人工智能导论

马尔可夫模型

郭兰哲

南京大学 智能科学与技术学院

Homepage: www.lamda.nju.edu.cn/guolz

Email: guolz@nju.edu.cn

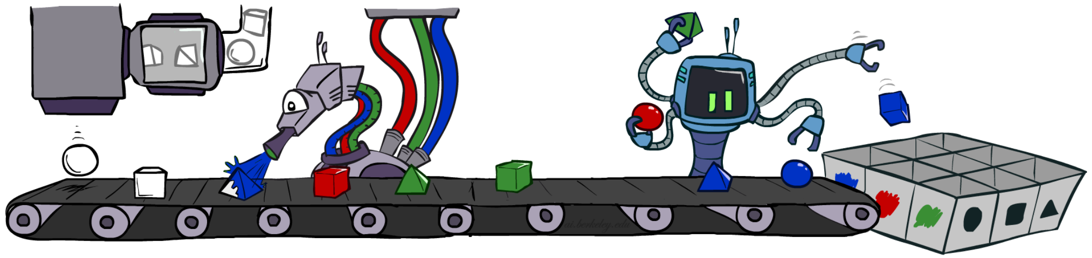
Probability Recap

- Conditional probability $$P(x|y) = \frac{P(x,y)}{P(y)}$$
- Product rule $$P(x,y) = P(x|y)P(y)$$
- Chain rule $$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1,\ldots,X_{i-1}) \end{aligned}$$
- X, Y independent if and only if: $\forall x, y : P(x,y) = P(x)P(y)$
- X and Y are conditionally independent given Z if and only if:

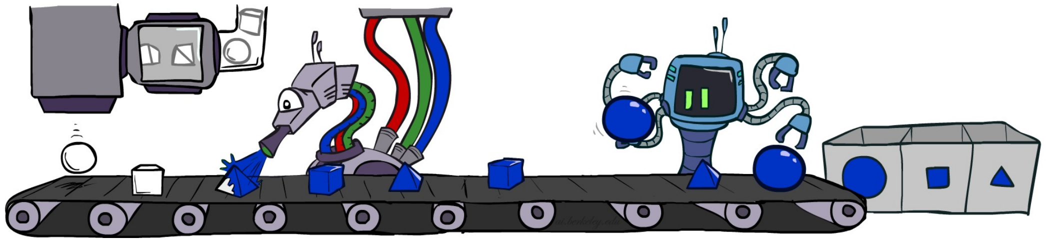
$$\forall x, y, z : P(x,y|z) = P(x|z)P(y|z) \qquad X \perp\!\!\!\perp Y | Z$$

Sampling Methods in Bayes Net

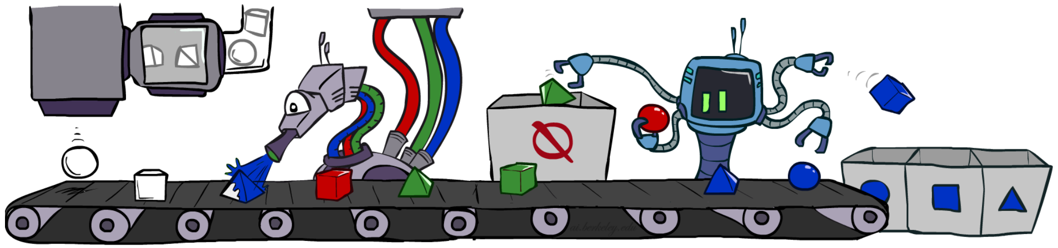
- Prior Sampling P :
  - 基于分布P(x_1,…,x_n) 生成样本



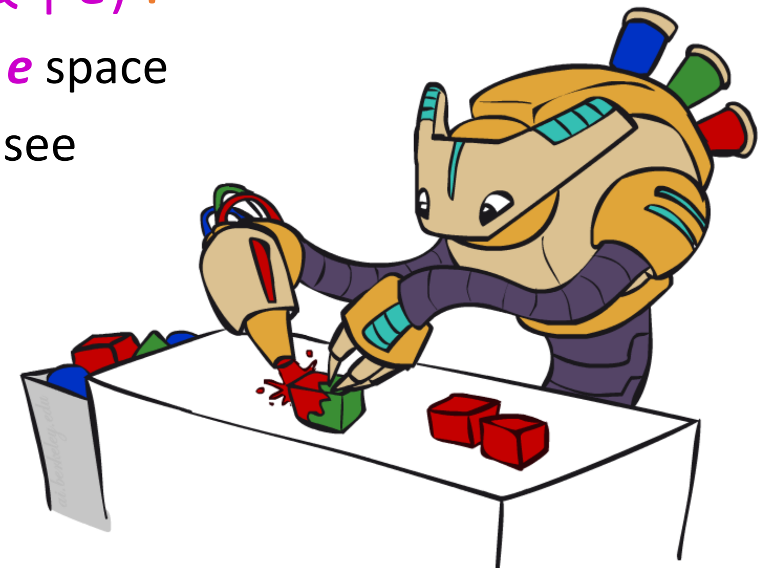
- Likelihood Weighting P(Q | **e**) :
  - 固定证据变量，对样本进行加权



- Rejection Sampling *P*(*Q* | ***e***) :
  - 拒绝掉不符合证据 ***e*** 的样本



- Gibbs sampling *P*(*Q* | ***e***) :
  - Wander around in ***e*** space
  - Average what you see



大纲

- 马尔可夫模型
- 马尔可夫模型中的前向推理
- 马尔可夫模型的平稳分布
- 隐马尔可夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然解释

大纲

- 马尔可夫模型
- 马尔可夫模型中的前向推理
- 马尔可夫模型的平稳分布
- 隐马尔可夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然解释

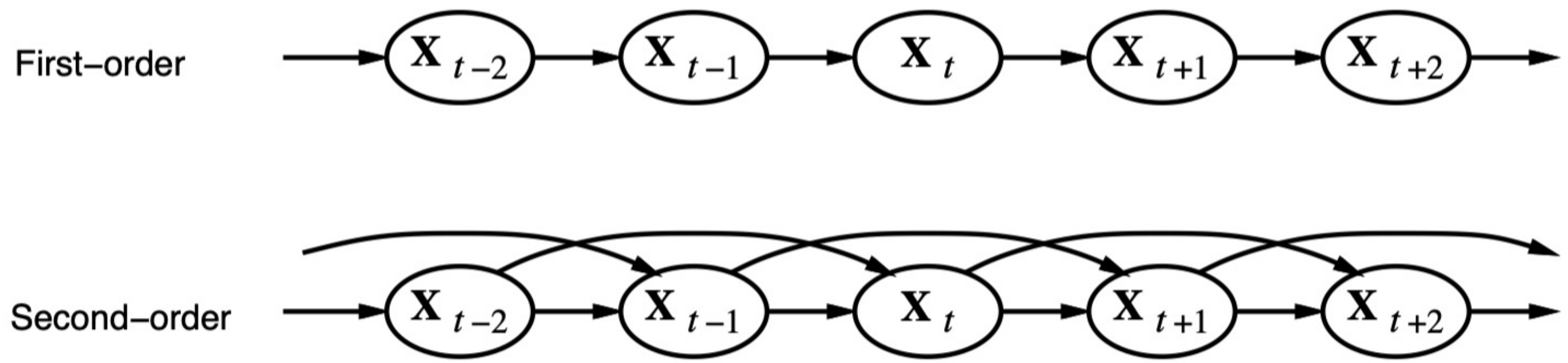
时间与不确定性

- 现实世界是动态变化的，需要对变化的世界进行建模
  - 医疗监护
  - 语音识别
  - 机器人定位
  - 气候监测
  - 搜索引擎
  - …
- 引入“时间”：解释现在、理解过去、预测未来

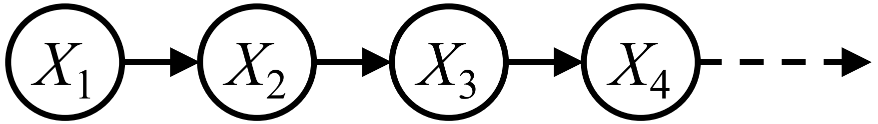
马尔可夫模型(Markov Model)

- 状态变量：给定时刻t，随机变量X的取值$X_1, \cdots, X_t$
- 初始状态概率：随机变量X的先验概率分布
- 转移模型(transition model)：指定世界如何随时间演变，给定过去的状态变量取值之后，确定最新状态的概率分布$P(X_t|X_{1,\cdots t-1})$

- 马尔可夫假设：当前状态只依赖于有限的固定数量的过去状态



马尔可夫模型(Markov Model)

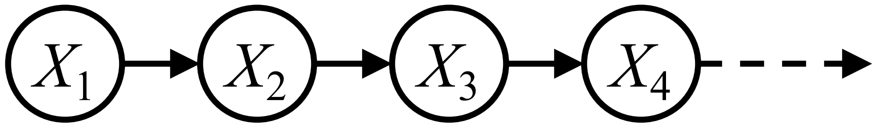


$$P(X_1) \qquad P(X_t|X_{t-1})$$

- 稳态假设(Stationarity assumption)：转移模型不变，即变化的过程是有本身不随时间变化的规律支配的

注意区分静态和稳态

马尔可夫模型的联合概率分布



$$P(X_1) \qquad\qquad P(X_t|X_{t-1})$$

- 联合概率分布：

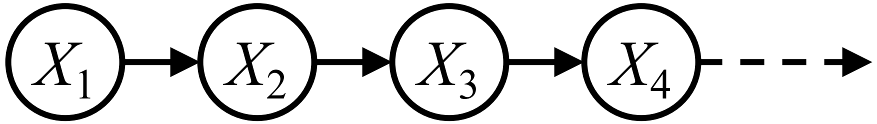
$$P(X_1,X_2,X_3,X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

- 一般化形式：

$$P(X_1,X_2,\cdots,X_{\mathrm{T}}) = P(X_1)P(X_2|X_1)\cdots P(X_T|X_{T-1})$$

$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

马尔可夫模型的条件独立性



- 隐含了什么样的(条件)独立性?

$$\mathrm{X}_t \perp X_1, \cdots, X_{t-2} | X_{t-1}$$

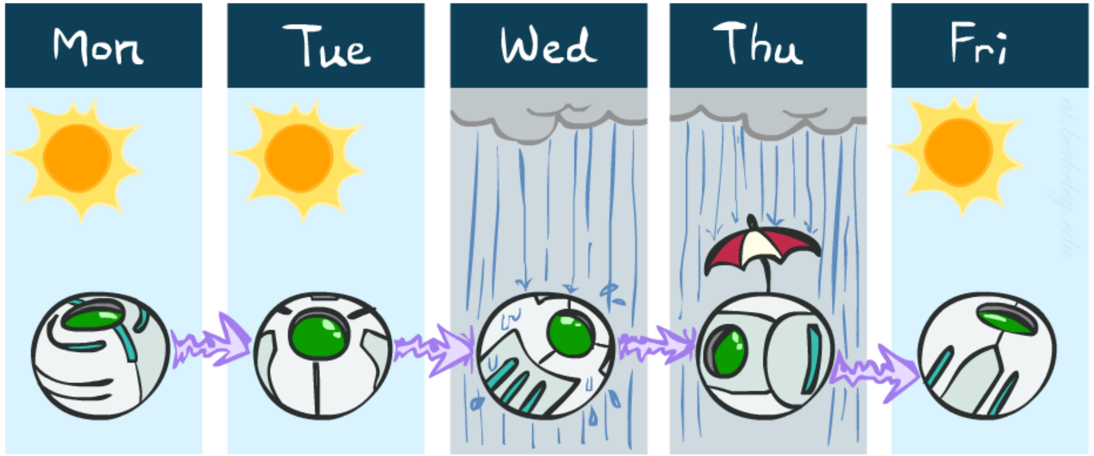
一阶马尔科夫性 -> 马尔科夫链

大纲

- 马尔可夫模型
- 马尔可夫模型中的前向推理
- 马尔可夫模型的平稳分布
- 隐马尔可夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然解释

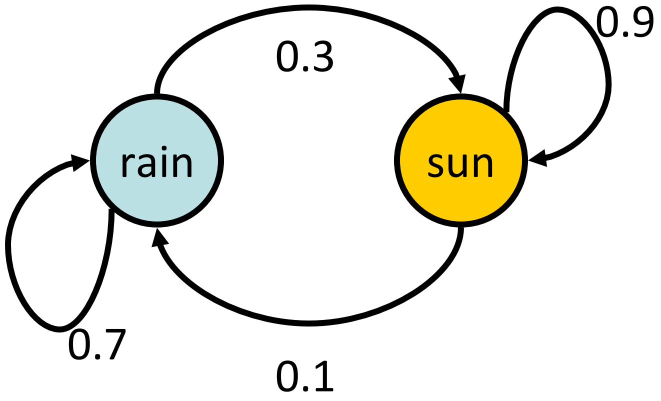
马尔可夫链样例：天气预报

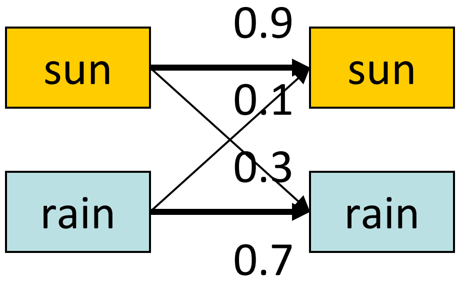
- 状态：$X = \{rain, sun\}$
- 起始分布：$[1.0\ sun, 0.0\ rain]$
- 状态转移：$P(X_t|X_{t-1})$



X_{t-1}	X_t	$P(X_t\|X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

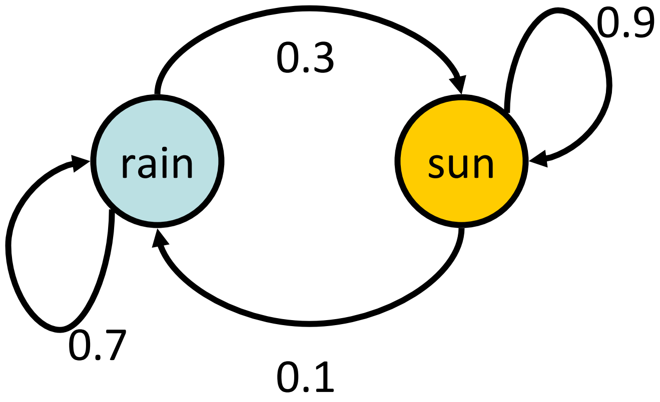
另外两种表示转移模型的方法：概率转移图和轨迹图





马尔可夫链样例：天气预报

- 起始分布: $[1.0\ sun, 0.0\ rain]$



- $P(X_2 = sun)$?

$$P(X_2 = sun) \quad = P(X_2 = sun|X_1 = sun)P(X_1 = sun) + P(X_2 = sun|X_1 = rain)P(X_1 = rain)$$

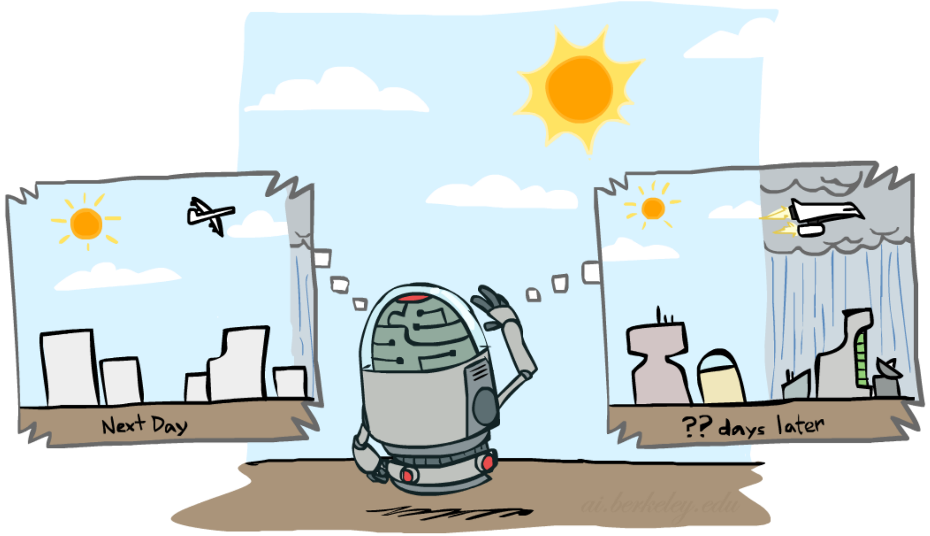
$$= 0.9 * 1.0 + 0.3 * 0.3 = 0.9$$

迷你前向算法(Mini-Forward Algorithm)

- T时刻的$P(X)$？

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \dashrightarrow$$

- $P(X_1)$已知

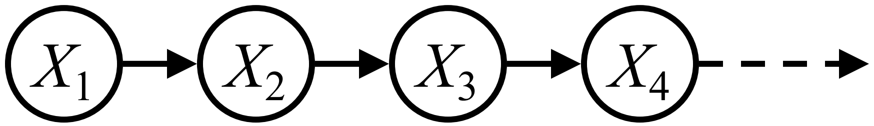


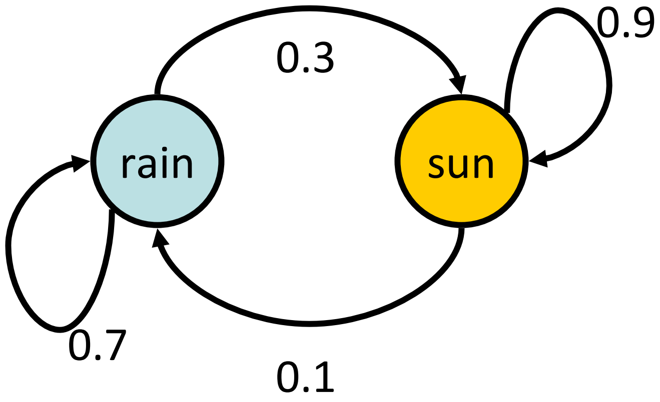
$$P(X_t) = \sum_{x_{t-1}} P(X_t|x_{t-1})P(x_{t-1})$$

前向模拟

迷你前向算法(Mini-Forward Algorithm)

- 第2天的概率分布$P(X_2)$如何计算





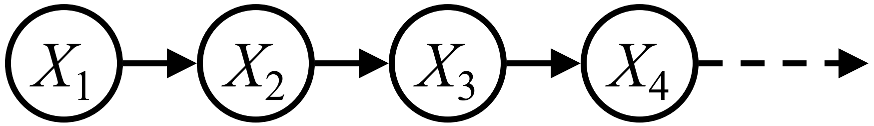
$$P(X_2 = sun) = P(X_2 = sun | X_1 = sun) P(X_1 = sun) + P(X_2 = sun | X_1 = rain) P(X_1 = rain)$$
$$= 0.9 * 1.0 + 0.3 * 0.0 = 0.9$$

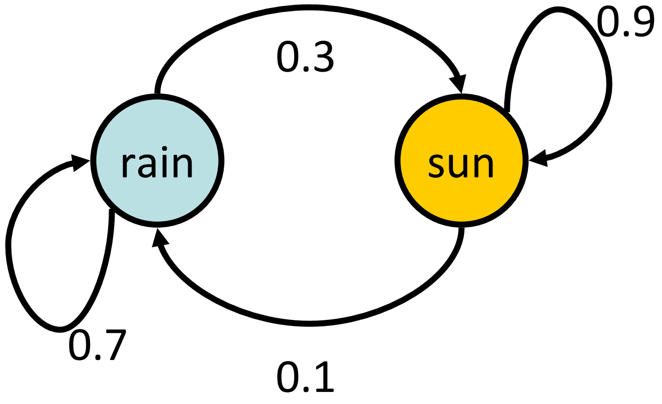
$$P(X_2 = rain) = P(X_2 = rain | X_1 = sun) P(X_1 = sun) + P(X_2 = rain | X_1 = rain) P(X_1 = rain)$$
$$= 0.1 * 1.0 + 0.7 * 0.0 = 0.1$$

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \qquad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle$$
$$\mathrm{P}(X_1) \qquad \mathrm{P}(X_2)$$

迷你前向算法(Mini-Forward Algorithm)

- 第3天的概率分布$P(X_3)$如何计算





$$\begin{aligned}P(X_3 = sun) &= P(X_3 = sun|X_2 = sun)P(X_2 = sun) + \\ &\quad P(X_3 = sun|X_2 = rain)P(X_2 = rain) \\ &= 0.9 * 0.9 + 0.3 * 0.1 = 0.84\end{aligned}$$

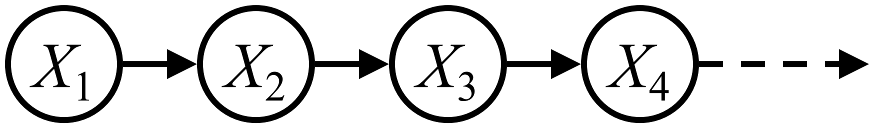
$$\begin{aligned}P(X_3 = rain) &= P(X_3 = rain|X_2 = sun)P(X_2 = sun) + \\ &\quad P(X_3 = rain|X_2 = rain)P(X_2 = rain) \\ &= 0.1 * 0.9 + 0.7 * 0.1 = 0.16\end{aligned}$$

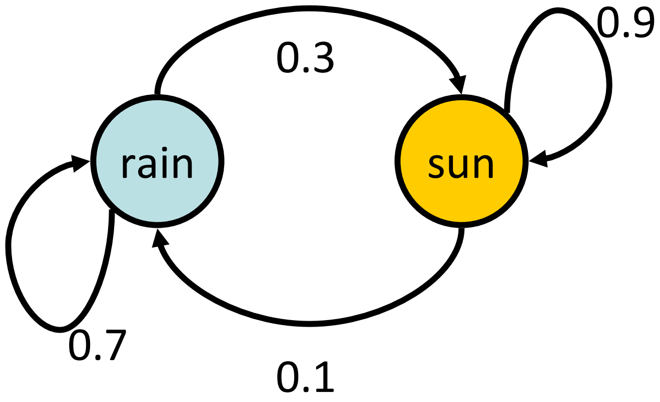
$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle$$

$\mathrm{P}(X_1)$ $\mathrm{P}(X_2)$ $\mathrm{P}(X_3)$

迷你前向算法(Mini-Forward Algorithm)

- 第4天的概率分布$P(X_4)$如何计算





$$P(X_4 = sun) = P(X_4 = sun|X_3 = sun)P(X_3 = sun) + P(X_4 = sun|X_3 = rain)P(X_3 = rain)$$
$$= 0.9 * 0.84 + 0.3 * 0.16 = 0.804$$

$$\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \qquad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle$$

$\mathrm{P}(X_3)$ $\qquad$ $\mathrm{P}(X_4)$

$$P(X_4 = rain) = P(X_4 = rain|X_3 = sun)P(X_3 = sun) + P(X_4 = rain|X_3 = rain)P(X_3 = rain)$$
$$= 0.1 * 0.84 + 0.7 * 0.16 = 0.196$$

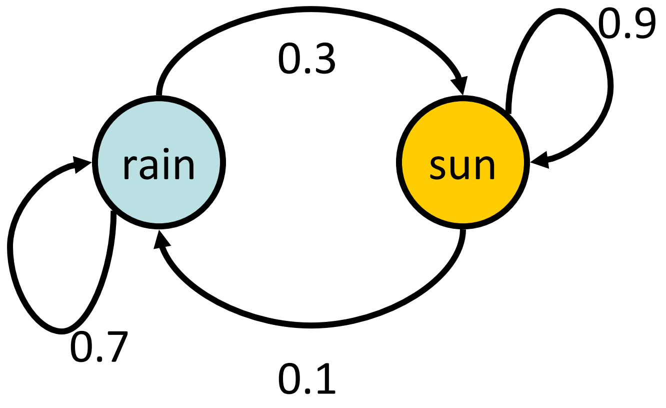
矩阵计算

- 第2天的概率分布$P(X_2)$

$$P(X_2 = sun) \;= P(X_2 = sun|X_1 = sun)P(X_1 = sun) + P(X_2 = sun|X_1 = rain)P(X_1 = rain)$$
$$= 0.9 * 1.0 + 0.3 * 0.0 = 0.9$$

$$P(X_2 = rain) \;= P(X_2 = rain|X_1 = sun)P(X_1 = sun) + P(X_2 = rain|X_1 = rain)P(X_1 = rain)$$
$$= 0.1 * 1.0 + 0.7 * 0.0 = 0.1$$

$$T^{\top} \times \left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle$$



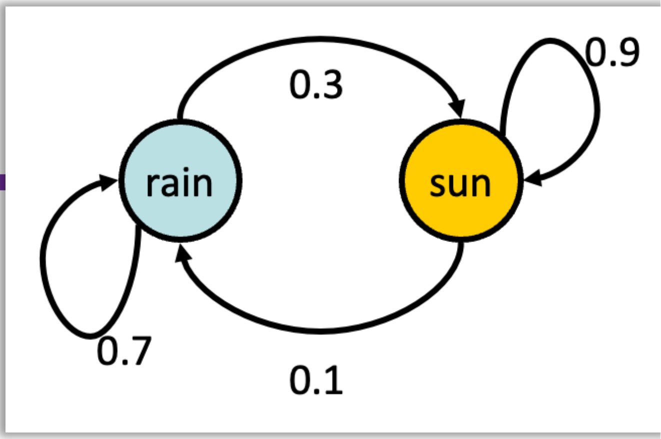
X_{t-1}	$P(X_t \| X_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

$$T_{ij} = P(X_t = j|X_{t-1} = i)$$

大纲

- 马尔可夫模型
- 马尔可夫模型中的前向推理
- 马尔可夫模型的平稳分布
- 隐马尔可夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然解释

Example



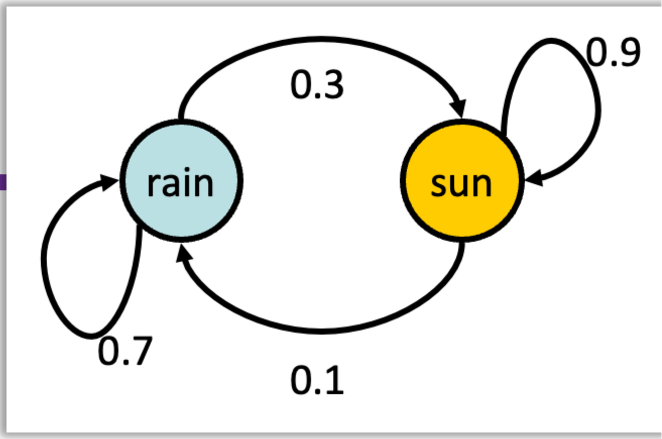
- 起始状态： *sun*

$$\begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix}_{P(X_1)} \quad \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix}_{P(X_2)} \quad \begin{pmatrix} 0.84 \\ 0.16 \end{pmatrix}_{P(X_3)} \quad \begin{pmatrix} 0.804 \\ 0.196 \end{pmatrix}_{P(X_4)} \Rightarrow P(X_\infty)$$

- 起始状态： *rain*

$$\begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix}_{P(X_1)} \quad \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix}_{P(X_2)} \quad \begin{pmatrix} 0.48 \\ 0.52 \end{pmatrix}_{P(X_3)} \quad \begin{pmatrix} 0.588 \\ 0.412 \end{pmatrix}_{P(X_4)} \Rightarrow P(X_\infty)$$

Example



- 起始状态： *sun*

$$\begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix} \quad \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix} \quad \begin{pmatrix} 0.84 \\ 0.16 \end{pmatrix} \quad \begin{pmatrix} 0.804 \\ 0.196 \end{pmatrix} \Rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

$P(X_1)$ $P(X_2)$ $P(X_3)$ $P(X_4)$ $P(X_\infty)$

- 起始状态： *rain*

$$\begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix} \quad \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} \quad \begin{pmatrix} 0.48 \\ 0.52 \end{pmatrix} \quad \begin{pmatrix} 0.588 \\ 0.412 \end{pmatrix} \Rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

$P(X_1)$ $P(X_2)$ $P(X_3)$ $P(X_4)$ $P(X_\infty)$

- 任意初始状态

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} \quad \cdots \quad \Rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

$P(X_1)$ $P(X_\infty)$

平稳分布（Stationary Distribution）

☐ 对于大多数马尔可夫链：

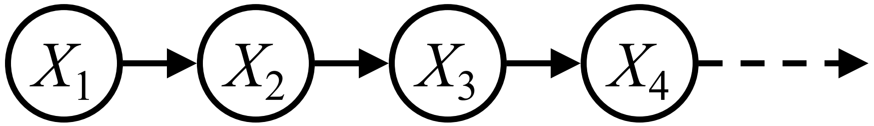
- 起始分布的影响会随时间逐渐减弱
- 最终的分布与起始分布无关

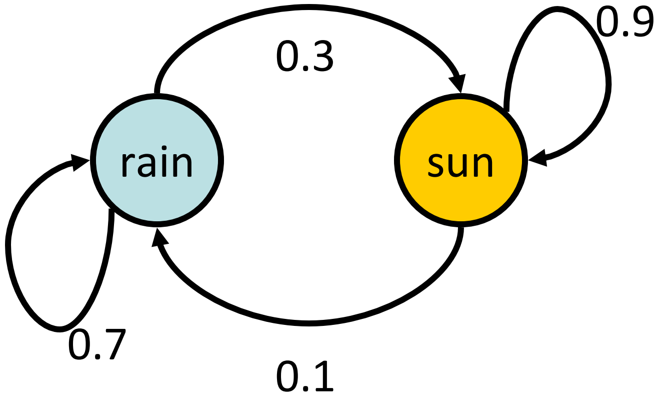
☐ 平稳分布：

- 最终到达的分布P_∞称为平稳分布
- 满足如下条件：

$$P_\infty(X) = P_{\infty+1}(X) = \sum P(X|x)P_\infty(x)$$

平稳分布的计算





$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$

$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

$$P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain)$$
$$P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain)$$

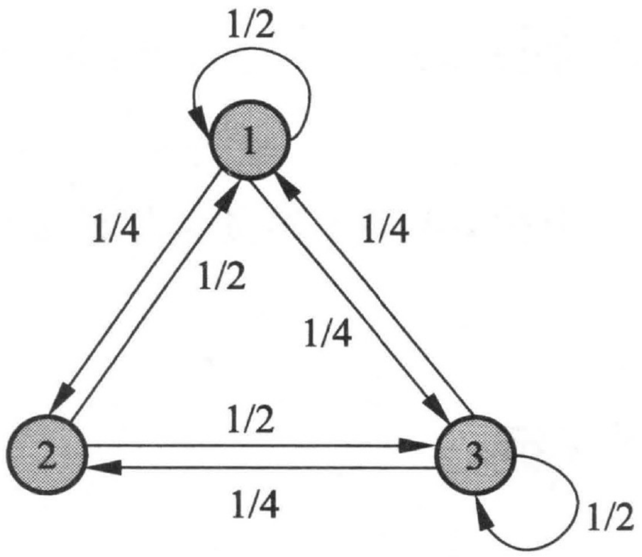
$$P_\infty(sun) = 3P_\infty(rain)$$
$$P_\infty(rain) = 1/3P_\infty(sun)$$

此外 $P_\infty(sun) + P_\infty(rain) = 1$ $\Rightarrow$

$$P_\infty(sun) = 3/4$$
$$P_\infty(rain) = 1/4$$

Example 1

- 给定下列状态转移模型，求对应马尔可夫链的平稳分布



$$x_1 = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{4}x_3$$

$$x_2 = \frac{1}{4}x_1 + \frac{1}{4}x_3$$

$$x_3 = \frac{1}{4}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3$$

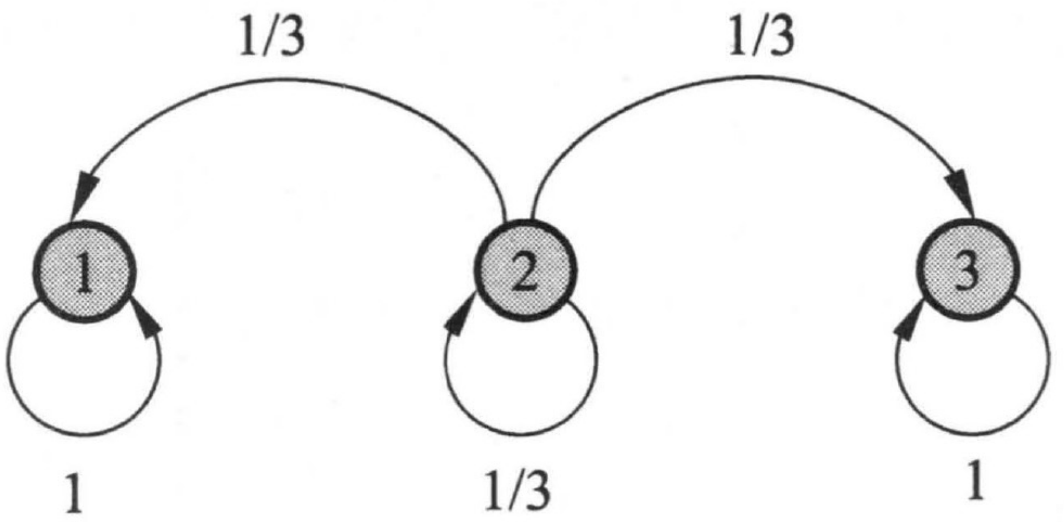
$$x_1 + x_2 + x_3 = 1$$

$$x_i \geqslant 0, \quad i = 1, 2, 3$$

解方程组可得： $(2/5 \quad 1/5 \quad 2/5)^{\mathrm{T}}$

Example 2

- 给定下列状态转移模型，求对应马尔可夫链的平稳分布



马尔可夫链可能存在一个平稳状态，也可能存在多个平稳状态，或者不存在平稳状态

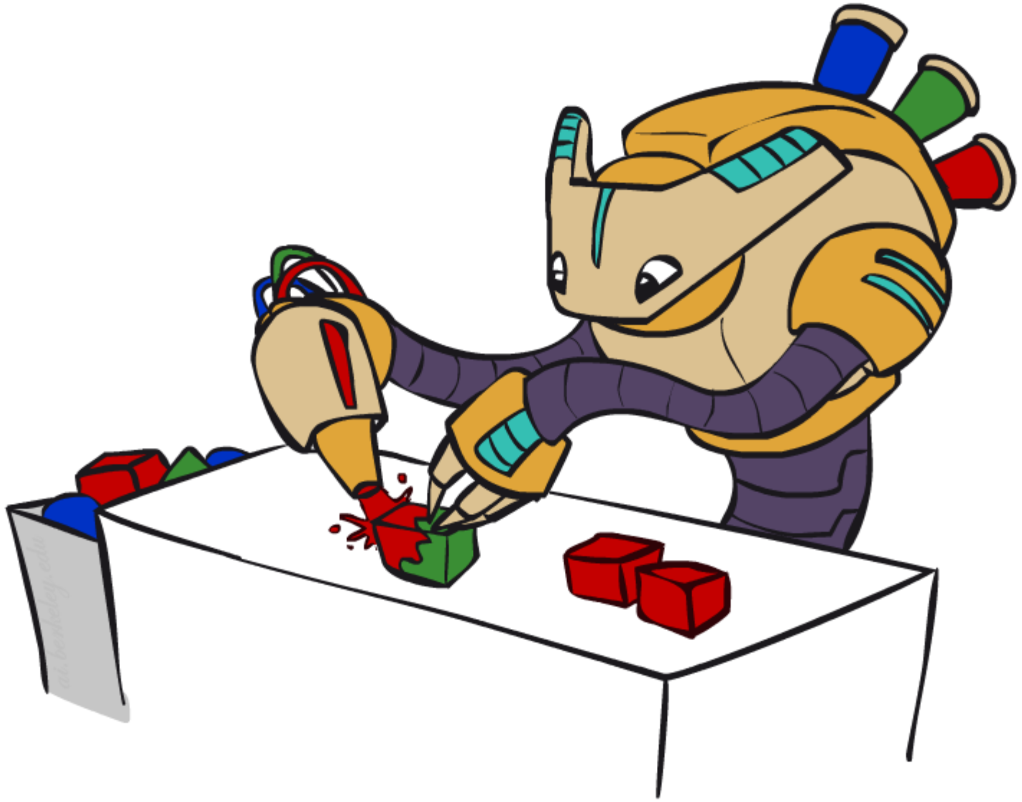
平稳分布的应用： Gibbs Sampling

- 状态转移：
  - With probability 1/n resample variable X_j according to

$$P(X_j|x_1,x_2,\cdots,x_{j-1},x_{j+1},\cdots x_n,e_1,\cdots,e_m)$$

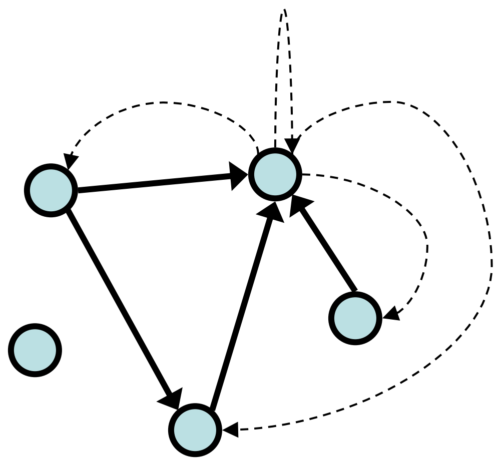
- 平稳分布：
  - Conditional distribution

$$P(X_1,X_2,\cdots,X_n|e_1,\cdots,e_m)$$



平稳分布的应用：网页排序PageRank算法

- 假设互联网是一个有向图，在其基础上定义一阶马尔科夫链，表示网页浏览者在互联网上随机浏览网页的过程
- 转移概率：
  - 以1-C的概率，随机跟随页面中的外链跳转（实线部分）。模拟通过点击外链完成页面跳转
  - 以C的概率，均匀且随机的跳转到任意其它页面（虚线部分）模拟通过地址栏输入网页地址完成访问，同时保障状态图的全连通性
- 直观上，如果指向一个网页的超链接越多，随机跳转到该网页的概率也就越高，该网页的PageRank 值就越高，这个网页也就越重要
- 1996 年由Page 和Brin提出，用于谷歌搜索引擎的网页排序



平稳分布的应用：扩散模型(Diffusion Model)

- Text-based image/art generation
  - Dall-E 2, Imagen, Stable Diffusion, MidJourney…

[Christian Beltrami/MidJourney]

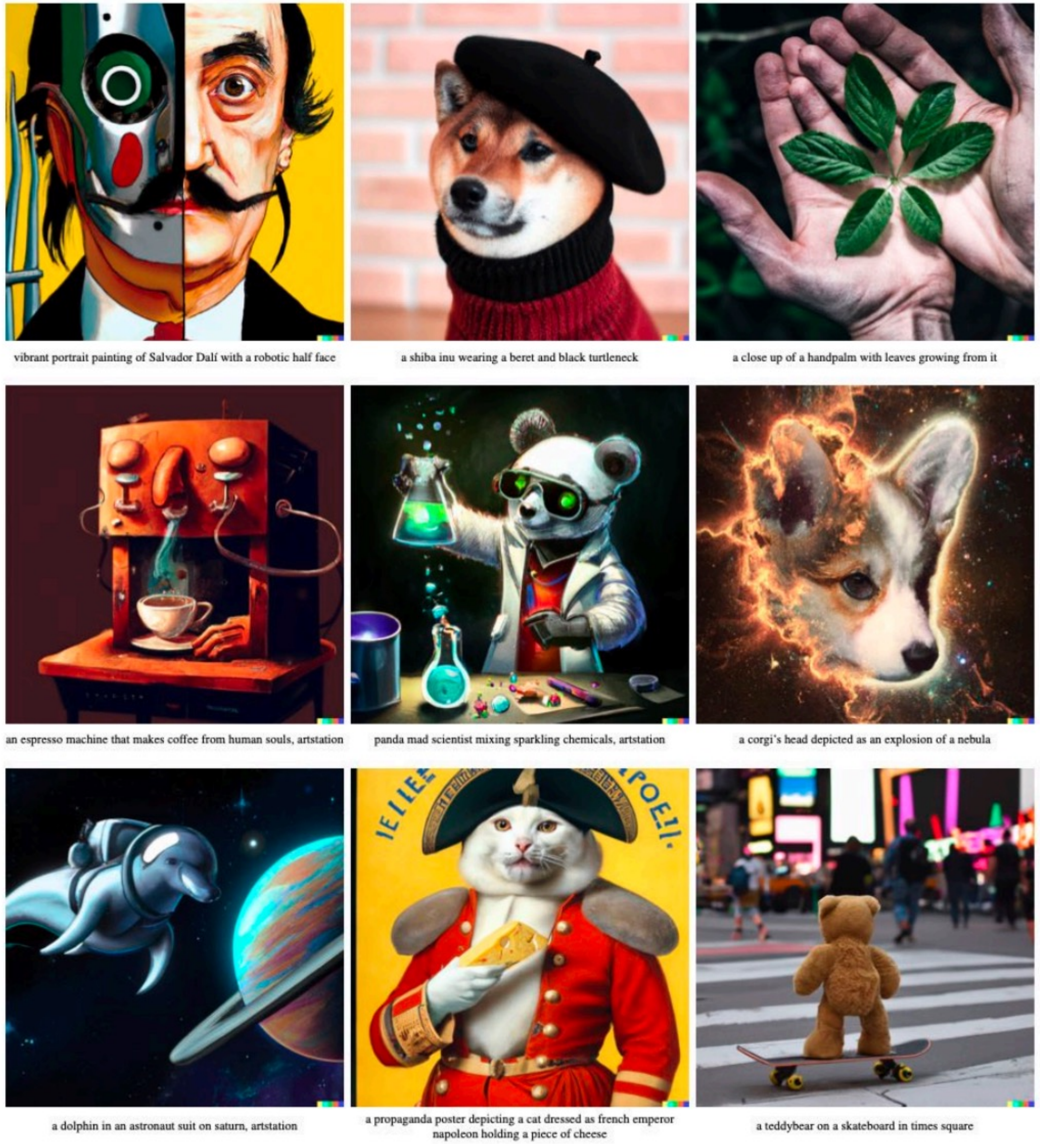
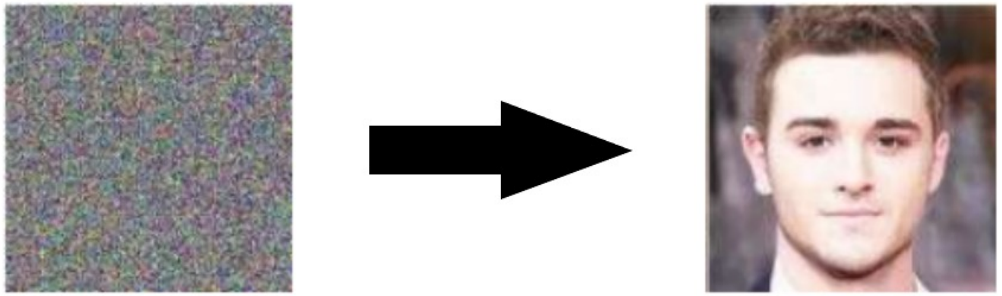
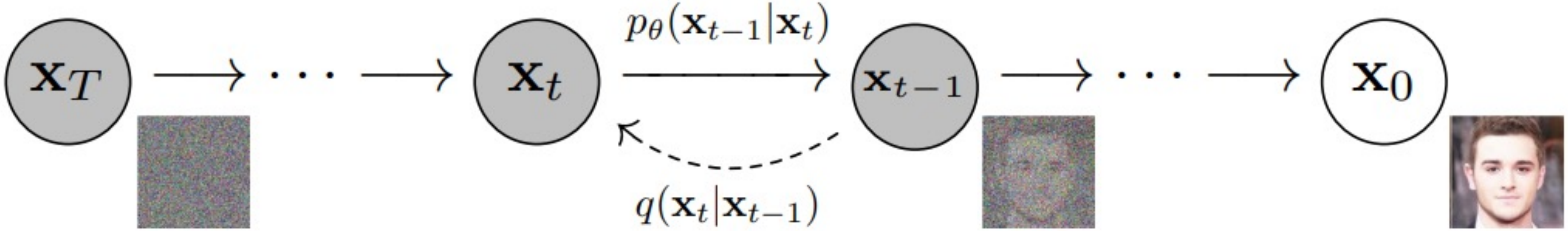


Figure 1: Selected 1024×1024 samples from a production version of our model. [OpenAI]

平稳分布的应用： Diffusion Model



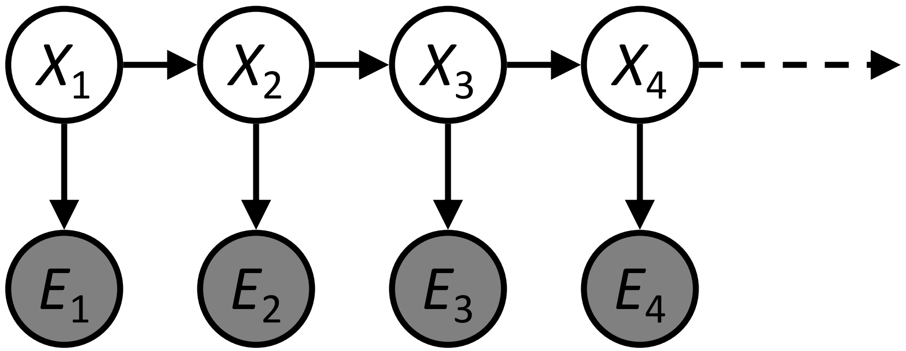


大纲

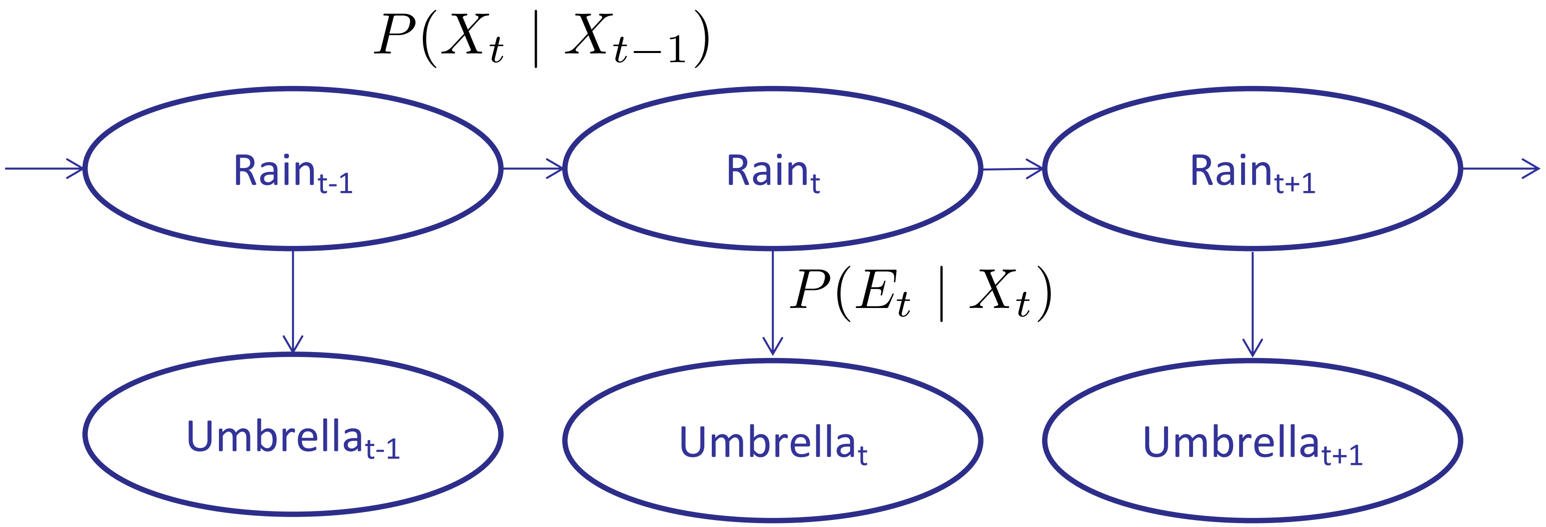
- 马尔可夫模型
- 马尔可夫模型中的前向推理
- 马尔可夫模型的平稳分布
- 隐马尔可夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然解释

隐马尔可夫模型(Hidden Markov Models)

- 隐马尔可夫模型(HMMs) :
  - 由一个隐藏的马尔可夫链随机生成不可观测的状态序列(state sequence)，再由各个状态生成一个观测从而产生观测序列(observation sequence)的过程



Example： Weather HMM



- 初始概率： $P(X_1)$
- 转移概率： $P(X_t|X_{t-1})$
- 观测(发射)概率： $P(E_t|X_t)$

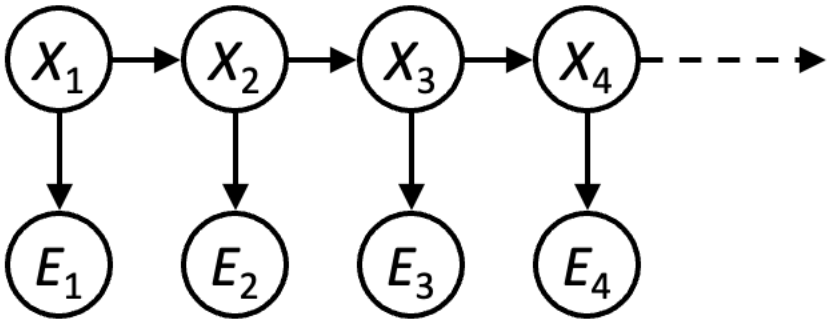
HMM的三要素

R_{t-1}	$P(R_t \mid R_{t-1})$
+r	0.7
-r	0.3

R_t	$P(U_t \mid R_t)$
+r	0.9
-r	0.2

隐马尔可夫模型的联合概率

- HMM的两条独立性假设：
  - 马尔可夫假设：隐藏的马尔可夫链在任意时刻t的状态只依赖其前一时刻的状态
  - 观测独立性假设：任意时刻的观测只依赖于该时刻的马尔可夫链状态



$$P(X_1, E_1, \cdots, X_T, E_T) = P(X_1)P(E_1|X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})P(E_t|X_t)$$

隐马尔可夫模型的条件独立性

$$P(X_1,E_1,\cdots,X_T,E_T) = P(X_1)P(E_1|X_1)\prod_{t=2}^{T}P(X_t|X_{t-1})P(E_t|X_t)$$

- 对于所有时刻 t，给定前一时刻隐状态，当前隐状态独立于所有过去的隐状态和观测状态

$$\mathrm{X}_t \perp X_1,E_1\cdots,X_{t-2},E_{t-2},E_{t-1}|X_{t-1}$$

- 给定当前时刻隐状态，当前时刻的可见状态独立于过去所有的隐状态和可见状态

$$E_t \perp X_1,E_1\cdots,X_{t-2},E_{t-2},E_{t-1}|X_t$$

隐马尔可夫模型的应用

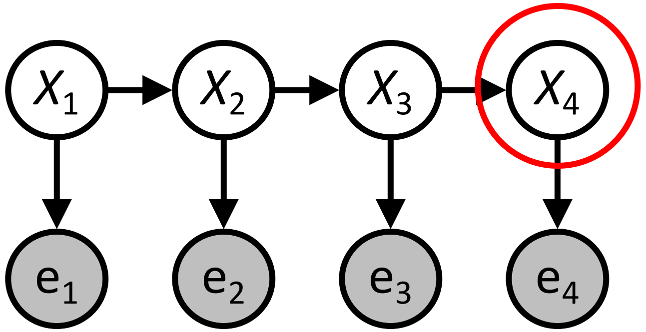
- 语音识别
  - 观测序列为声学信号
  - 状态序列为语言文字
- 机器翻译
  - 观测序列为源语言文字
  - 状态序列为目标语言文字
- 机器人跟踪定位
  - 观测序列为连续的传感器信号
  - 状态序列为机器人所处的位置

大纲

- 马尔可夫模型
- 马尔可夫模型中的前向推理
- 马尔可夫模型的平稳分布
- 隐马尔可夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然解释

隐马尔可夫模型的概率推理：过滤任务

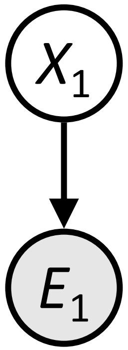
Filtering: $P(X_t | e_{1:t})$



- 观测序列：$e_{1:t} = e_1, \cdots, e_t$
- 隐变量的概率分布：$B(X_t) = P(X_t | e_{1:t})$

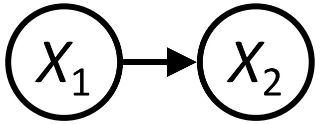
给定当前所有观测序列，计算最新时刻隐状态的后验概率分布

过滤任务：基本过程



$$P(X_1|e_1)$$

$$\begin{aligned} P(x_1|e_1) &= P(x_1, e_1)/P(e_1) \\ &\propto_{X_1} P(x_1, e_1) \\ &= P(x_1)P(e_1|x_1) \end{aligned}$$



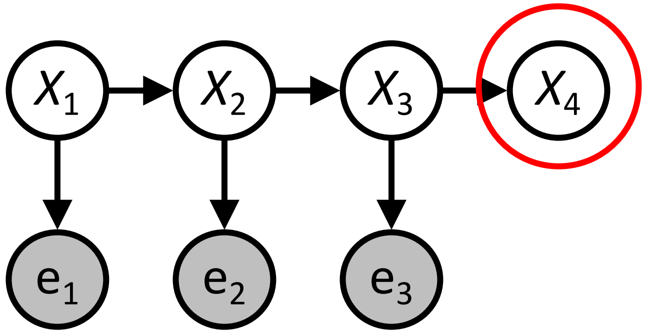
$$P(X_2)$$

$$\begin{aligned} P(x_2) &= \sum_{x_1} P(x_1, x_2) \\ &= \sum_{x_1} P(x_1)P(x_2|x_1) \end{aligned}$$

时间推移

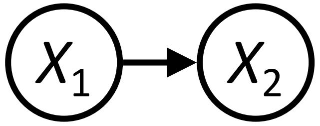
- 假设已知 t 时刻的状态分布：

$$B(X_t) = P(X_t|e_{1:t})$$



- 如何计算： $P(X_{t+1}|e_{1:t})$

$$\begin{aligned}
P(X_{t+1}|e_{1:t}) &= \sum_{x_t} P(X_{t+1}, x_t|e_{1:t}) \\
&= \sum_{x_t} P(X_{t+1}|x_t, e_{1:t})\, P(x_t|e_{1:t}) \\
&= \sum_{x_t} P(X_{t+1}|x_t)\, P(x_t|e_{1:t})
\end{aligned}$$



$$\begin{aligned}
P(x_2) &= \sum_{x_1} P(x_1, x_2) \\
&= \sum_{x_1} P(x_1)P(x_2|x_1)
\end{aligned}$$

- $B'(X_{t+1}) = \sum_{x_t} P(X_{t+1}|x_t)\, B(X_t)$

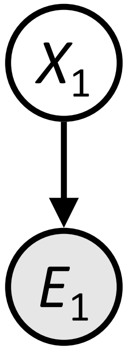
观测过程

- 假设已知 $t+1$ 时刻的状态分布：

$$B'(X_{t+1}) = P(X_{t+1}|e_{1:t})$$

- 如何计算： $B(X_{t+1}) = P(X_{t+1}|e_{1:t+1})$

$$\begin{aligned}P(X_{t+1}|e_{1:t+1}) &= P(X_{t+1}, e_{t+1}|e_{1:t})/P(e_{t+1}|e_{1:t}) \\ &\propto_{X_{t+1}} P(X_{t+1}, e_{t+1}|e_{1:t}) \\ &= P(e_{t+1}|e_{1:t}, X_{t+1})P(X_{t+1}|e_{1:t}) \\ &= P(e_{t+1}|X_{t+1})P(X_{t+1}|e_{1:t})\end{aligned}$$



$$P(X_1|e_1)$$

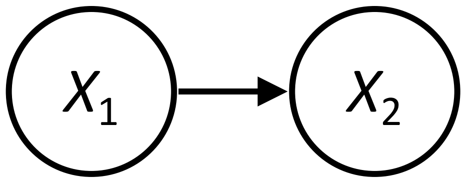
$$\begin{aligned}P(x_1|e_1) &= P(x_1, e_1)/P(e_1) \\ &\propto_{X_1} P(x_1, e_1) \\ &= P(x_1)P(e_1|x_1)\end{aligned}$$

- $B(X_{t+1}) = P(e_{t+1}|X_{t+1})B'(X_{t+1})$

在线更新

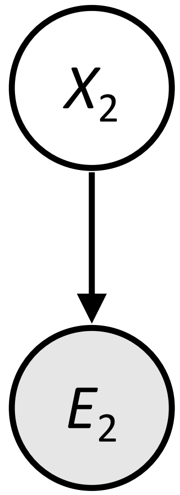
- 每一时刻，从当前状态分布出发：$B(X_t) = P(X_t|e_{1:t})$

- 时间推移： $B'(X_{t+1}) = P(X_{t+1}|e_{1:t})$



$$P(X_{t+1}|e_{1:t}) = \sum_{x_t} P(X_{t+1}|x_t)P(X_t|e_{1:t})$$

- 观测过程更新： $B(X_{t+1}) = P(X_{t+1}|e_{1:t+1})$



$$P(X_{t+1}|e_{1:t+1}) = P(e_{t+1}|X_{t+1})P(X_{t+1}|e_{1:t})$$

前向算法

$$P(X_{t+1}|e_{1:t}) = \sum_{x_t} P(X_{t+1}|x_t)P(X_t|e_{1:t})$$

$$P(X_{t+1}|e_{1:t+1}) = P(e_{t+1}|X_{t+1})P(X_{t+1}|e_{1:t})$$

$$P(X_{t+1}|e_{1:t+1}) = P(e_{t+1}|X_{t+1})\sum_{x_t} P(X_{t+1}|x_t)P(X_t|e_{1:t})$$

$$\boldsymbol{f}_{1:t+1} = \text{FORWARD}(\boldsymbol{f}_{1:t}, e_{t+1})$$

简化的矩阵算法

- $P(X_{t+1}|e_{1:t+1}) = P(e_{t+1}|X_{t+1})\sum_{x_t} P(X_{t+1}|x_t)P(X_t|e_{1:t})$

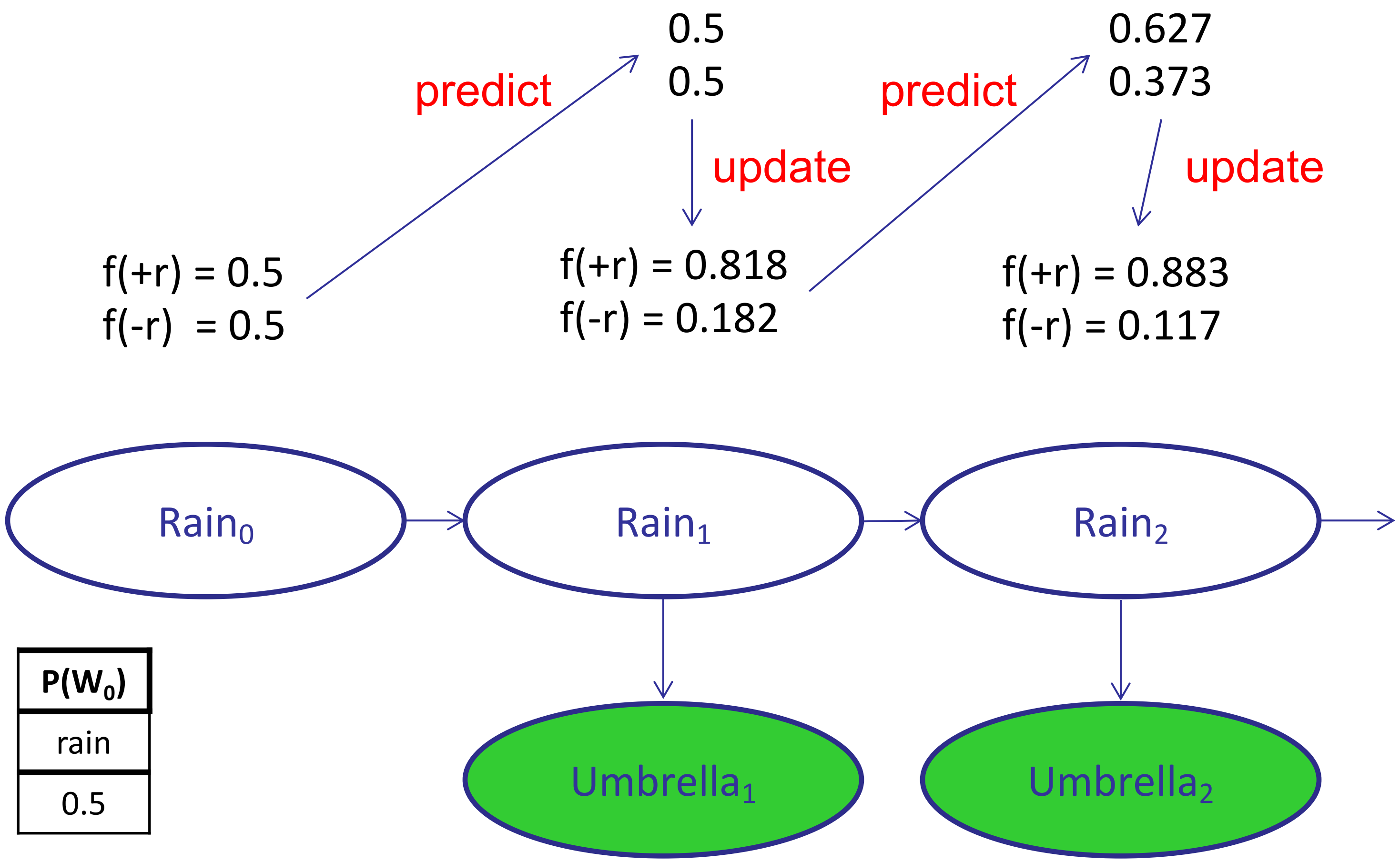
- 转移矩阵T，观测矩阵O
  - $T_{ij} = P(X_{t+1} = j|X_t = i)$
  - $O_{ii} = P(e_t|X_t = i), \mathrm{U}_1 = \mathrm{true}, \mathrm{O}_1 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.2 \end{pmatrix}$

- $f_{1:t+1} = O_{t+1}T^T f_{1:t}$

X_{t-1}	$P(X_t\|X_{t-1})$	
	r	-r
+r	0.7	0.3
-r	0.3	0.7

W_t	$P(U_t\|W_t)$	
	true	false
+r	0.9	0.1
-r	0.2	0.8

Example : Weather HMM



P(W$_0$)
rain
0.5

X_{t-1}	$P(X_t \mid X_{t-1})$	
	r	-r
+r	0.7	0.3
-r	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
+r	0.9	0.1
-r	0.2	0.8

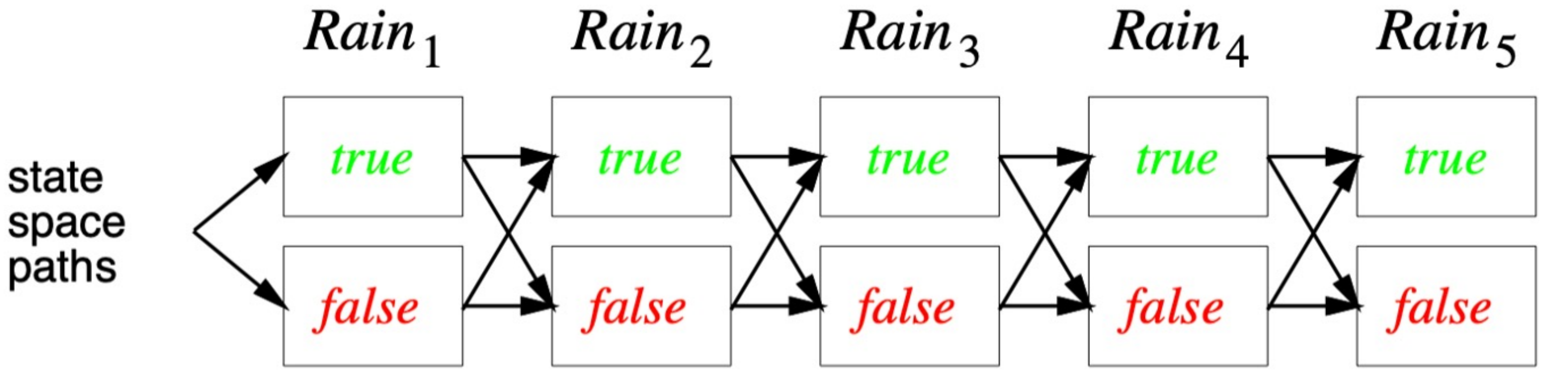
大纲

- 马尔可夫模型
- 马尔可夫模型中的前向推理
- 马尔可夫模型的平稳分布
- 隐马尔可夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然解释

最可能解释： $\text{argmax}_{X_{1\ldots t}} P(X_{1:t}|e_{1,\ldots t})$

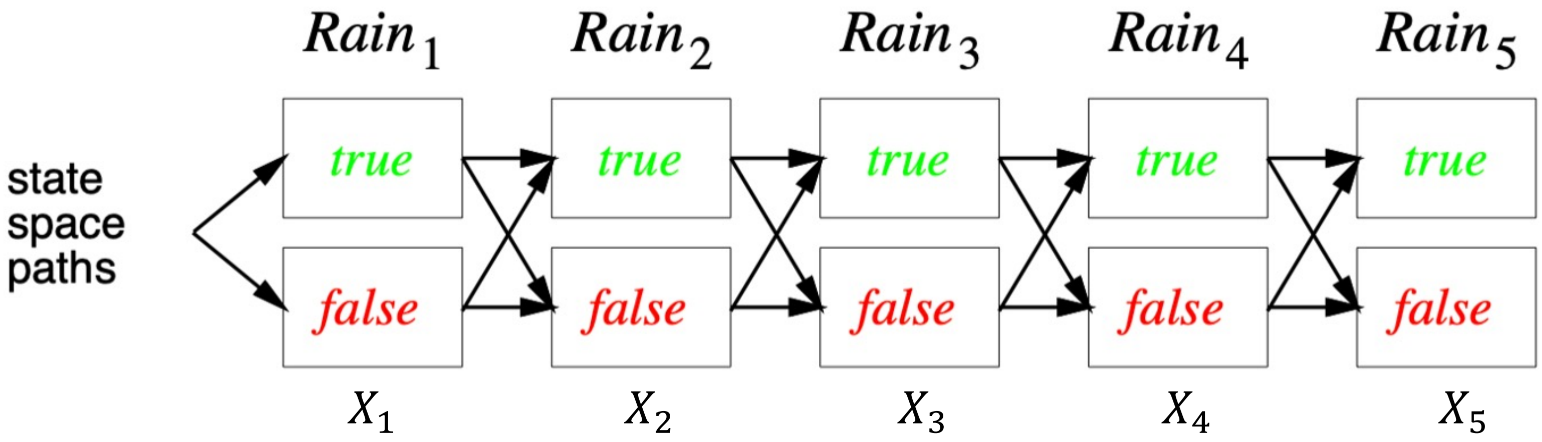
- 假设观察到的雨伞的情况是：[true, true, false, true, true]， 最有可能的天气序列是什么？

Most likely sequence ≠ sequence of most likely states!!!



维特比算法：到达X_{t+1}的最优路径包含到达X_t的最优路径

最可能解释：$\text{argmax}_{x_{1\ldots t}} P(X_{1:t} | e_{1,\ldots t})$



- 每条边代表状态的转移：$x_{t-1} \rightarrow x_t$

- 每条边具有一定的权重：$P(x_t | x_{t-1})P(e_t | x_t)$

- 每条路径都是一段状态序列，路径中每条边的概率乘积代表对应的状态序列在给定观测序列下的联合概率

- <u>前向算法</u>计算开始状态到结束状态的所有路径的概率和
- <u>维特比算法</u>计算开始状态到结束状态所有路径中概率最大的那一条

最可能解释： $\text{argmax}_{X_{1\ldots t}} P\left(X_{1:t} \middle| e_{1,\ldots t}\right)$

$$\max_{\mathbf{x}_1 \ldots \mathbf{x}_t} \mathbf{P}(\mathbf{x}_1, \ldots, \mathbf{x}_t, \mathbf{X}_{t+1} | \mathbf{e}_{1:t+1})$$
$$= \mathbf{P}(\mathbf{e}_{t+1} | \mathbf{X}_{t+1}) \max_{\mathbf{x}_t} \left(\mathbf{P}(\mathbf{X}_{t+1} | \mathbf{x}_t) \underbrace{\max_{\mathbf{x}_1 \ldots \mathbf{x}_{t-1}} P(\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}, \mathbf{x}_t | \mathbf{e}_{1:t}))}_{m_{1:t}}$$

前向算法：加和

$\boldsymbol{f}_{1:t+1}$ = FORWARD($\boldsymbol{f}_{1:t}$, e_{t+1})
$= P(e_{t+1}|X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t) \, \boldsymbol{f}_{1:t}$

维特比算法：最大

$\boldsymbol{m}_{1:t+1}$ = VITERBI($\boldsymbol{m}_{1:t}$, e_{t+1})
$= P(e_{t+1}|X_{t+1}) \max_{x_t} P(X_{t+1} | x_t) \, \boldsymbol{m}_{1:t}$

小结

- 马尔可夫性与马尔可夫模型
- 马尔可夫模型的平稳分布
- 隐马尔科夫模型
- 隐马尔可夫模型的前向推理
- 隐马尔可夫模型的极大似然序列
- 了解相关应用