫模型学习

HMM参数学习

隐马尔科夫模型参数

$$P(S_t \mid S_{t-1}) = \frac{P(S_{t-1} S_t)}{P(S_{t-1})} \qquad P(O_t \mid S_t) = \frac{P(O_t, S_t)}{P(S_t)}$$

训练思路：

通过观察序列 $O = O_1 O_2 \ldots O_T$作为训练数据，用最大似然估计，使得观察序列 O 的概率 $p(O|\mu)$ 最大。

9.2.3 隐马尔科夫模型学习

情况1： 产生观察序列 O 的状态 $Q = q_1 q_2 \ldots q_T$ 已知，可以采用有监督的学习方法，用最大似然估计来计算 μ 的参数：

$$\bar{\pi}_i = \delta(q_1, S_i)$$

$$\bar{a}_{ij} = \frac{Q\text{中从状态}q_i\text{转移到}q_j\text{的次数}}{Q\text{中所有从状态}q_i\text{转移到另一状态(包括}q_j\text{自身)的总数}} = \frac{\sum_{t=1}^{T-1} \delta(q_t, S_i) \times \delta(q_{t+1}, S_j)}{\sum_{t=1}^{T-1} \delta(q_t, S_i)}$$

其中，δ(x, y) 为克罗奈克(Kronecker)函数，当 x= y 时, δ(x, y)=1, 否则δ(x, y) = 0。

$$\bar{b}_j(k) = \frac{Q\text{中从状态}q_j\text{输出符号}v_k\text{的次数}}{Q\text{到达}q_j\text{的总次数}} = \frac{\sum_{t=1}^{T} \delta(q_t, S_j) \times \delta(O_t, v_k)}{\sum_{t=1}^{T} \delta(q_t, S_j)}$$

其中，v_k 是模型输出符号集中的第 k 个符号。

9.2.3 隐马尔科夫模型学习

情况2：HMM 中的状态序列Q 是观察不到的，这时，最大似然估计方法不可行。可以采用无监督的EM学习方法。

解决方法：

期望最大化EM 算法。根据**EM 算法**调节模型的参数 π_i a_{ij} ，$b_{j(k)}$，使得观察序列O的概率P(O|M)最大，主要使用**前向后向算法**（鲍姆-韦尔奇 Baum-Welch ）算法。

（略）

内 容 提 要

9.1 序列标注问题

- 9.2 隐马尔科夫模型HHM
  - 9.2.1 马尔科夫模型
  - 9.2.2 隐马尔科夫模型（结构）
  - 9.2.3 隐马尔科夫模型学习
  - 9.2.4 隐马模型应用
- 9.3 条件随机场 CRF
- 9.4 神经网络 RNN+ CRF

9.2.4 隐马模型应用

HMM模型在自然语言处理中有着广泛的应用

★ **分词：** HMM的评估问题：当分词出现多种可能时，求观察序列 $O=O_1O_2...O_T$ 的概率， 结果取 概率最大的序列 ；解码问题：用序列标注直接进行分词

★ **词性标注**： 相当HMM的解码问题。即求观察序列 $O=O_1O_2...O_T$ 下，概率最大的标注序列 **argmax P(Q|O ,μ)**

★ **其他：**如 短语识别、语音识别 ....

HMM模型相关工具　HTK 地址：http://htk.eng.cam.ac.uk/

9.2.4 隐马模型应用

1. 应用中HMM中各部分与自然语言的对应关系：

观察序列$O=O_1 O_2 \ldots O_T$：处理的语言单位，一般为 **词**

状态序列 S =S1S2…ST ：与语言单位对应的句法信息，一般为 **词类**

模型参数：初始状态概率、状态转移概率、发射概率 需要学习获得

9.2.4 隐马模型应用

2. 参数学习（获得）方法：

情况1：有大规模分词和词性标注语料　用最大似然估计方法计算各概率

$$\bar{\pi}_{pos_i} = \frac{\text{POS}_i\text{出现在句首的次数}}{\text{所有句首的个数}}$$

$$\bar{a}_{ij} = \frac{\text{从词类POS}_i\text{转移到词类POS}_j\text{的次数}}{\text{所有从状态POS}_i\text{转移到另一POS(包括POS}_j\text{)的总数}}$$

$$\bar{b}_j(k) = \frac{\text{从状态POS}_j\text{输出词}w_k\text{的次数}}{\text{状态POS}_j\text{出现的总次数}}$$

注：
- $\bar{a}_{ij}$ 中状态数目POS：为语料中词类标记符号的总个数

  不同的语料标记个数不同，如　滨州树库汉语33个词类标记；北大语料106个标记

- $\bar{b}_j(k)$ 为所有词类到各单词的概率

  如某单词无某一词类，则该词类到此单词的发射概率为 0

9.2.4 隐马模型应用

情况2： 无标注语料：

- 需要一部有词性标注的词典
  - 获取词类个数(POS状态数)；
  - 获取对应每种词类的词汇数(观察符号数)
- 利用无监督**EM迭代算法**获取初始 状态概率、状态转移概率 和 发射概率

9.2.4 隐马模型应用

例： HMM模型在词性标注中的应用

设，有如下从语料库训练得到的词性转移概率矩阵和词语生成概率矩阵

词性转移概率

词性	估计
$PROB(ART \mid \varnothing)$	0.71
$PROB(N \mid \varnothing)$	0.29
$PROB(N \mid ART)$	1
$PROB(V \mid N)$	0.43
$PROB(N \mid N)$	0.13
$PROB(P \mid N)$	0.44
$PROB(N \mid V)$	0.35
$PROB(ART \mid V)$	0.65
$PROB(ART \mid P)$	0.74
$PROB(N \mid P)$	0.26

词语生成概率

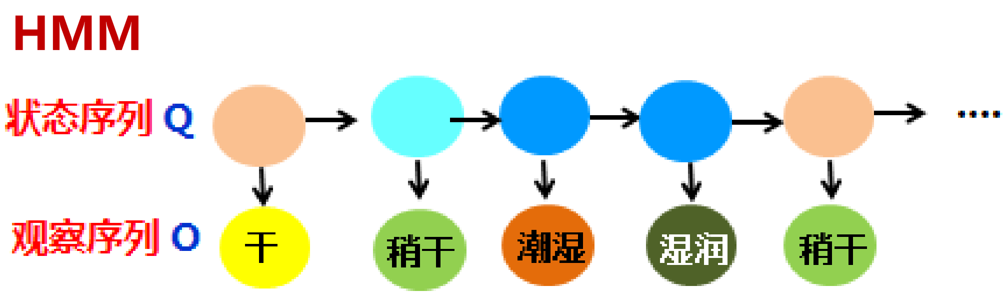
$PROB(the \mid ART)$	0.54
$PROB(flies \mid N)$	0.025
$PROB(flies \mid V)$	0.076
$PROB(like \mid V)$	0.1
$PROB(like \mid P)$	0.068
$PROB(like \mid N)$	0.012
$PROB(a \mid ART)$	0.360
$PROB(a \mid N)$	0.001
$PROB(flower \mid N)$	0.063
$PROB(flower \mid V)$	0.05
$PROB(birds \mid N)$	0.076

试对 “ flies like a flower ” 进行词性标注

9.2.4 隐马模型应用

解：　**问题求解目标**：对每个词标出其词性

该问题属于序列标注问题，可用HMM模型进行标注



观察集（词集）： flies， like， a， flower

状态集（词性集）： N，V，P，ART

- 共有 256 种可能标注结果
- 可用 Viterbi 搜索算法 解码

词性转移概率

词性	估计
PROB(ART | ø)	0.71
PROB(N | ø)	0.29
PROB(N | ART)	1
PROB(V | N)	0.43
PROB(N | N)	0.13
PROB(P | N)	0.44
PROB(N | V)	0.35
PROB(ART | V)	0.65
PROB(ART | P)	0.74
PROB(N | P)	0.26

词语生成概率

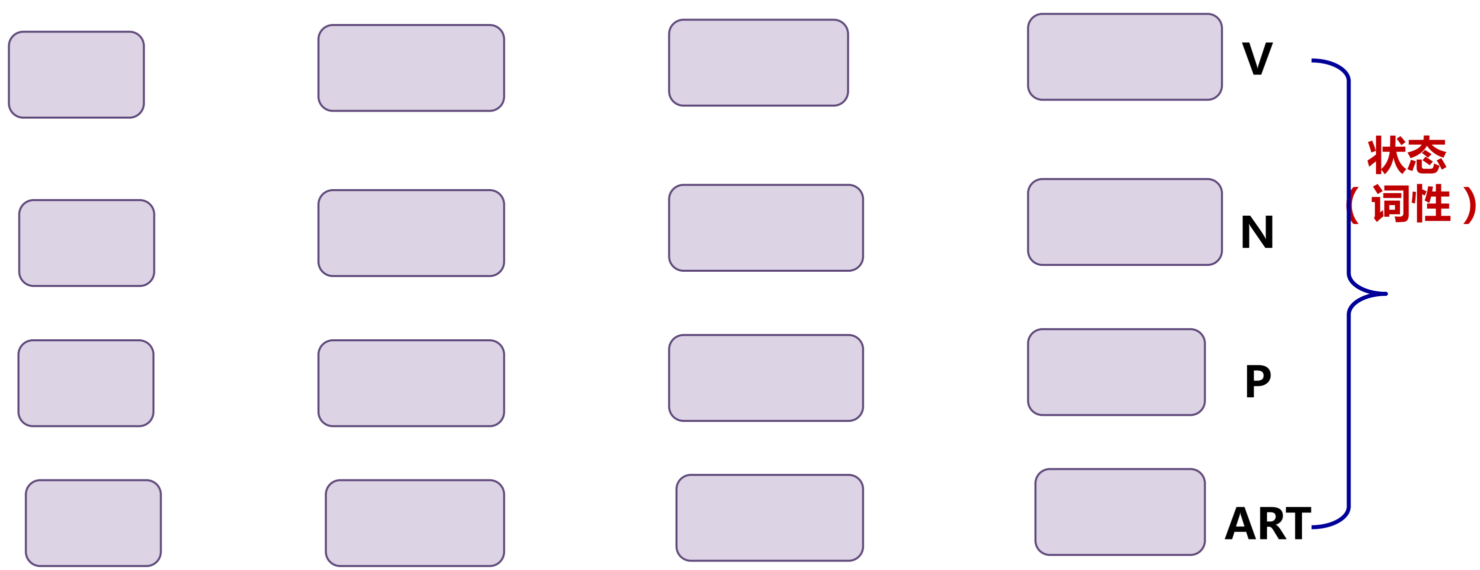
PROB(*the* | ART)	0.54
PROB(*flies* | N)	0.025
PROB(*flies* | V)	0.076
PROB(*like* | V)	0.1
PROB(*like* | P)	0.068
PROB(*like* | N)	0.012
PROB(*a* | ART)	0.360
PROB(*a* | N)	0.001
PROB(*flower* | N)	0.063
PROB(*flower* | V)	0.05
PROB(*birds* | N)	0.076

Viterbi 搜索算法

1. 初始化： $\delta_1(i)=\pi_i b_i(O_1)$， $\varphi_1(i)=0$， $1\le i\le N$
2. 递归： $\delta_t(j)=[\max\limits_{1\le i\le N}\delta_{t-1}(i)a_{ij}]b_j(O_t)$， $2\le t\le T, 1\le j\le N$

   $\varphi_t(j)=\arg\max\limits_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t)$， $2\le t\le T, 1\le j\le N$
3. 终结： $p^*=\max\limits_{1\le i\le N}[\delta_T(i)]$ ， $q_T^*=\arg\max\limits_{1\le i\le N}[\delta_T(i)]$
4. 路径回溯： $q_t^*=\varphi_{t+1}(q_{t+1}^*)$， $t=T-1,T-2,\dots,1$

观察序列（词）　flies　like　a　flower



词性转移概率

词性	估计
$PROB(ART \mid \phi)$	0.71
$PROB(N \mid \phi)$	0.29
$PROB(N \mid ART)$	1
$PROB(V \mid N)$	0.43
$PROB(N \mid N)$	0.13
$PROB(P \mid N)$	0.44
$PROB(N \mid V)$	0.35
$PROB(ART \mid V)$	0.65
$PROB(ART \mid P)$	0.74
$PROB(N \mid P)$	0.26

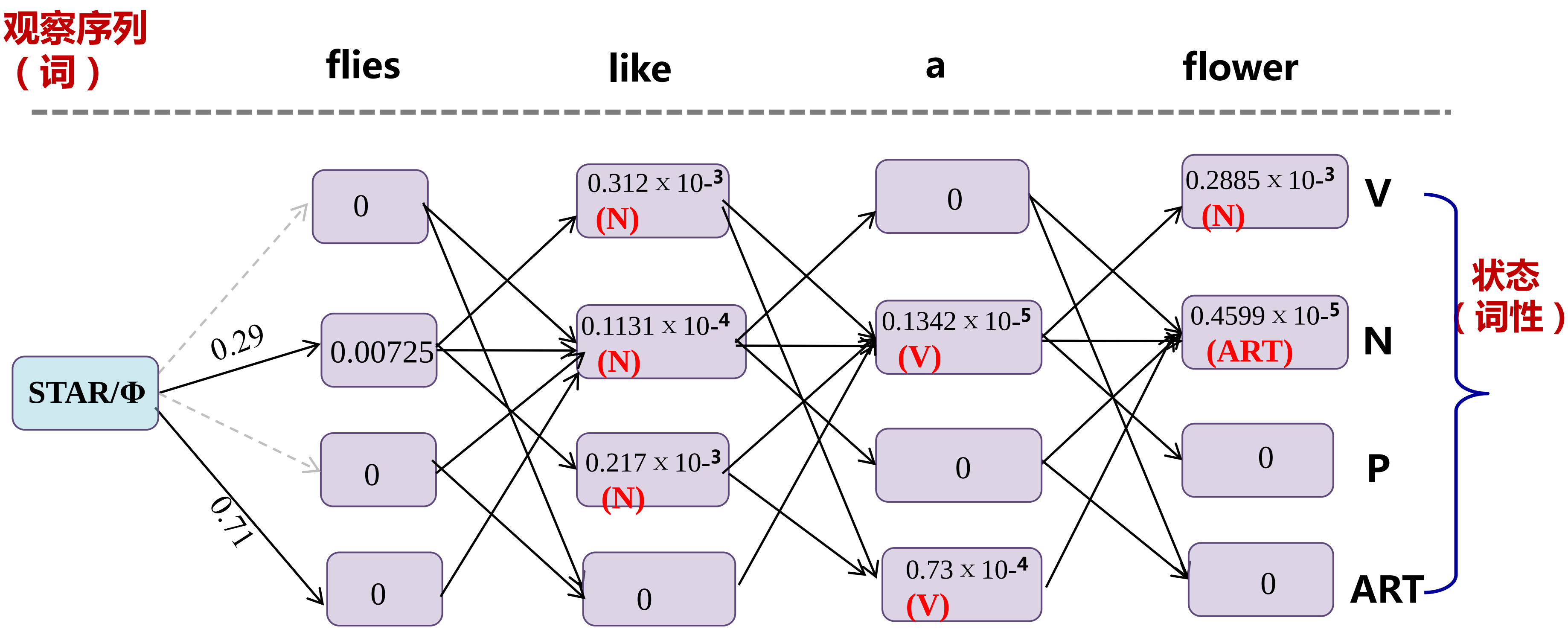
词语生成概率

$PROB(the \mid ART)$	0.54
$PROB(flies \mid N)$	0.025
$PROB(flies \mid V)$	0.076
$PROB(like \mid V)$	0.1
$PROB(like \mid P)$	0.068
$PROB(like \mid N)$	0.012
$PROB(a \mid ART)$	0.360
$PROB(a \mid N)$	0.001
$PROB(flower \mid N)$	0.063
$PROB(flower \mid V)$	0.05
$PROB(birds \mid N)$	0.076

Viterbi 搜索算法

1. 初始化： $\delta_1(i)=\pi_i b_i(O_1)$， $\varphi_1(i)=0$， $1 \le i \le N$
2. 递归： $\delta_t(j)=[\max_{1\le i\le N}\delta_{t-1}(i)a_{ij}]b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$

   $\varphi_t(j)=\arg\max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t)$， $2 \le t \le T, 1 \le j \le N$
3. 终结： $p^* = \max_{1\le i\le N}[\delta_T(i)]$ ， $q_T^* = \arg\max_{1\le i\le N}[\delta_T(i)]$
4. 路径回溯： $q_t^* = \varphi_{t+1}(q_{t+1}^*)$， $t = T-1, T-2, \ldots, 1$



词性转移概率

词性	估计
PROB(ART | ø)	0.71
PROB(N | ø)	0.29
PROB(N | ART)	1
PROB(V | N)	0.43
PROB(N | N)	0.13
PROB(P | N)	0.44
PROB(N | V)	0.35
PROB(ART | V)	0.65
PROB(ART | P)	0.74
PROB(N | P)	0.26

词语生成概率

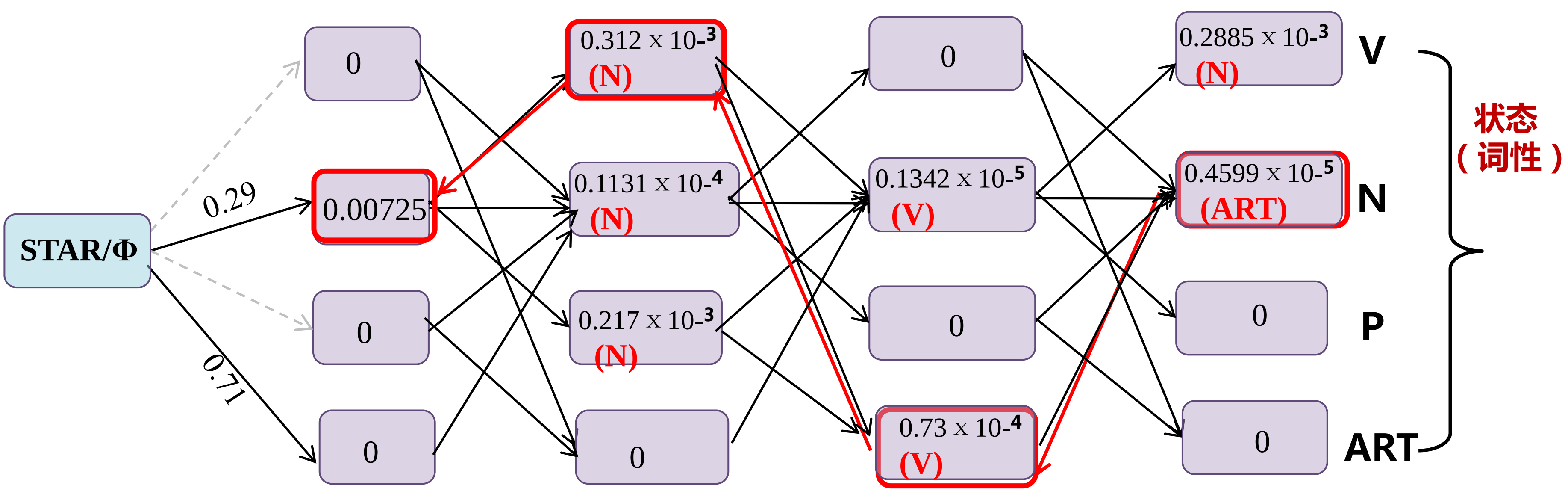
PROB(*the* | ART)	0.54
PROB(*flies* | N)	0.025
PROB(*flies* | V)	0.076
PROB(*like* | V)	0.1
PROB(*like* | P)	0.068
PROB(*like* | N)	0.012
PROB(*a* | ART)	0.360
PROB(*a* | N)	0.001
PROB(*flower* | N)	0.063
PROB(*flower* | V)	0.05
PROB(*birds* | N)	0.076

Viterbi 搜索算法

1. 初始化：$\delta_1(i)=\pi_i b_i(O_1)$，$\varphi_1(i)=0$，$1\le i\le N$
2. 递归：$\delta_t(j)=[\max\limits_{1\le i\le N}\delta_{t-1}(i)a_{ij}]b_j(O_t)$，$2\le t\le T, 1\le j\le N$

   $\varphi_t(j)=\arg\max\limits_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t)$，$2\le t\le T, 1\le j\le N$
3. 终结：$p^*=\max\limits_{1\le i\le N}[\delta_T(i)])$ ，$q_T^*=\arg\max\limits_{1\le i\le N}[\delta_T(i)]$
4. 路径回溯：$q_t^*=\varphi_{t+1}(q_{t+1}^*)$，$t=T-1,T-2,\dots,1$

观察序列（词）　flies /N　like /V　a /ART　flower /N　结果



9.2.4 隐马模型应用

HMM 等生产式模型存在的问题

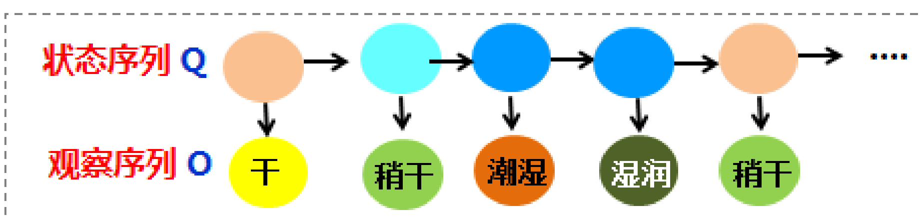
1. 由于生成模型定义的是联合概率，必须列举所有观察序列的可能值，这对多数领域来说是比较困难的。

   在自然语言处理中，常知道各种各样但又不完全确定的信息，需要一个统一的模型将这些信息综合起来。

2. 输出独立性假设要求序列数据严格相互独立才能保证推导的正确性，导致其不能考虑上下文特征

   在自然语言处理中，常常需要考虑上下文关系。

HMM



解决方案 → 条件随机场（CRF）

参考文献：

http://wenku.baidu.com/view/3cf29130f111f18583d05a57.html

http://wenku.baidu.com/view/9121f528bd64783e09122b80.html

http://wenku.baidu.com/view/bbd57f82fc4ffe473268ab59.html?from=search

邹博，熵导论与最大熵模型，机器学习班

李航，统计学习方法　清华大学出版社

宗成庆，统计自然语言处理（第2版）

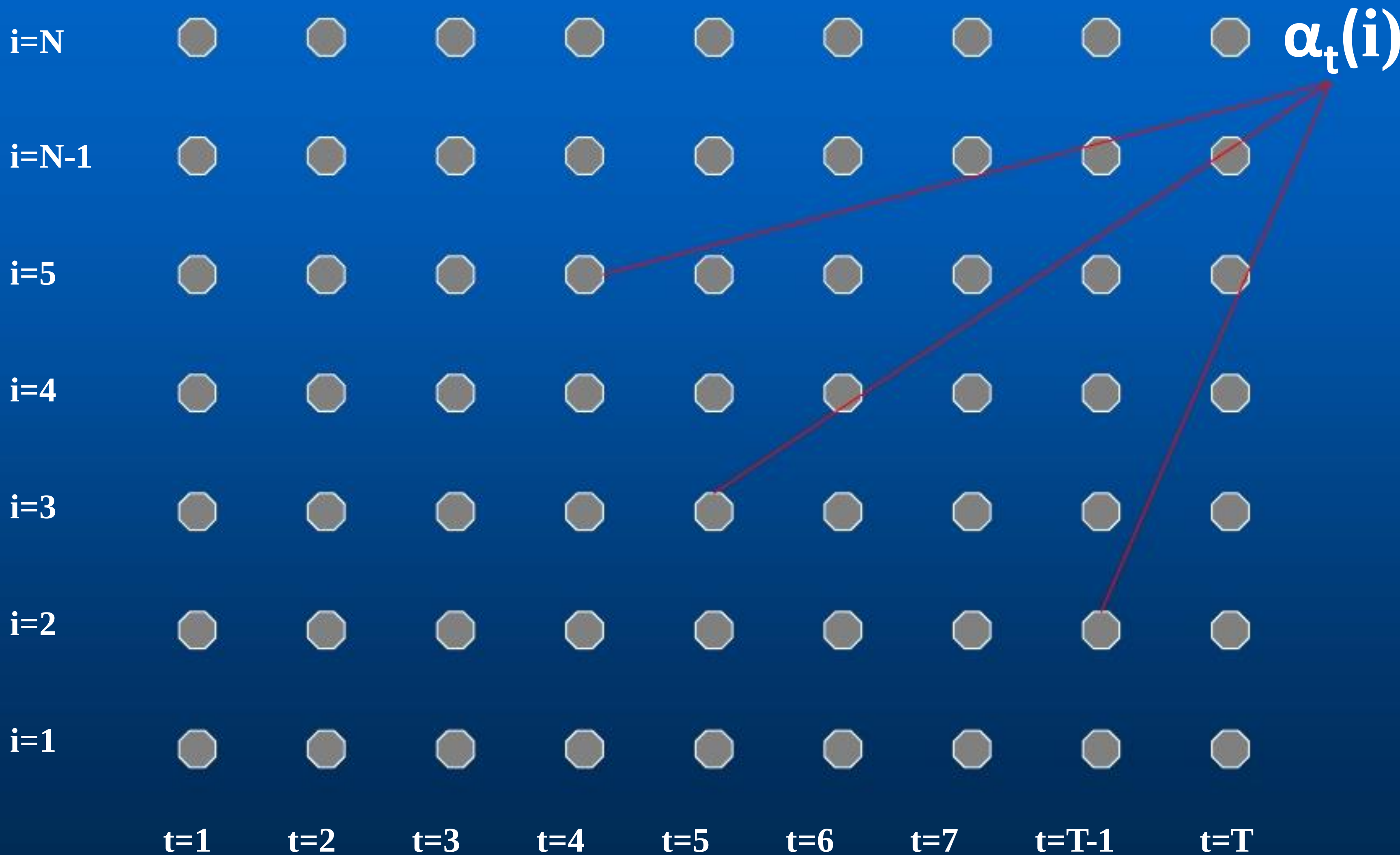
在 此 表 示 感 谢！

谢谢各位！

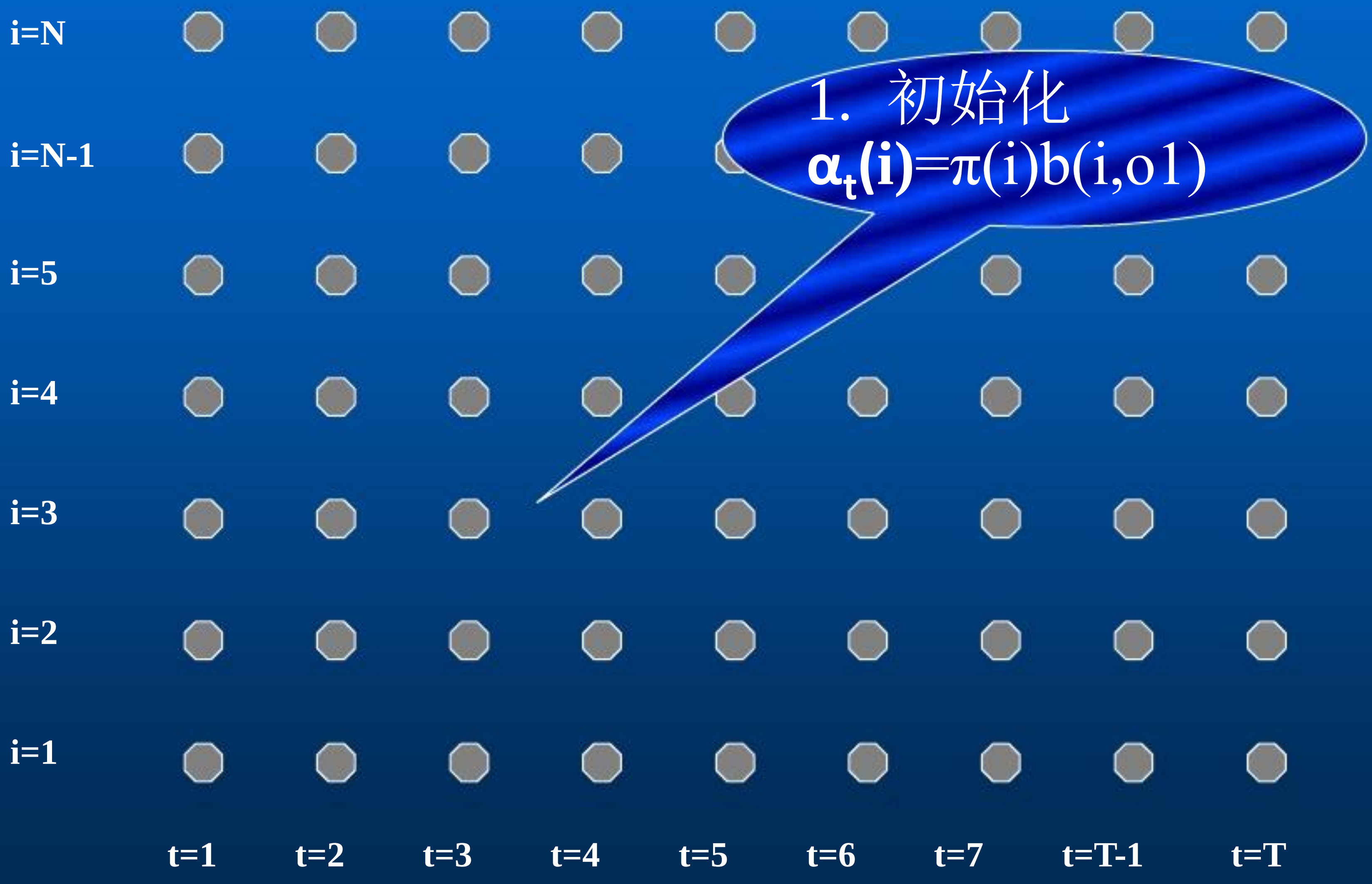
附录：

前向计算过程演示

前向算法过程演示



前向算法过程演示

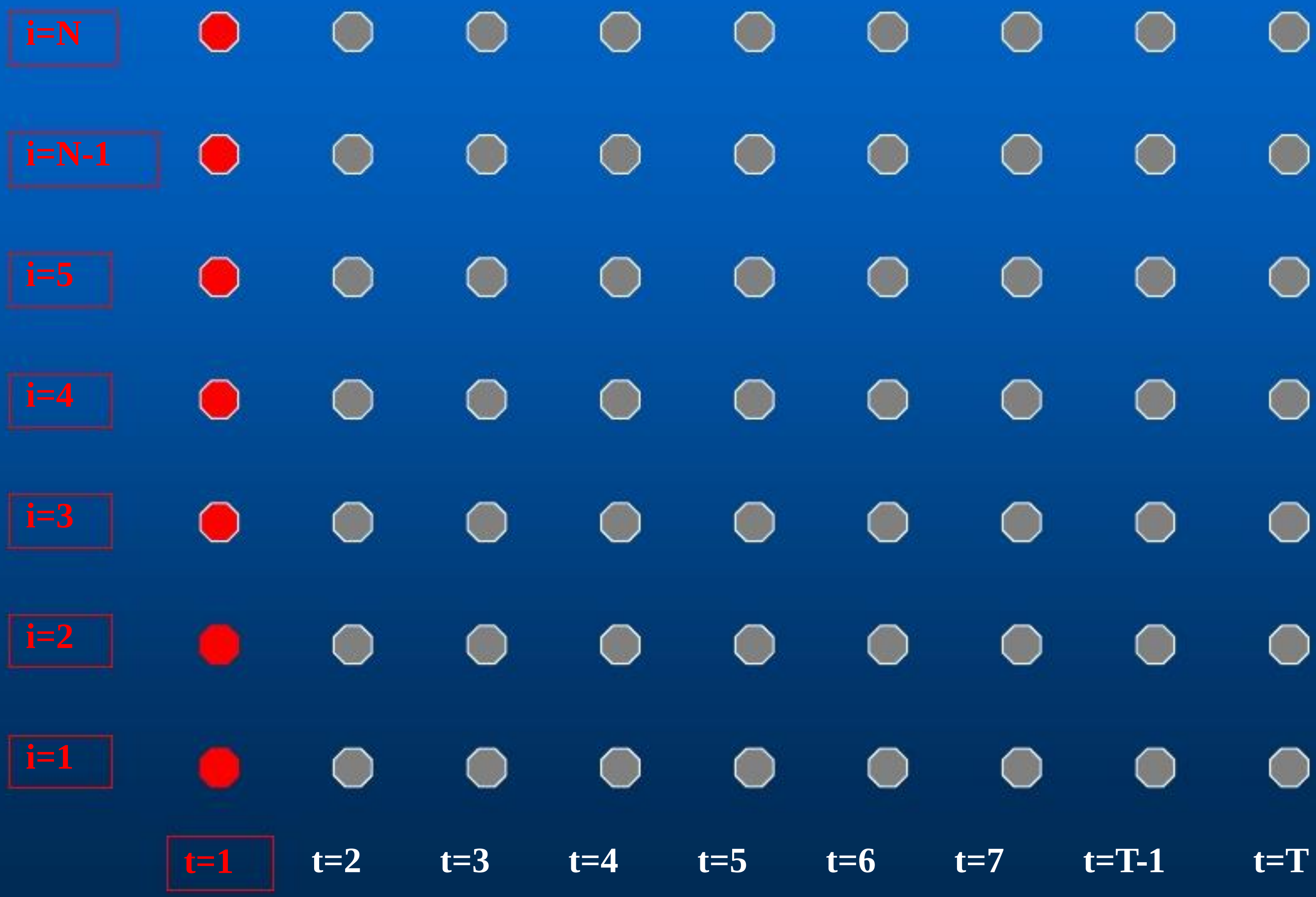

i=N
i=N-1
i=5
i=4
i=3
i=2
i=1
1. 初始化
$\alpha_t(i)=\pi(i)b(i,o1)$
t=1
t=2
t=3
t=4
t=5
t=6
t=7
t=T-1
t=T

前向算法过程演示

前向算法过程演示

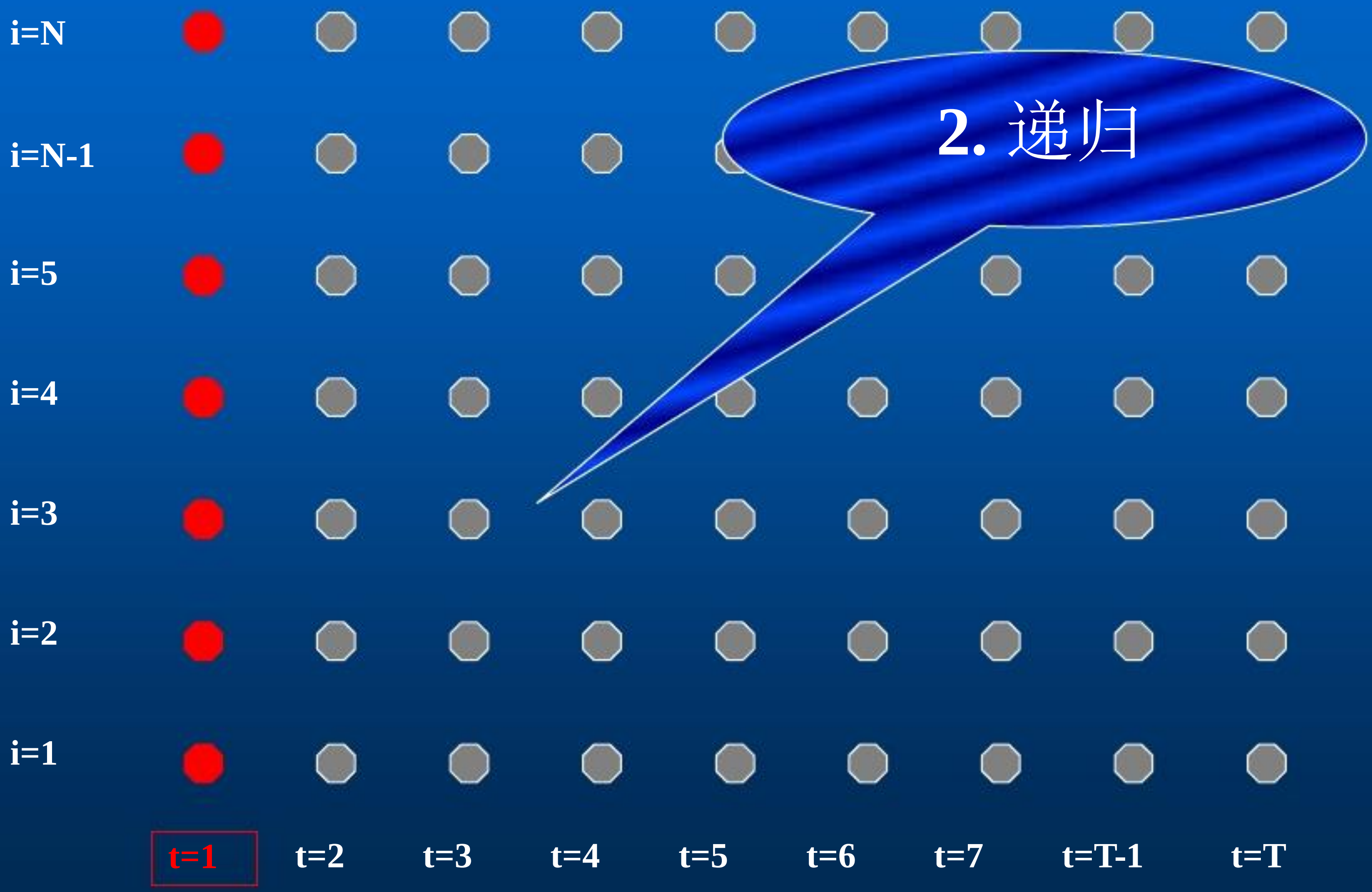
前向算法过程演示

前向算法过程演示

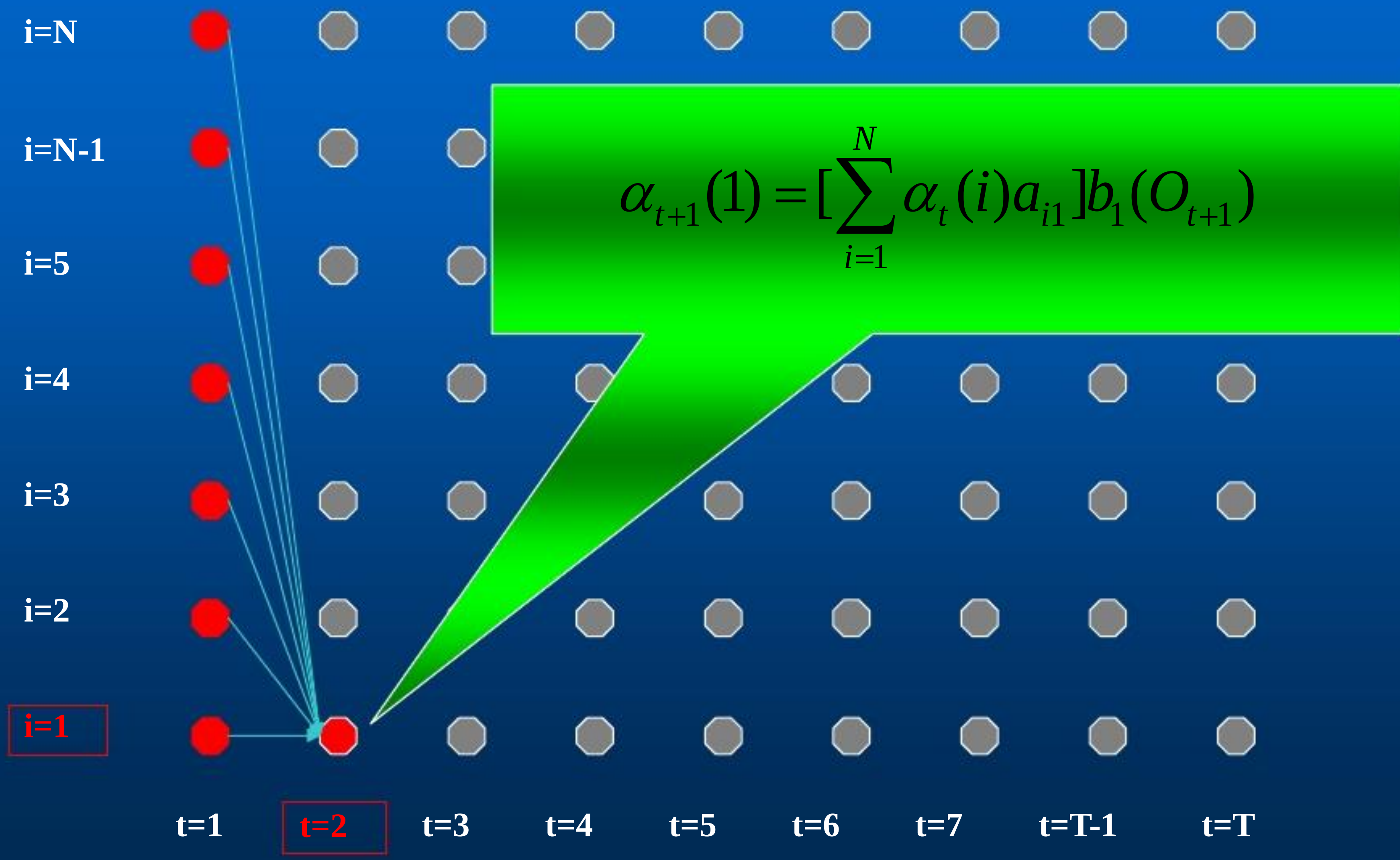


前向算法过程演示

$$\alpha_{t+1}(j)=[\sum_{i=1}^{N}\alpha_t(i)a_{ij}]b_j(O_{t+1})$$

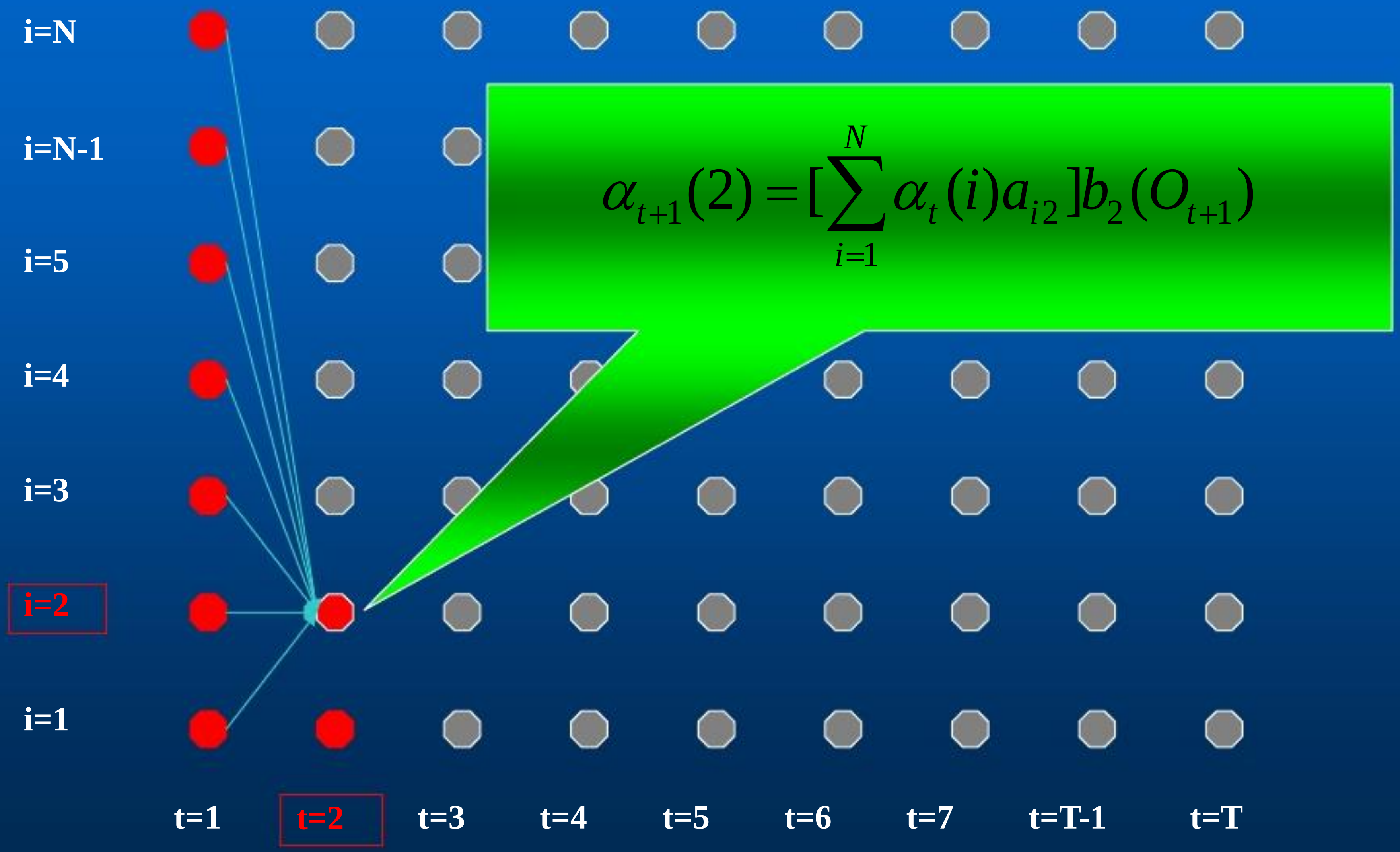


前向算法过程演示

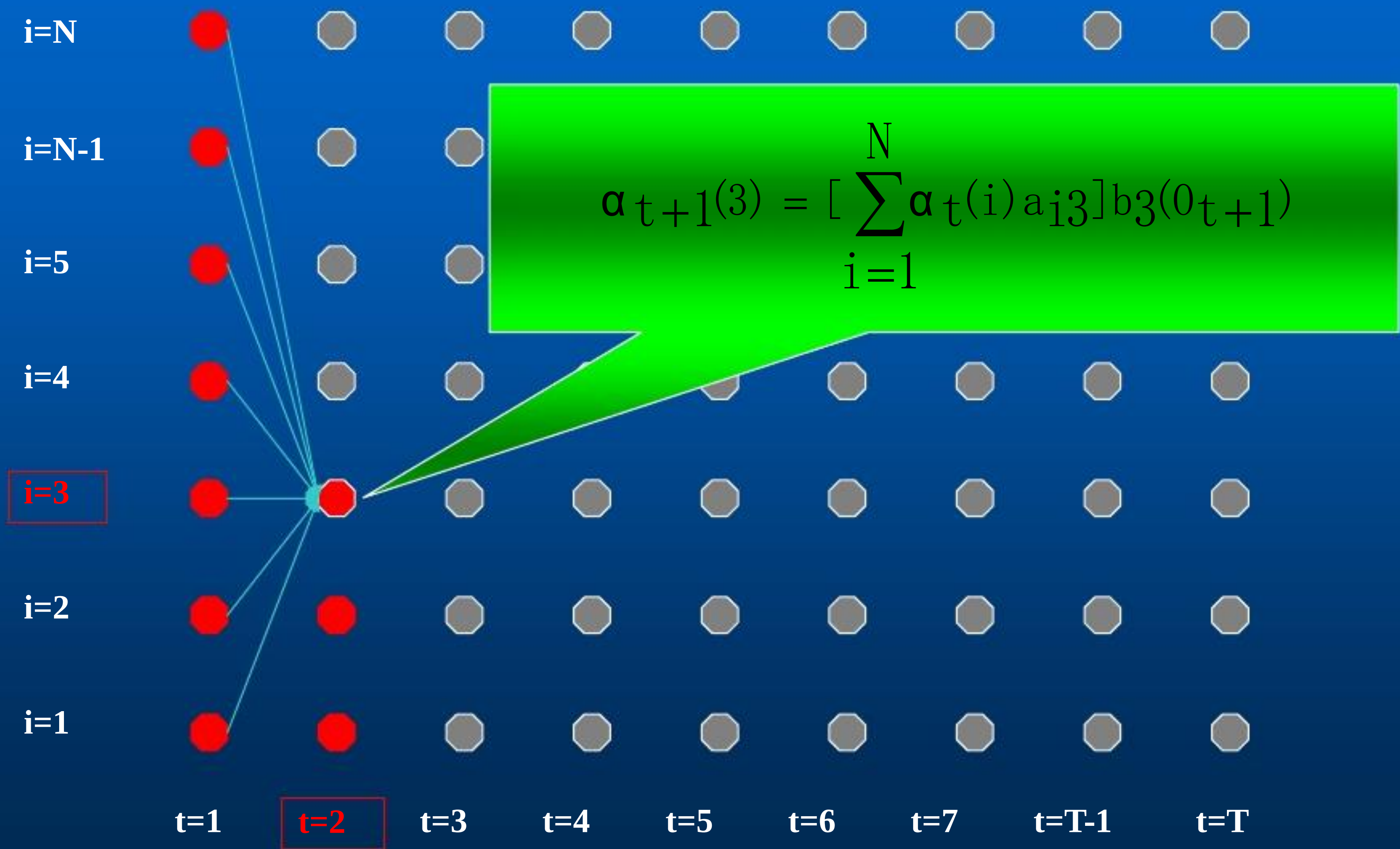


i=N
i=N-1
i=5
i=4
i=3
i=2
i=1
$$\alpha_{t+1}(1) = [\sum_{i=1}^{N} \alpha_t(i) a_{i1}] b_1(O_{t+1})$$
t=1
t=2
t=3
t=4
t=5
t=6
t=7
t=T-1
t=T

前向算法过程演示

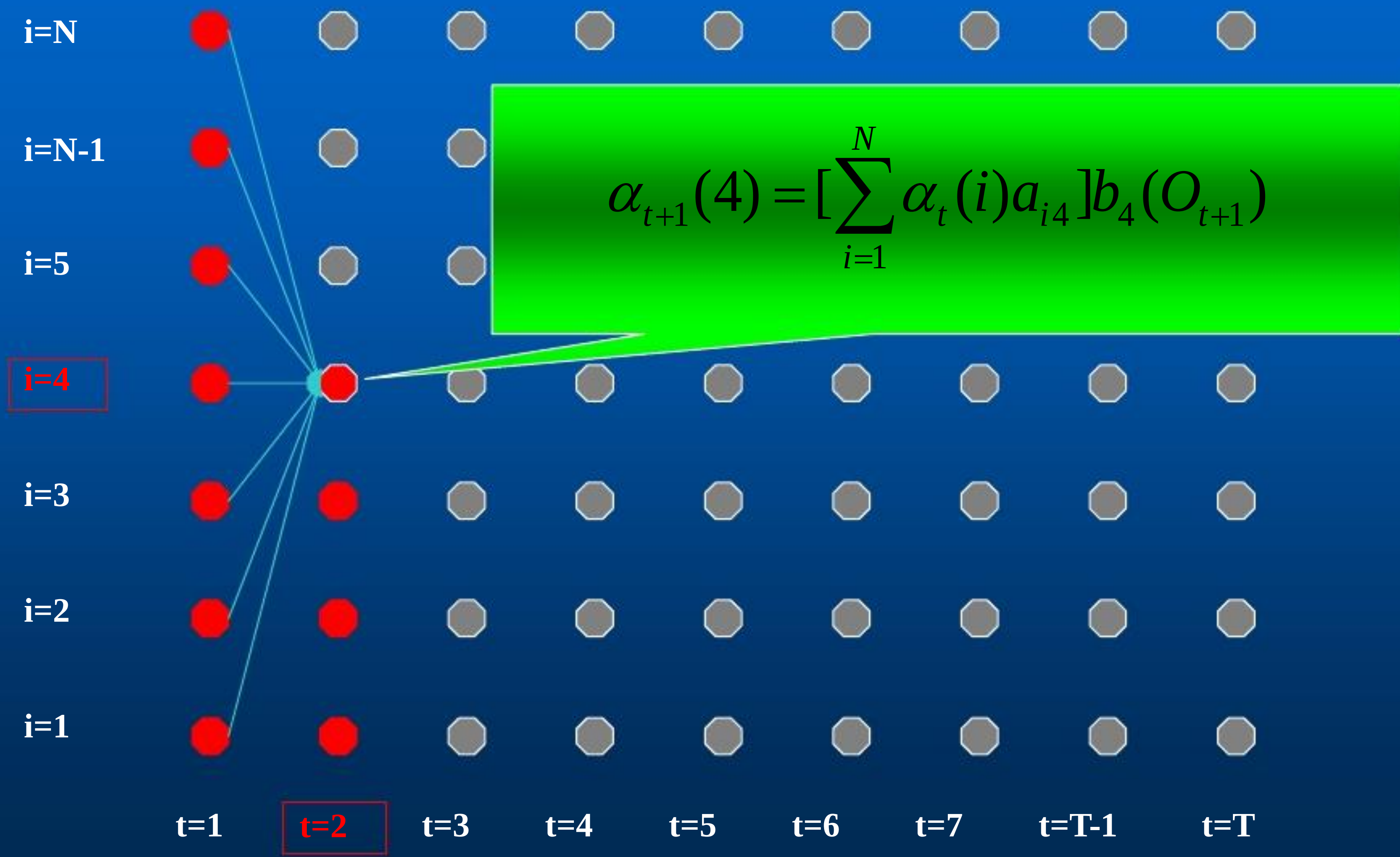

i=N
i=N-1
i=5
i=4
i=3
i=2
i=1
$\alpha_{t+1}(2) = [\sum_{i=1}^{N} \alpha_t(i) a_{i2}] b_2(O_{t+1})$
t=1
t=2
t=3
t=4
t=5
t=6
t=7
t=T-1
t=T

前向算法过程演示



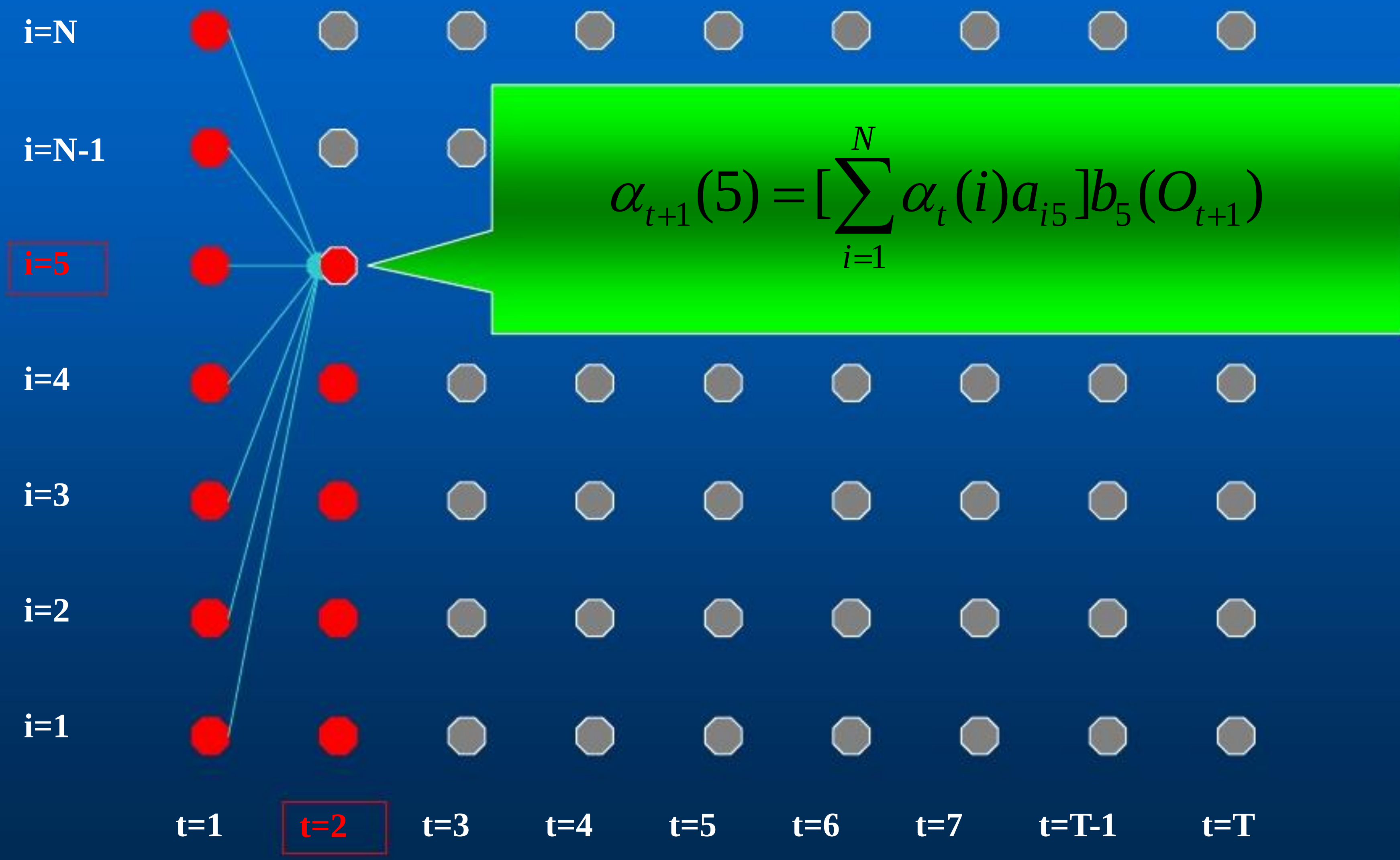
$$\alpha_{t+1}(3) = \left[\sum_{i=1}^{N} \alpha_t(i)\, a_{i3}\right] b_3(O_{t+1})$$

前向算法过程演示



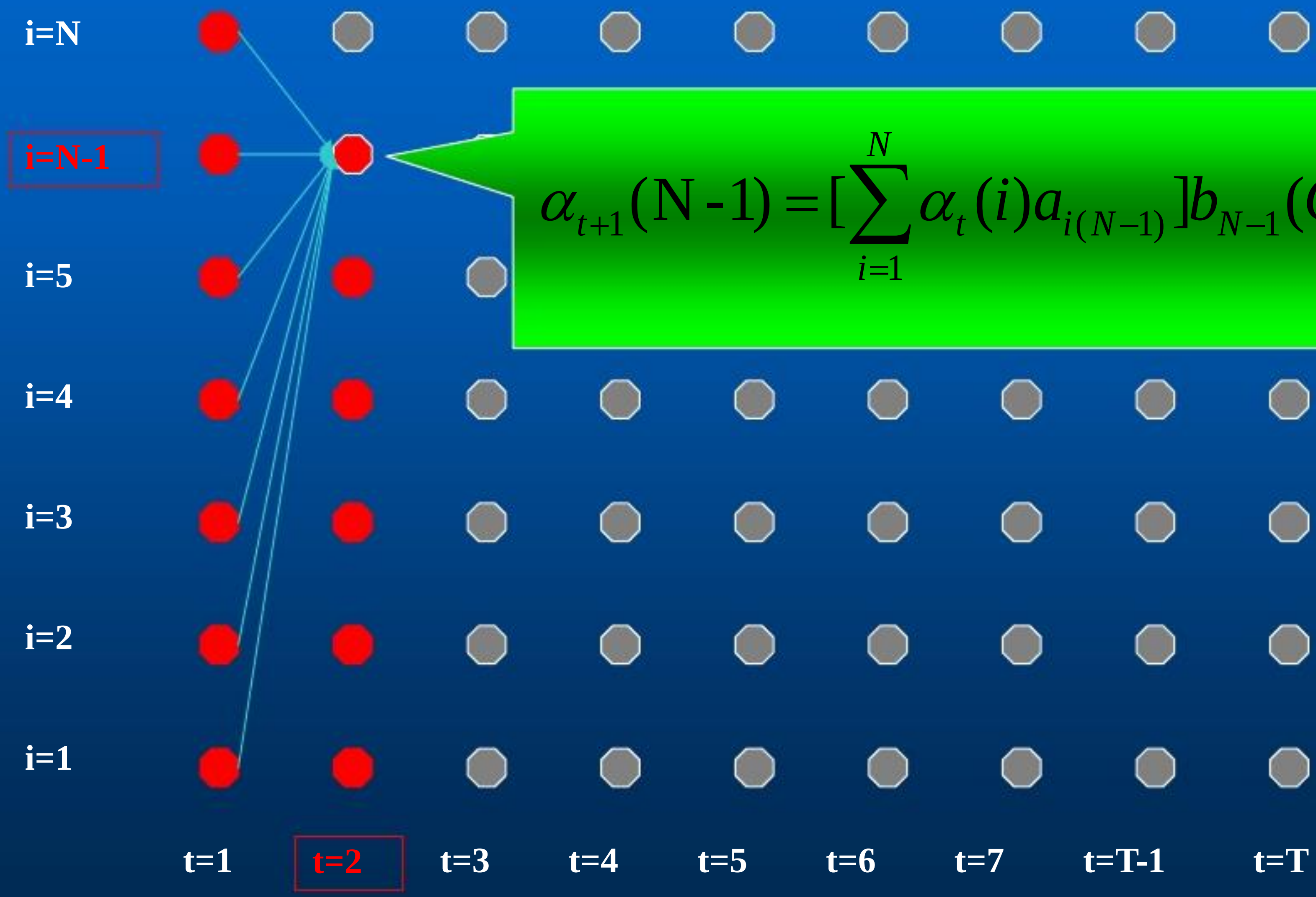
$$\alpha_{t+1}(4) = [\sum_{i=1}^{N} \alpha_t(i) a_{i4}] b_4(O_{t+1})$$

前向算法过程演示



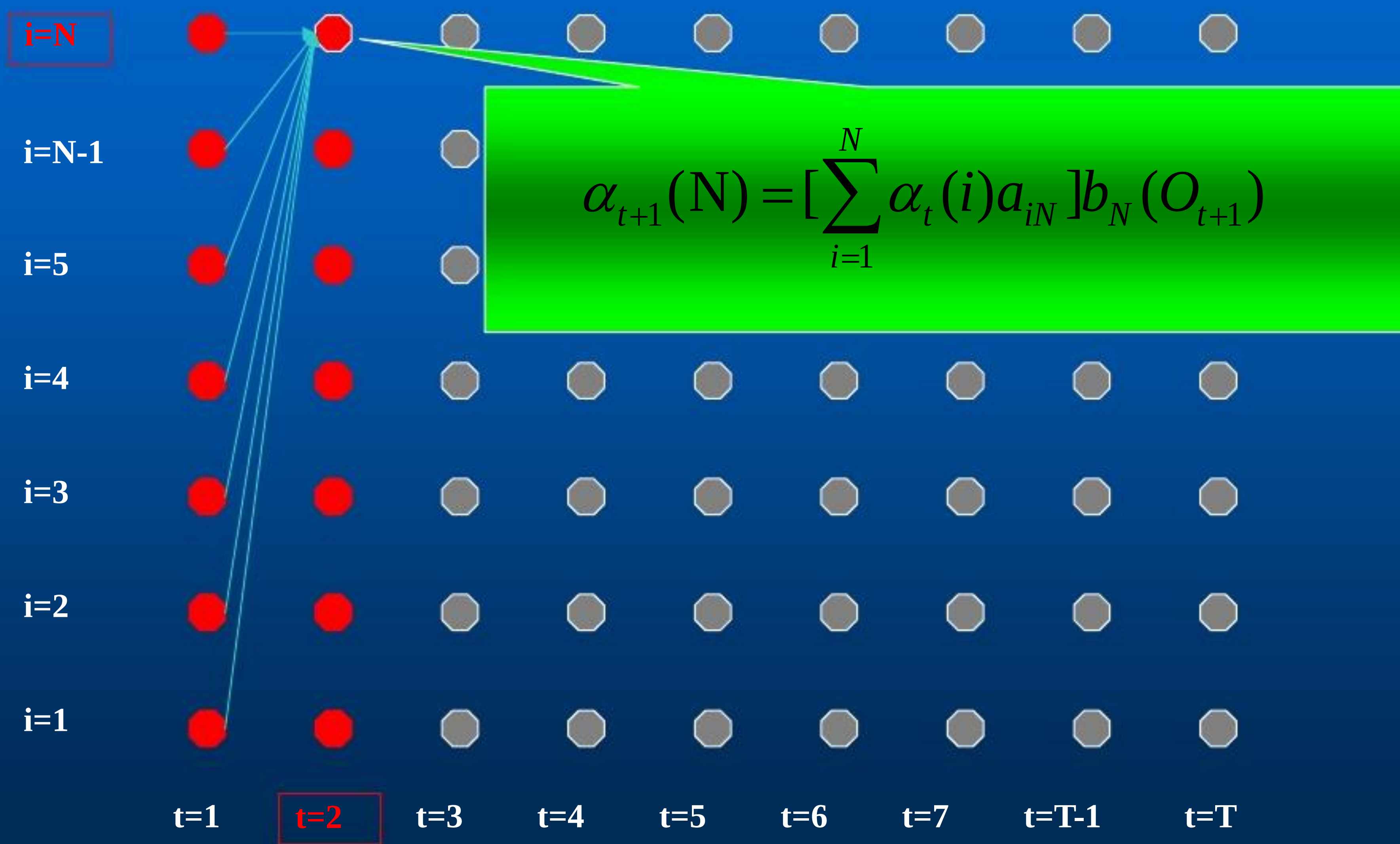
$$\alpha_{t+1}(5)=[\sum_{i=1}^{N}\alpha_t(i)a_{i5}]b_5(O_{t+1})$$

前向算法过程演示



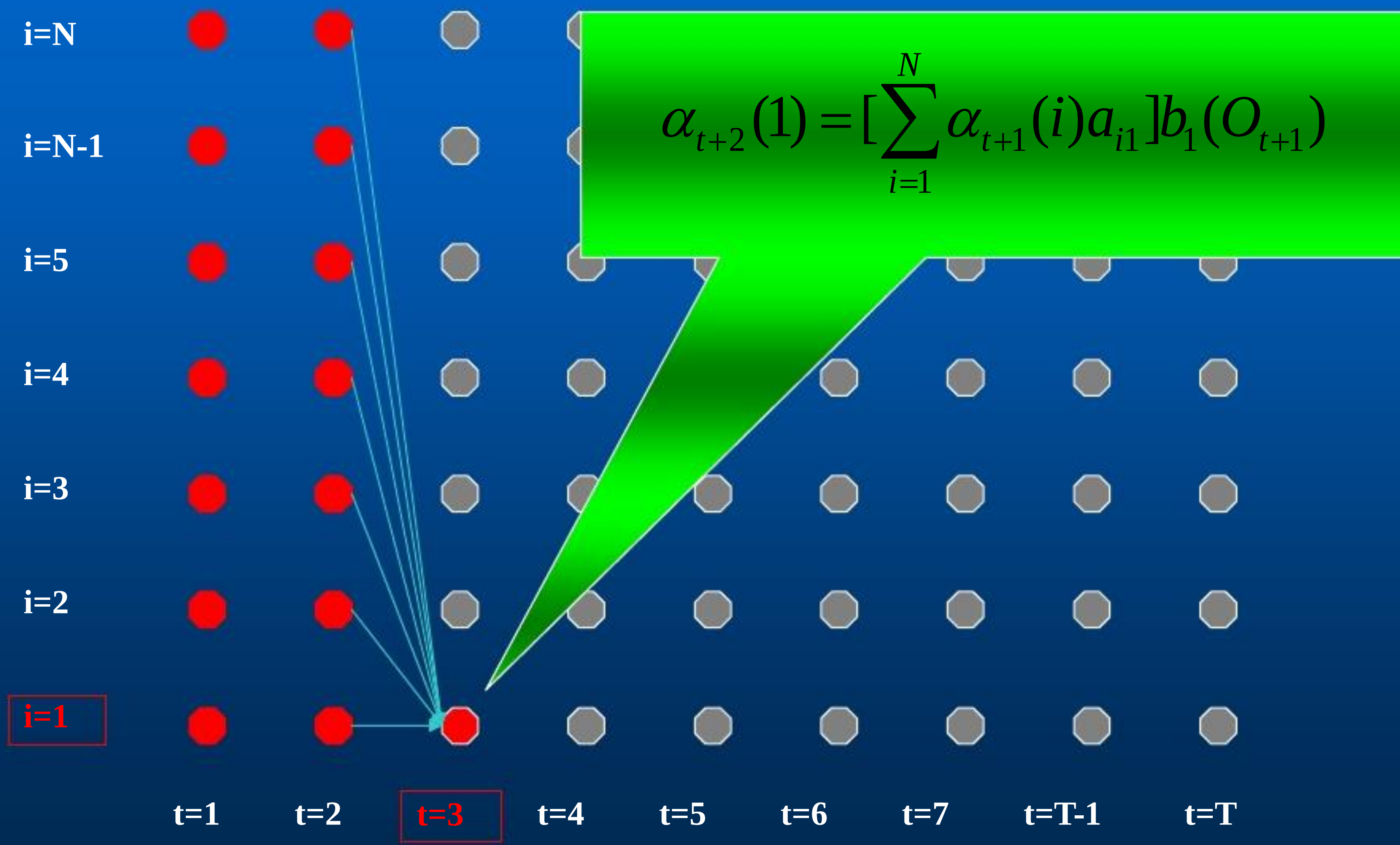
$$\alpha_{t+1}(\mathrm{N-1}) = [\sum_{i=1}^{N} \alpha_t(i) a_{i(N-1)}] b_{N-1}(O_{t+1})$$

前向算法过程演示



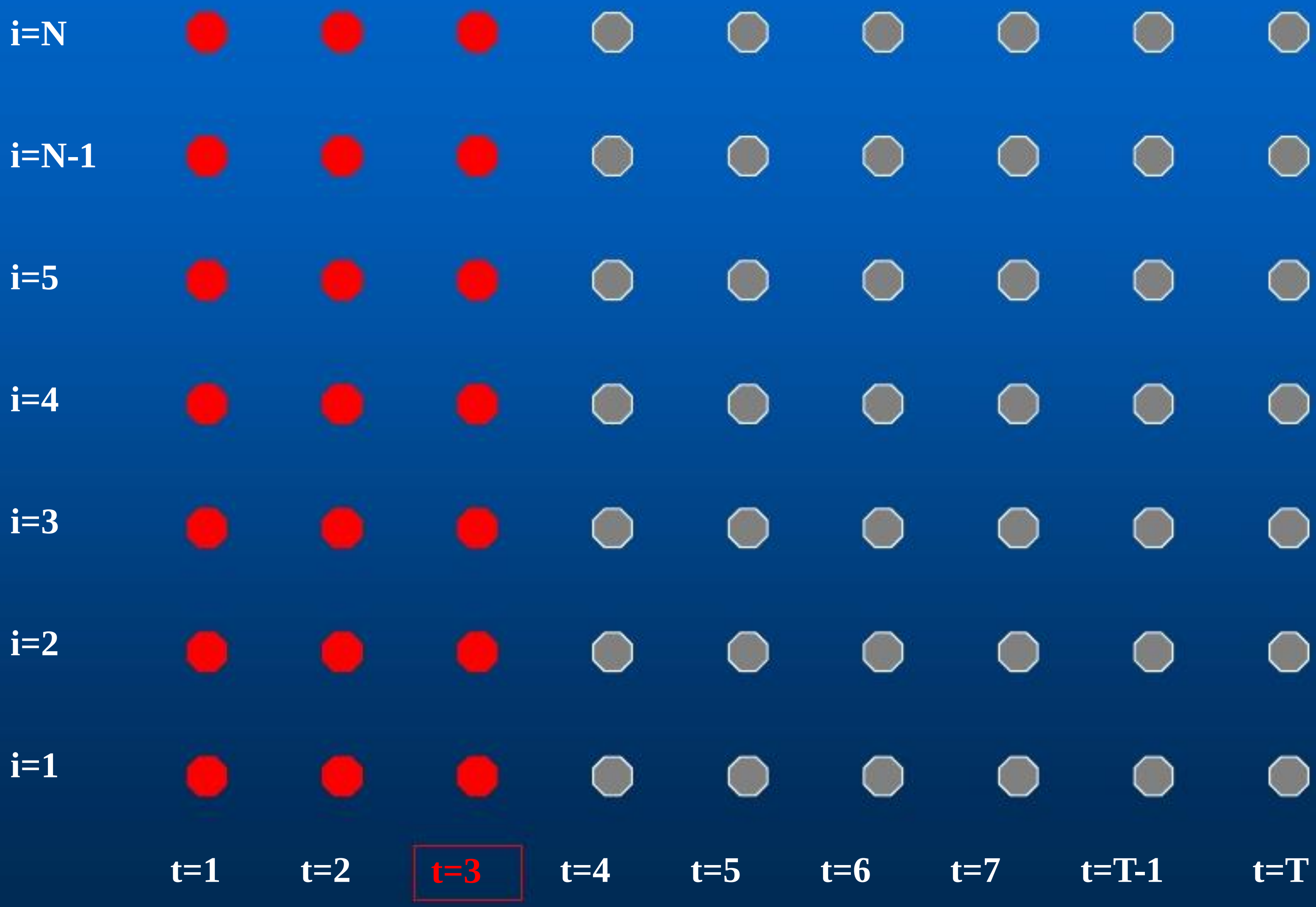
$$\alpha_{t+1}(\mathrm{N})=[\sum_{i=1}^{N}\alpha_t(i)a_{iN}]b_N(O_{t+1})$$

前向算法过程演示

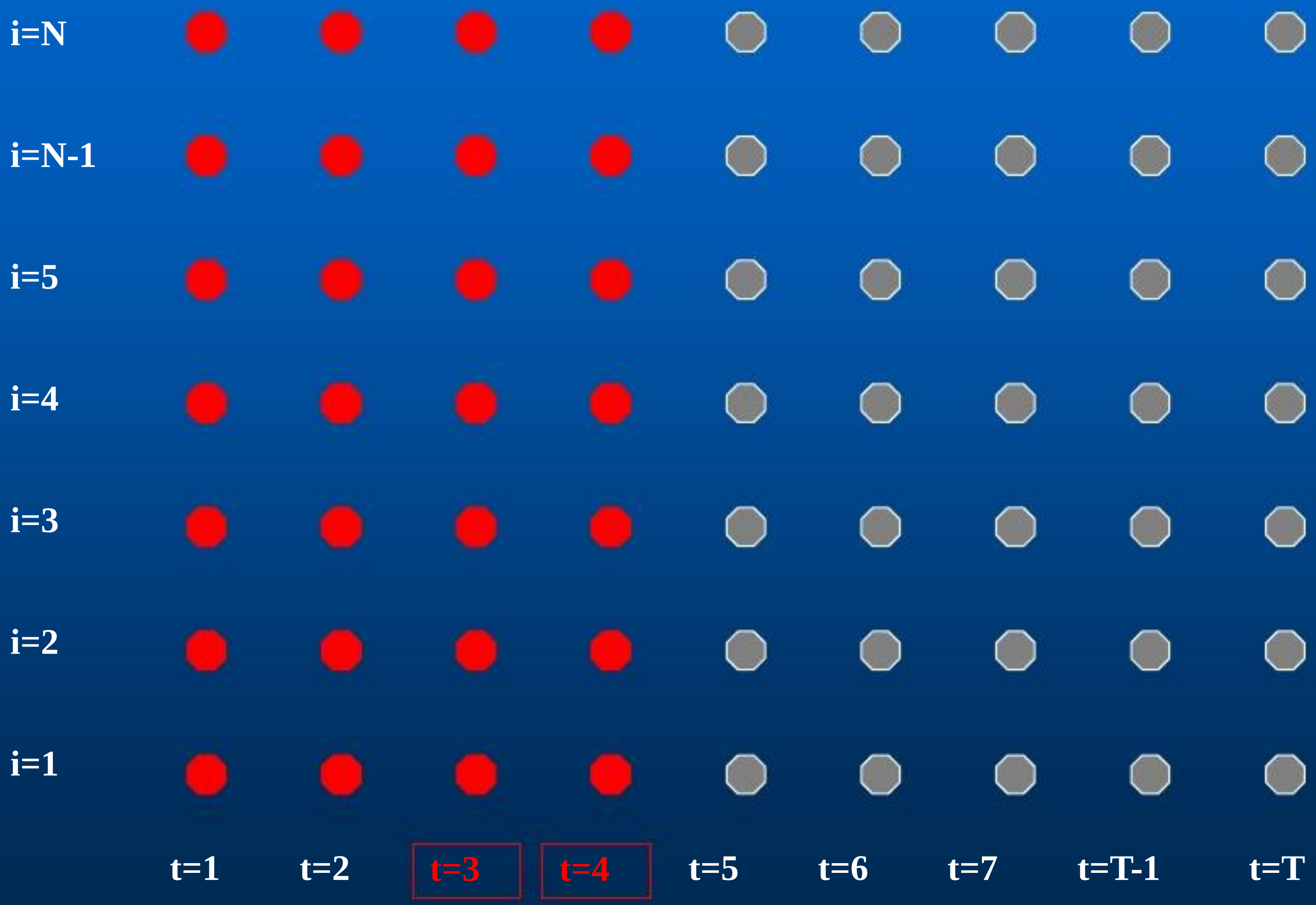


$$\alpha_{t+2}(1) = [\sum_{i=1}^{N} \alpha_{t+1}(i) a_{i1}] b_1(O_{t+1})$$
i=N
i=N-1
i=5
i=4
i=3
i=2
i=1
t=1
t=2
t=3
t=4
t=5
t=6
t=7
t=T-1
t=T

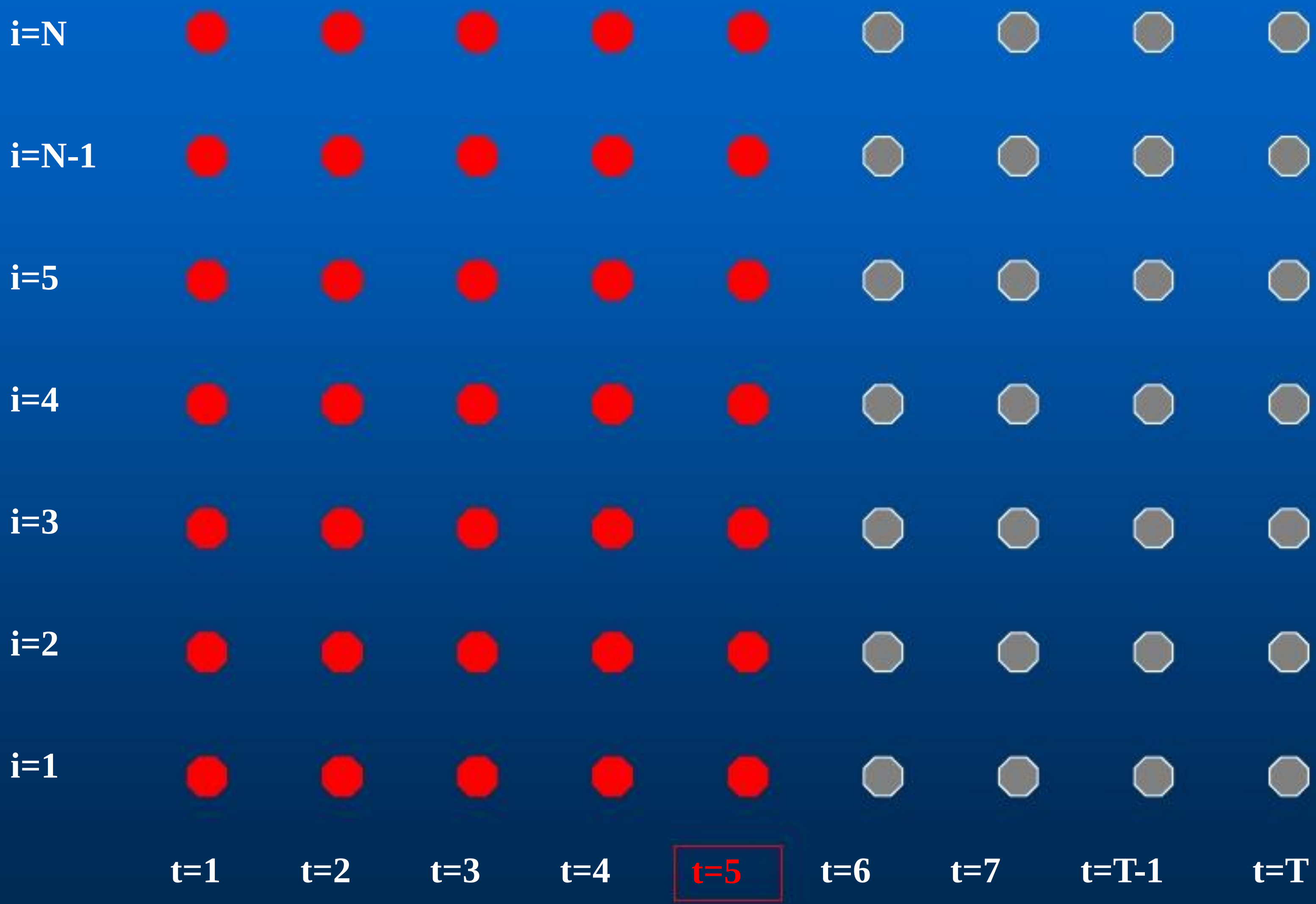
前向算法过程演示



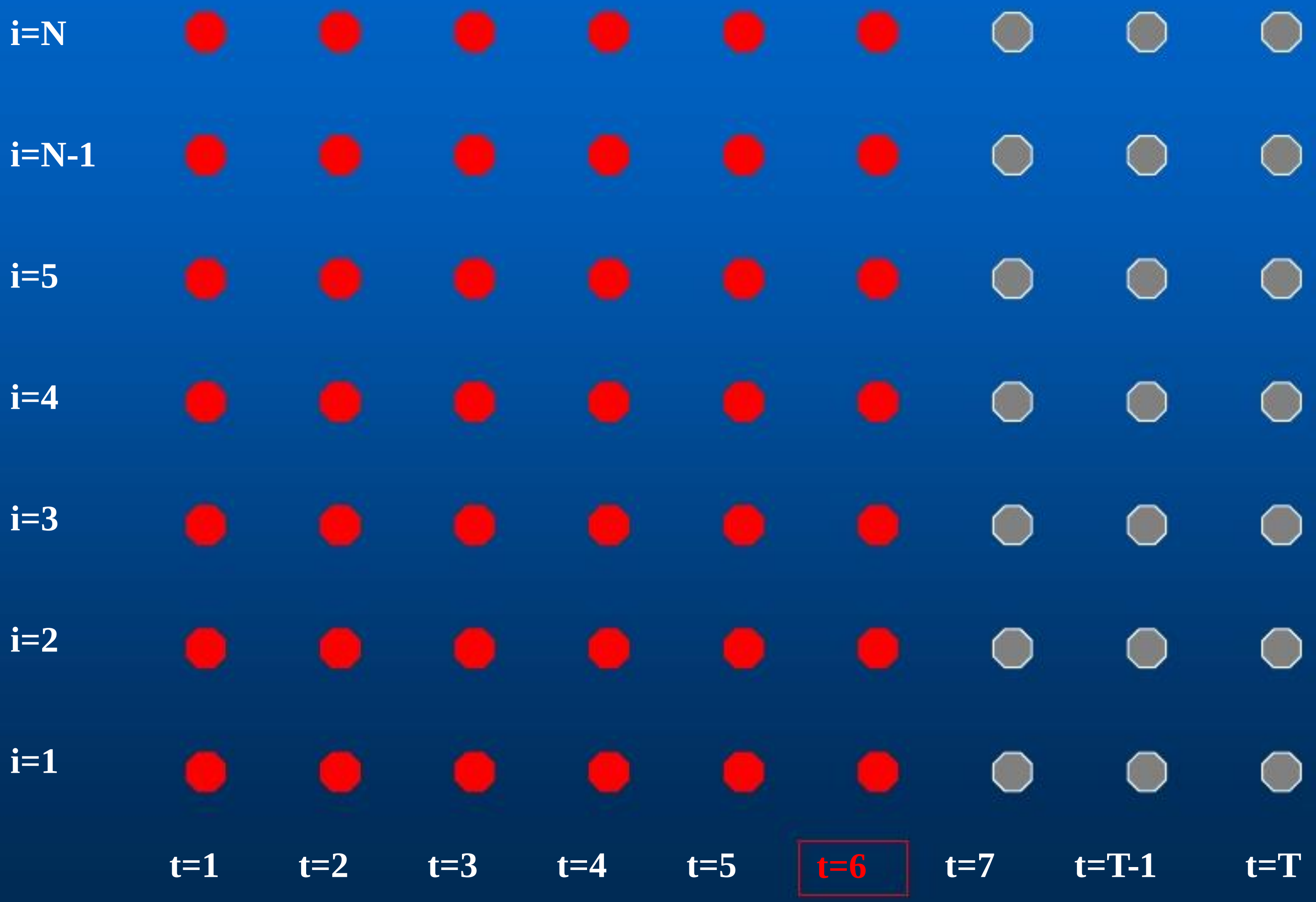
前向算法过程演示



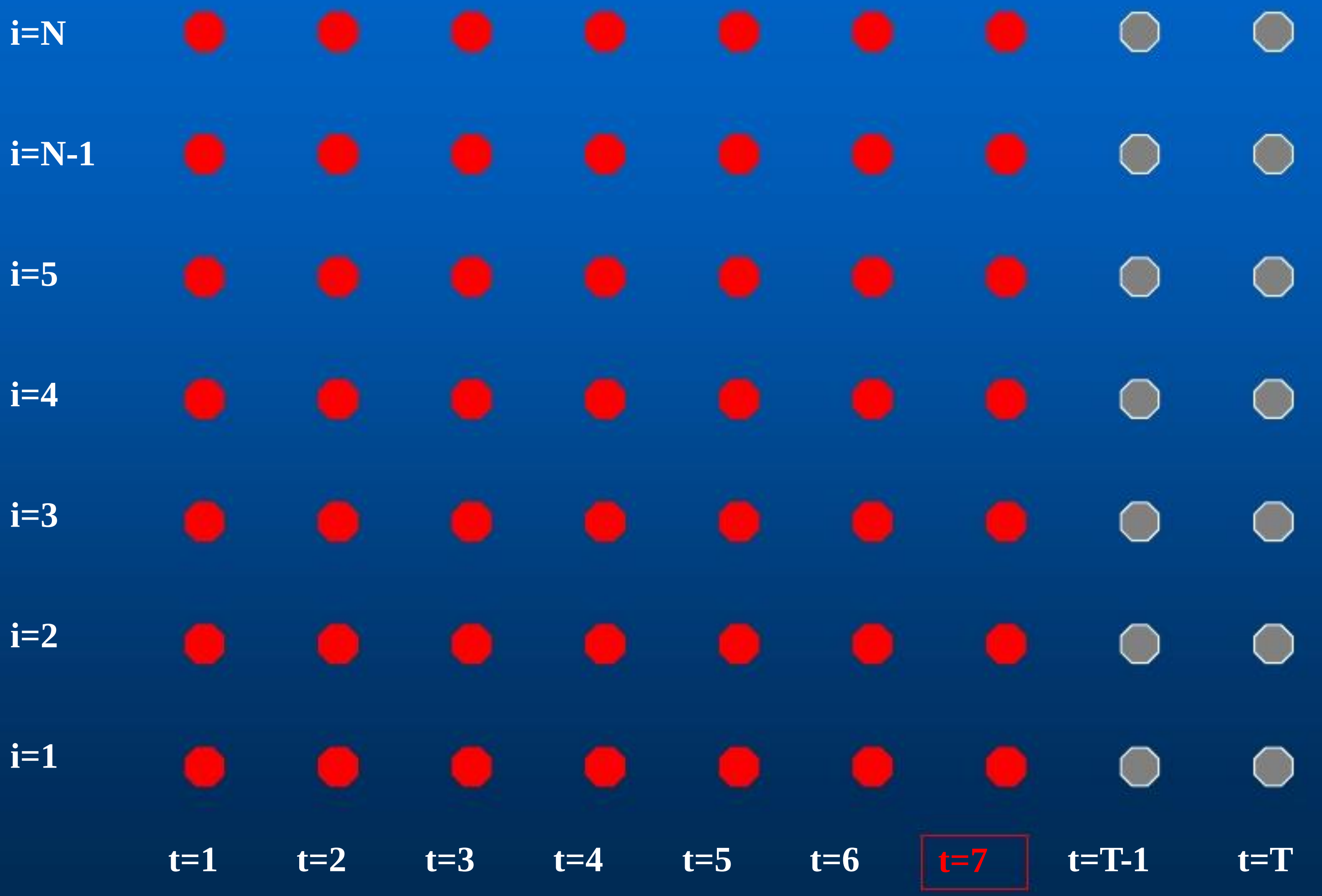
前向算法过程演示



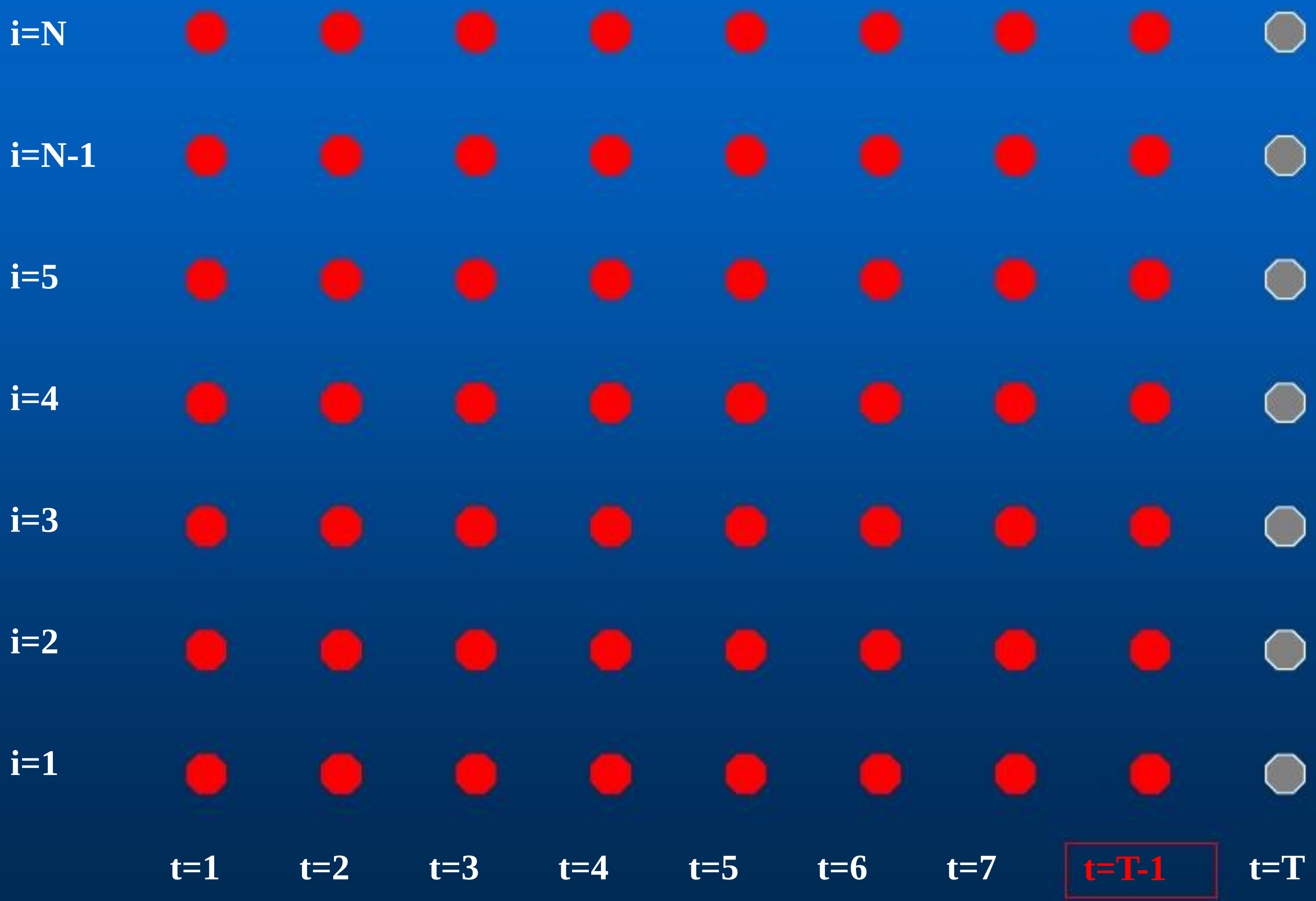
前向算法过程演示



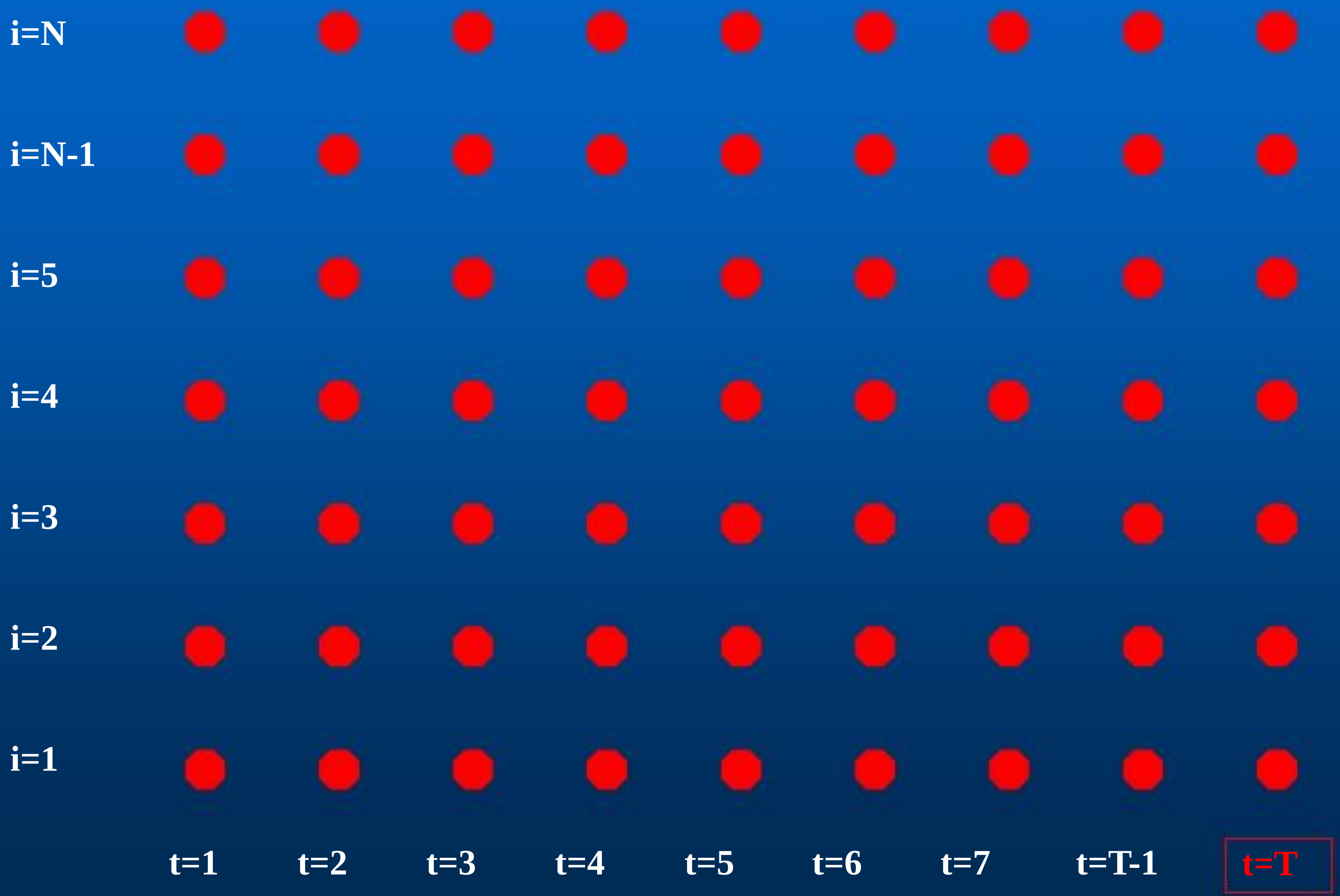
前向算法过程演示



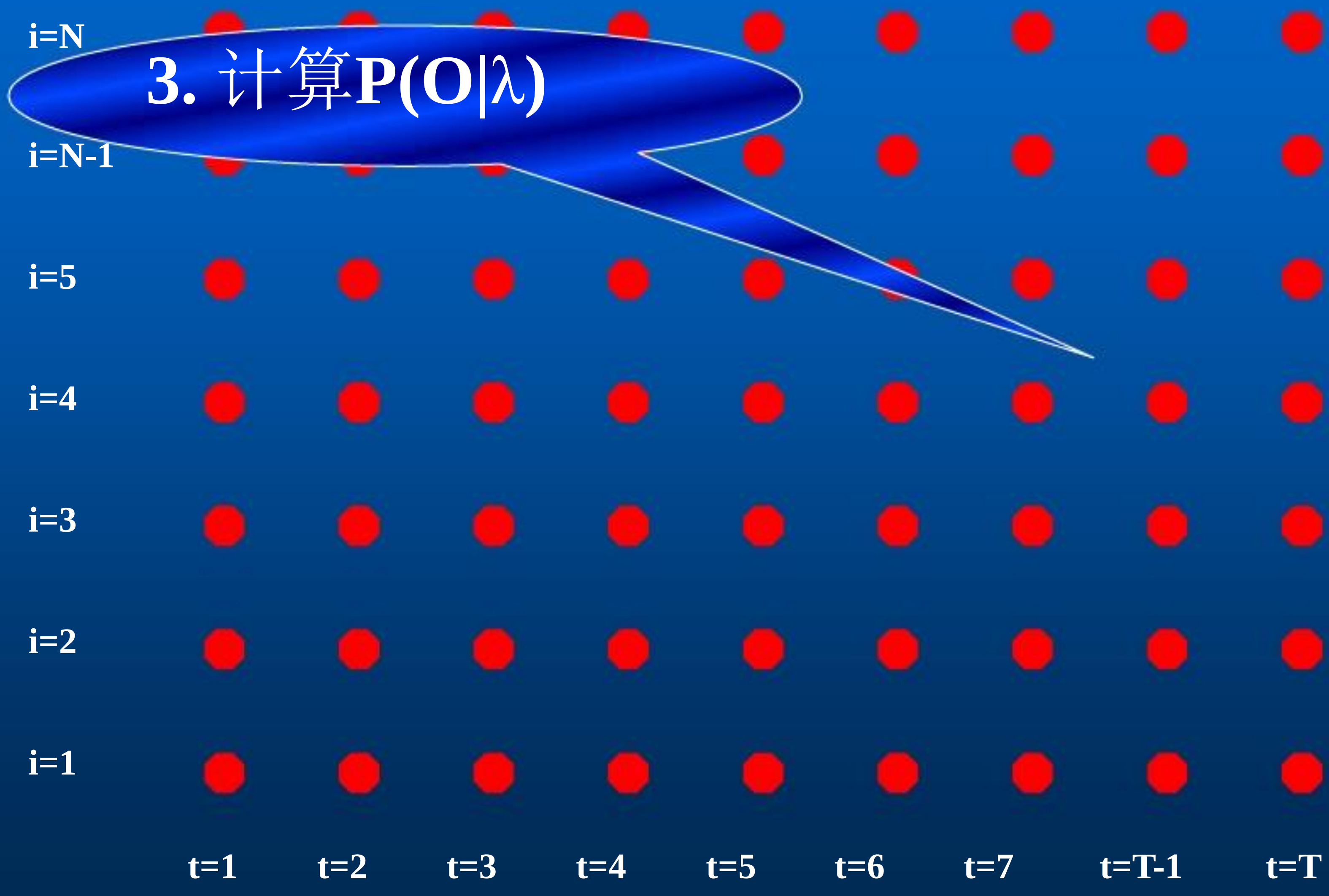
前向算法过程演示



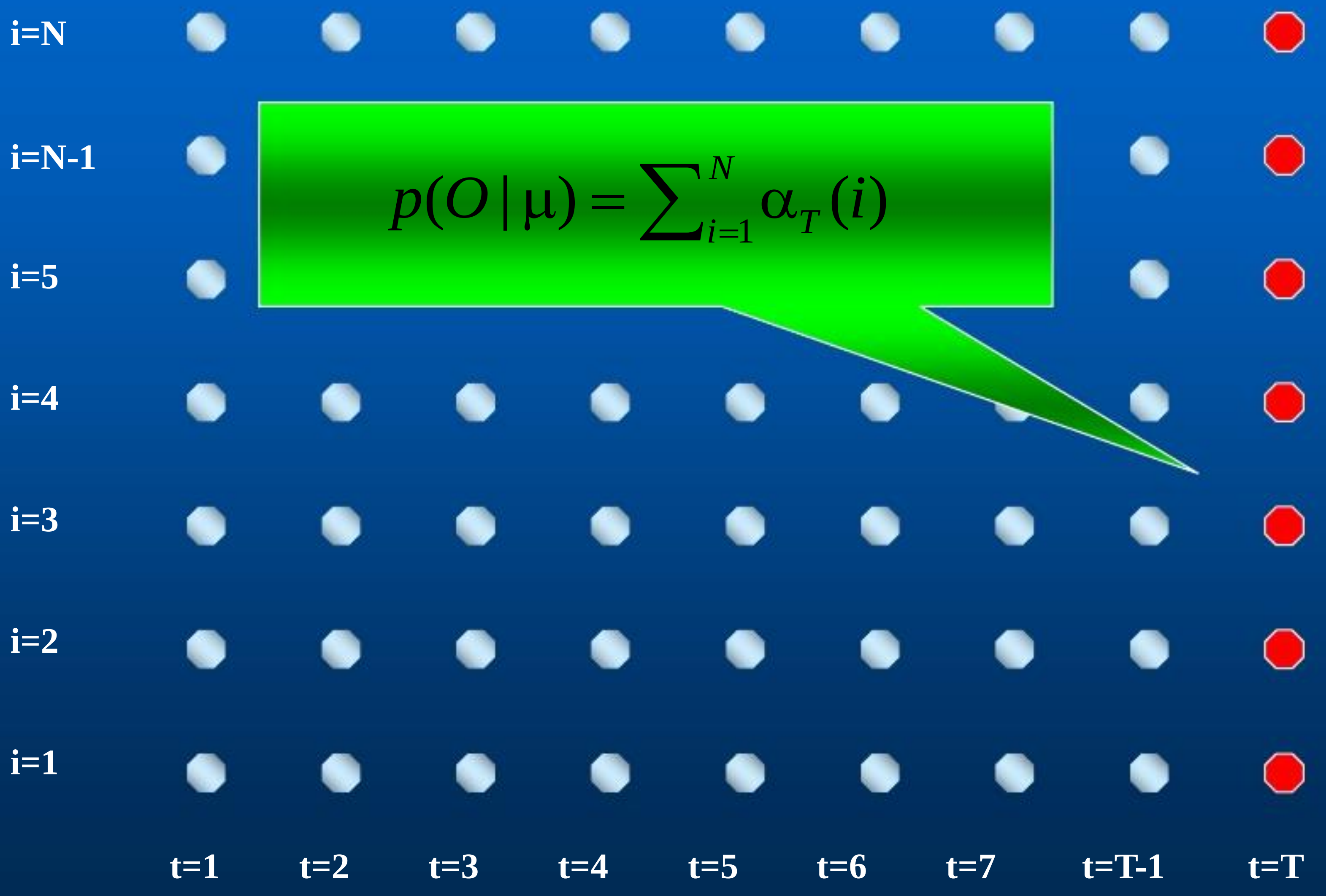
前向算法过程演示



前向算法过程演示



前向算法过程演示



$$p(O|\mu)=\sum_{i=1}^{N}\alpha_T(i)$$