计算机应用数学

Quaternijkon

计算机应用数学课程笔记

September 10, 2024

目录

1. 随机游走与马尔可夫链

1.1. 引言

Definition

随机游走 在有向图上：从一个起始顶点生成一系列顶点，每次随机选择一个出边，沿着这条边移动到一个新的顶点，并重复这个过程。 正式定义如下：

$$p(t)P = p(t+1)$$

其中，$p(t)$ 是一个行向量，它的每个分量表示在时间 t 时每个顶点的概率质量分布，P 是所谓的转移矩阵， $P_{i,j}$ 是游走从顶点 i 选择顶点 j 的概率。

Example

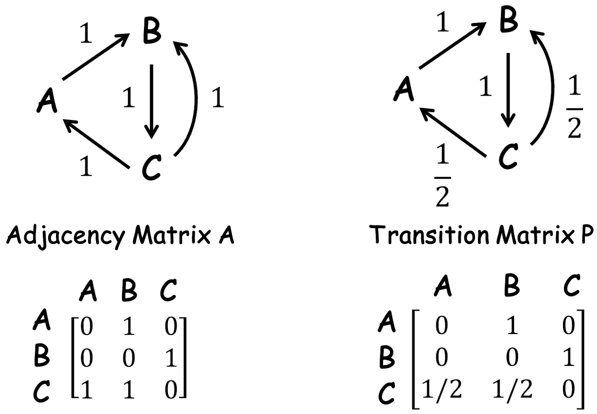

图 1　邻接矩阵与转移矩阵

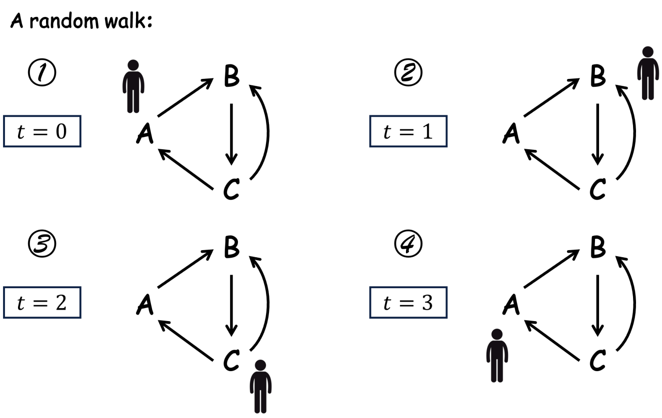

图 2　随机游走

Markov 链

有限的状态集合。

p_{xy}：从状态x到状态y的转移概率，$\sum_y p_{xy} = 1$。

Markov 链可以表示为有向图，其中从顶点 x 到顶点 y 的权重为p_{xy}。

Random walk	Markov chain
图 Graph	随机过程 Stochastic process
顶点 Vertex	状态 State
强连通 Strongly connected	持续 Persistent
非周期的 Aperiodic	非周期的 Aperiodic
强连通且非周期的 Strongly connected and aperiodic	遍历的 Ergodic
无向图 Undirected graph	时间可逆的 Time reversible

表 1 随机游走与马尔可夫链

我们将在本节中介绍以下内容：
- 示例：PageRank 和 Markov 决策过程。
- 平稳分布。
- 收敛性。
- Markov 过程。

Example

PageRank

将网页看作一个图：每个网页是一个顶点，超链接是边。

目标：根据重要性对网页进行排序。

Insight

一个网页的链接越多，它就越重要。

将入链看作投票，著名网站有更多的入链。

此外，来自重要网页的链接权重更大。

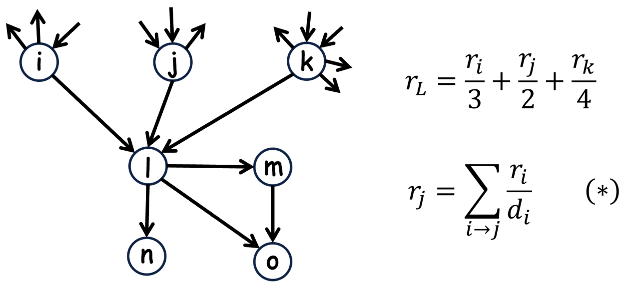

图 3

随机邻接矩阵

d_i 是节点 i 的度。

如果 $i \to j$，则 $M_{ji} = \frac{1}{d_i}$。

排序向量

r_i 是页面 i 的重要性得分。

公式 (*) 可以重写为：

$r = M \cdot r$

Example

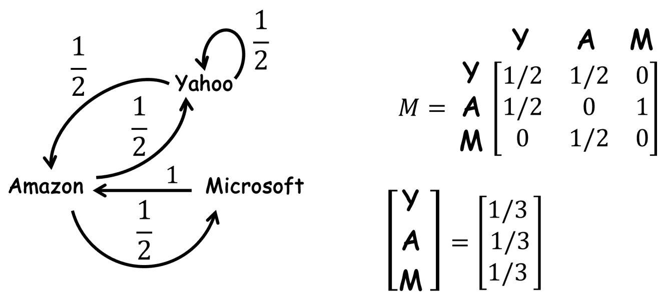

图 4

第一次迭代

$$\begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{1}{6} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$PR(Y)^1 = \frac{1}{2}PR(Y)^0 + \frac{1}{2}PR(A)^0 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3}$$

$$PR(A)^1 = \frac{1}{2}PR(Y)^0 + 1 \cdot PR(M)^0 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{3} = \frac{3}{6}$$

$$PR(M)^1 = \frac{1}{2}PR(A)^0 = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

第二次迭代

$$\begin{bmatrix} \frac{5}{12} \\ \frac{1}{3} \\ \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{1}{6} \end{bmatrix}$$

...

收敛

$$\begin{bmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{bmatrix}$$

Markov 过程（Markov 决策过程）

$$\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathbb{P}, \gamma$$

$\mathcal{S}$：状态集合。

$\mathcal{A}$：动作集合。

$\mathcal{R}$：在状态s下执行动作a的奖励$r(s, a)$。

$\mathbb{P}$：在状态 s 下执行动作 a 后转移到下一个状态 s' 的转移概率 $P(s' \mid s, a)$。

γ：折扣因子。

MDP（Markov 决策过程）：

$t=0$初始状态$s_0 \sim p(s_0)$
对于$t=0$到结束：
    执行动作a_t
    获得奖励$r_t \sim R(\cdot \mid s_t, a_t)$
    获得下一个状态$s_{t+1} \sim P(\cdot \mid s_t, a_t)$
    代理获得奖励r_t和状态s_{t+1}

算法 1 Markov 决策过程

目标：最大化长期奖励（累计奖励）$\sum_{t \geq 0} D^t r_t$。

Example

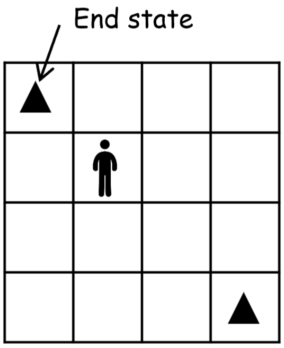

图 5

动作集合 = {左, 右, 上, 下} 到达空白格的奖励 ⟶ 使用最小化的动作数到达终点状态。

1.2. 平稳分布 Stationary Distribution

设 p_t 是随机游走经过 t 步后的概率分布。通过以下公式定义长期平均概率分布 a_t：

$$a_t = \frac{1}{t}(p_0 + p_1 + \cdots + p_{t-1})$$

Markov 链的基本定理：

对于一个连通的 Markov 链，它收敛于一个极限概率向量 x，满足：

$$XP = x; \sum_i x_i = 1 \Rightarrow X[P - I, 1] = [0, 1]$$

引理 1.3.1 设 P 是一个连通的 Markov 链的转移概率矩阵。通过在矩阵 $P - I$ 上增加一列 1 的列构造出的 $n \times (n+1)$ 矩阵 $A = [P - I, 1]$ 的秩为 n。

证明：作业

定理 1.3.2 设 P 是连通 Markov 链的转移概率矩阵，则存在一个唯一的概率向量 π 满足 $\pi P = \pi$。此外，对于任何初始分布，$\lim_{t\to\infty} a_t$ 存在且等于 π。

证明：考虑 a_t 和 a_{t+1} 的差，$a_t - a_{t+1} = a_t P$：

$$
\begin{aligned}
a_t P - a_t &= \frac{1}{t}[p_0 P + p_1 P + \cdots + p_{t-1} P] - \frac{1}{t}[p_0 + p_1 + \cdots + p_{t-1}] \\
&= \frac{1}{t}[p_1 + p_2 + \cdots + p_t] - \frac{1}{t}[p_0 + p_1 + \cdots + p_{t-1}] \\
&= \frac{1}{t}(p_t - p_0)
\end{aligned}
$$

因此，$b_t = a_t P - a_t$ 满足 $|\ b_t| \le \frac{2}{t}$，并且当 $t \to \infty$ 时趋于 0。

根据引理 1.3.1，$A = [P - I, 1]$ 的秩为 n。 由于 A 的所有行和为 0，$n \times n$ 矩阵 B 中除了最后一列以外的所有列是可逆的。

令 c_t 由 $b_t = a_t P - a_t$ 去掉第一列得到，使得 $a_t B = [c_t, 1]$。

因此 $a_t \to [c_t, 1] \to [0, 1]$ 并且 $a_t \to [0, 1]B^{-1}$。

因此 $a_t \to \pi$，我们得出 π 是一个概率向量。

由于 $a_t[P - I] = b_t \to 0$ ，我们得到 $\pi[P - I] = 0$。

由于 A 的秩为 n，这是唯一的解，如所要求的。

引理 1.3.3 对于在强连通图上的随机游走，若边上带有概率，向量 π 满足 $\pi_x p_{xy} = \pi_y p_{yx}$ 对于所有 x和y，且 $\sum_x \pi_x = 1$，那么 π 是随机游走的平稳分布。

证明： $\pi_x p_{xy} = \pi_y p_{yx}$，两边求和，$\pi_x = \sum_y \pi_y p_{yx}$，因此 (pi) 满足 $\pi = \pi P$。(By Theorem 1.3.2 ...)

1.3. 无向图上随机游走的收敛性

下一个问题：游走需要多长时间开始反映 Markov 过程的平稳概率？

示例： 这需要很长时间才能收敛。游走很难通过图的两个部分之间的窄通道。

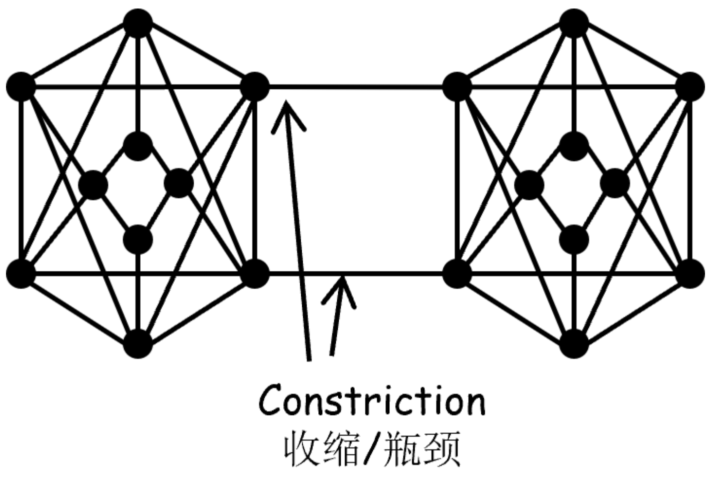

图 6

我们在下面定义了 Markov 链的收缩的一个组合度量，称为归一化导通率。

定义 1.3.1 设 $\varepsilon > 0$。Markov 链的 $\varepsilon - mixing$时间是最小的整数 t，使得对于任何初始分布 P_0，第 t 步的平均概率分布与平稳分布之间的 1-范数距离最多为 ε。

$$
|\ a_t - \pi\ | \le \varepsilon
$$

定义 1.3.2 对于一个顶点子集 S，令 $\pi(S)$ 表示 $\sum_{x \in S} \pi_x$。归一化导通率定义为：

$$\Phi(S) = \frac{\sum_{(x,y)\in(S,\bar{S})} \pi_x p_{xy}}{\min(\pi(S),\pi(\bar{S}))}$$

其中，$\bar{S} = V - S$ 。$\pi(S)$ 是平稳分布下，Markov 链处于某状态属于 S 的概率。

定义 1.3.3 Markov 链的归一化导通率，记作 Φ ，定义为：

$$\Phi = \min_S \Phi(S)$$

定理 1.3.4. 在无向图上，随机游走的 $\varepsilon - mixing$时间为：

$$\Phi\left(\frac{\ln\left(\frac{1}{\pi_{\min}}\right)}{\Phi^2\varepsilon^3}\right)$$

其中，$\pi_{\min}$ 是任何状态的最小平稳概率。

使用归一化导通率证明收敛性。

接下来，我们应用定理 1.3.4 通过一些例子说明归一化导通率如何限制收敛速度。

① **一个一维的格子**

图 7

n 个顶点路径，两端都有自环。

平稳概率在所有顶点上是均匀的 $\frac{1}{n}$。

具有最小归一化导通率的集合是：
- 具有 $\pi \le \frac{1}{2}$ 的集合；
- 包含前 $\frac{n}{2}$ 个顶点的集合。

从集合 S 到集合 $\bar{S}$ 的边的总导通率是：

$\pi_m p_{m,m+1} = \Omega\left(\frac{1}{n}\right), \left(m = \frac{n}{2}\right)$

$\pi(S) = \frac{1}{2}$

因此，$\Phi(\bar{S}) = 2\pi_m p_{m,m+1} = \Omega\left(\frac{1}{n}\right)$

根据定理 1.3.4，对于 $\varepsilon = \frac{1}{100}$，经过 $O(n^2 \log n)$ 步之后，$||\, a_t - \pi \,|| \le \frac{1}{100}$。

此图没有快速收敛性。

1.4. 4. 单位边权重的无向图上的随机游走

我们使用这种特殊类型的图来回答以下问题：

- 随机游走从 x 到达 y 的期望时间是多少？
- 从 x 到 y 并返回的期望时间是多少？
- 到达每个顶点的期望时间是多少？

$$\begin{tikzcd} X \arrow[r, bend left, "\text{random walk}"] & Y \end{tikzcd}$$

① **命中时间**

h_{xy} —— 也称为发现时间。

引理 1.3.5. 从路径上的一个端点开始随机游走，穿过有 n 个顶点的路径到达另一端的期望时间是 $\Theta H(n^2)$。

证明:

$$h_{12} = 1$$

$$h_{i,i+1} = \tfrac{1}{2} + \tfrac{1}{2}(1 + h_{i-1,i+1})$$

$$= 1 + \tfrac{1}{2}(h_{i-1,i} + h_{i,i+1})$$

$$= 2 + h_{i-1,i}$$

因此，$h_{i,i+1} = 2i - 1$，$2 \le i \le n-1$

要从 1 走到 n，

$$h_{1,n} = \textstyle\sum_{i=1}^{n-1} h_{i,i+1}$$

$$= \textstyle\sum_{i=1}^{n-1}(2i-1)$$

$$= 2\textstyle\sum_{i=1}^{n-1} i - \sum_{i=1}^{n-1} 1$$

$$= 2\tfrac{n(n-1)}{2} - (n-1)$$

$$= (n-1)^2$$

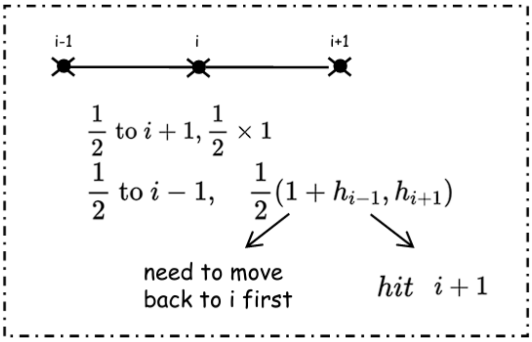

图 8

引理 1.3.6 设随机游走从顶点 1 到顶点 n，在包含 n 个顶点的链中。令 $t(i)$ 为在顶点 i 停留的期望时间。那么：

$$t(i) = \begin{cases} n-1, & i = 1 \\ 2(n-i), & 2 \le i \le n-1 \\ 1, & i = n \end{cases}$$

证明 现在 $t(n) = 1$，因为游走到达 n 时会停止。当游走到达 $n-1$ 时，一半的时间它会继续走向 n。因此，$t(n-1) = 2$。对于 $3 \le i \le n-1$，

$$t(i) = \tfrac{1}{2}[t(i-1) + t(i+1)]$$

$$t(1) = \frac{1}{2}t(2) + 1$$

$$t(2) = t(1) + \frac{1}{2}t(3)$$

因此我们得到

$t(i+1) = 2t(i) - t(i-1)$

因此，$t(i) = 2(n-i)$ 对于 $3 \le i \le n-1$。

$t(2) = 2(n-2)$，$t(1) = n-1$。

因此，在顶点停留的总时间是

$$n - 1 + 2(1 + 2 + ... + n - 2) + 1 = (n - 1 + 1 + 2\frac{(n-1)(n-2)}{2} + 1) = (n-1)^2 + 1$$

这比 h_{1n} 多出 1。

② **往返时间**

$$commute(x, y) = h_{xy} + h_{yx}$$

③ **覆盖时间**

$Cover(x, G) \to$ 从顶点 x 开始的随机游走到达每个顶点至少一次的期望时间。

$$Cover(G) = \max_x Cover(x, G)$$

定理 1.3.7. 设 G 是一个有 n 个顶点和 m 条边的图。覆盖时间 $Cover(G)$ 的上界为 $4m(n-1)$。

证明. 进行一次从某个顶点 Z 开始的深度优先搜索。T 是结果生成的深度优先搜索生成树。深度优先搜索覆盖每个顶点。注意，生成树中的每条边在两个方向上都被遍历了两次。

$$Cover(Z, G) \le \sum_{(x,y)\in T,(y,x)\in T} h_{xy}$$

推论. 如果 x 和 y 是相邻的, 则 $h_{xy} + h_{yx} \le 2m$, 其中 m 是边的数量。该推论表明 $h_{xy} \le 2m$。由于深度优先搜索树中有 $n-1$ 条边，并且每条边都被遍历两次，$Cover(Z) \le 4m(n-1)$。因此，$Cover(G) \le 4m(n-1)$。

1.5. 更多关于 Markov 的内容

△ **一个简单的 Markov 链** $\langle S, P \rangle$ S: 状态, P: 概率

总奖励

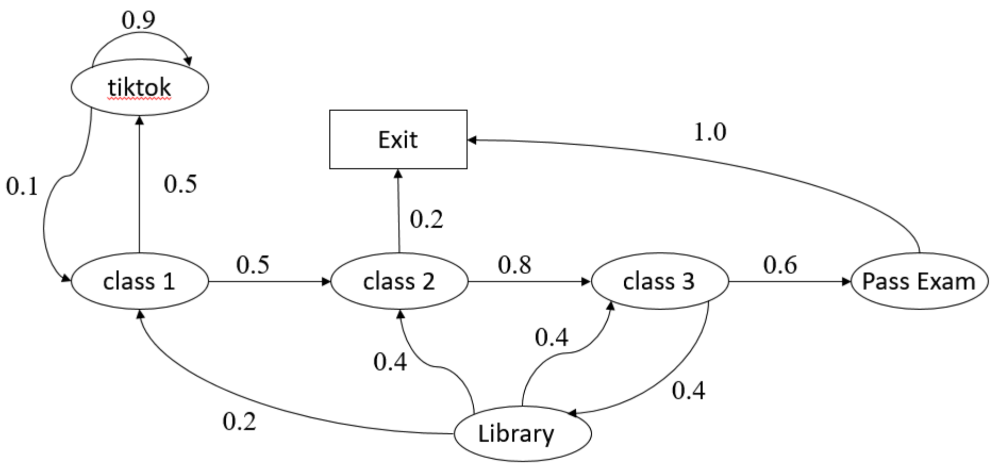

图 9

大量路径的例子：

$C_1, C_2, C_3, pass$

$C_1, TikTok, C_1, C_2, C_3, pass$

$C_1, C_2, C_3, Library, C_2, pass$

……

△ **Markov 奖励过程** $\langle S, P, R, \gamma \rangle$ R: 奖励, γ: 折扣因子

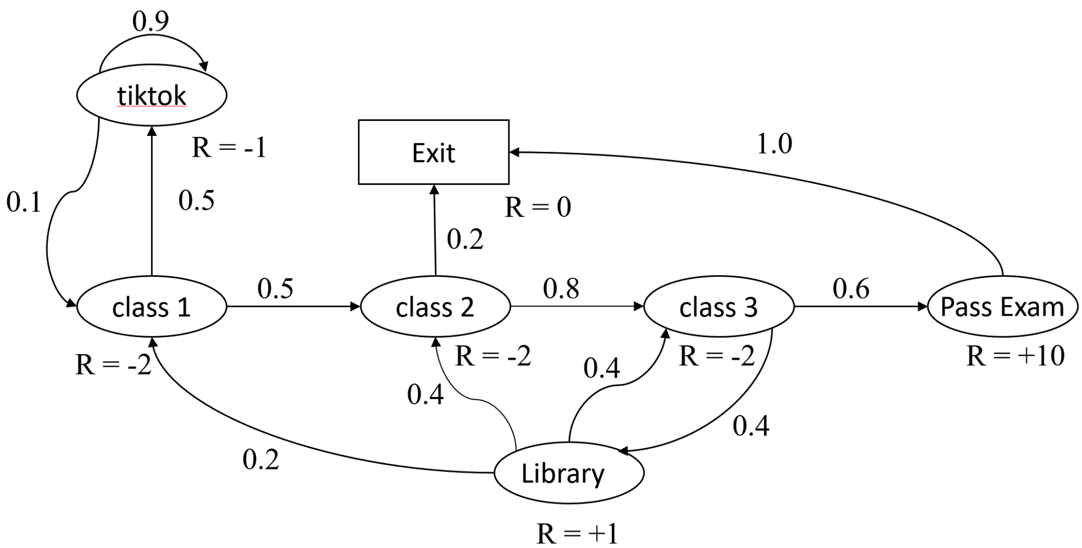

图 10

总奖励

$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots$

$= \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$

状态的价值函数

$V(s) = \mathbb{E}[G_t \mid S_t = s]$

= 从该状态开始的期望奖励，即不同路径的平均奖励。

路径：$C_1, C_2, C_3, Pass, Exit$

$S_1 = C_1$ 且 $\gamma = 1/2$

$V_{C_1} = -2 - 2 \cdot \frac{1}{2} - 2 \cdot \frac{1}{4} - 10 \cdot \frac{1}{8} = -2.25$

路径：$C_1, TikTok, TikTok, C_1, C_2, Exit$

$V_{C_1} = -2 - 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{4} - 2 \cdot \frac{1}{8} - 2 \cdot \frac{1}{16} = -3.125$

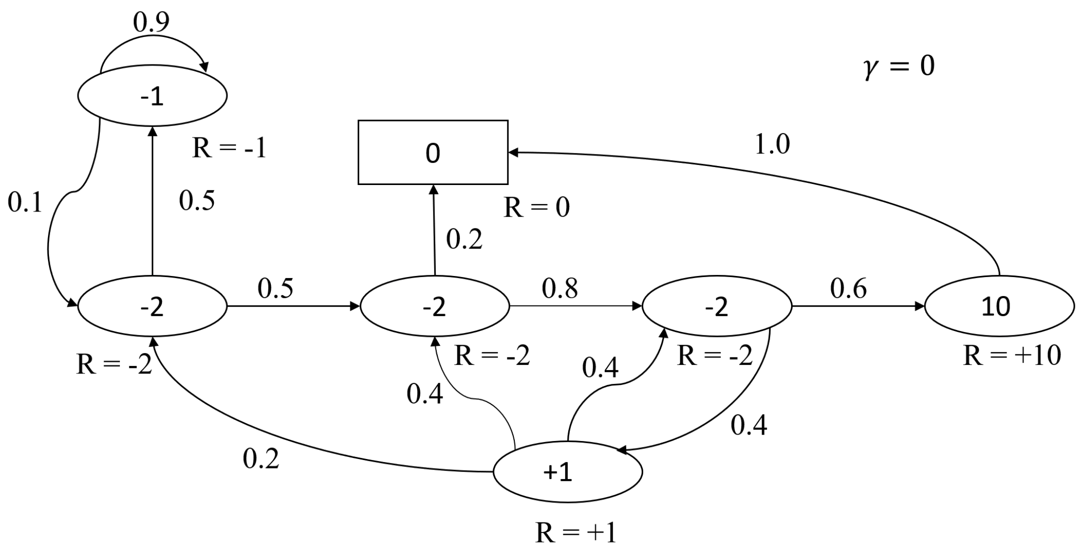

图 11

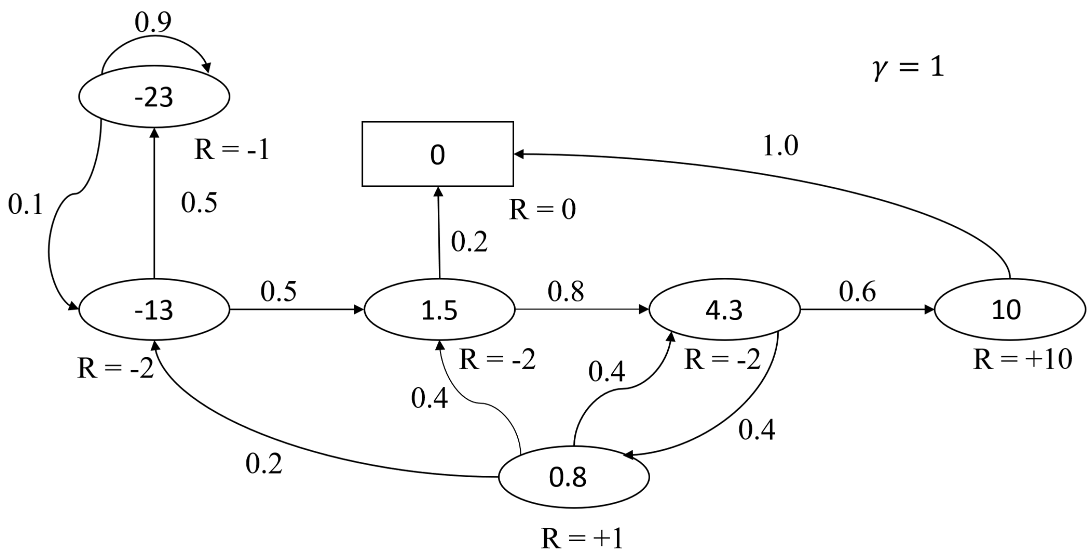

图 12

Bellman 期望方程

$$
\begin{aligned}
V(s) &= \mathbb{E}[G_t \mid s_t = s] \\
&= \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \mid s_t = s] \\
&= \mathbb{E}[R_{t+1} + \gamma (R_{t+2} + \gamma R_{t+3} \cdots) \mid s_t = s] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \mid s_t = s] \\
&= \mathbb{E}[R_{t+1} + \gamma V(s_{t+1}) \mid s_t = s]
\end{aligned}
$$

使用 s' 表示 $t+1$ 的可能状态，

$$V(s) = R_s + \gamma \sum_{s' \in S} P_{ss'} V(s')$$

对于 class 3

$$4.3 = -2 + 0.6 \times 10 + 0.4 \times 0.8$$

△ **Markov 决策过程** $\langle S, A, P, R, \gamma \rangle$ A : 动作

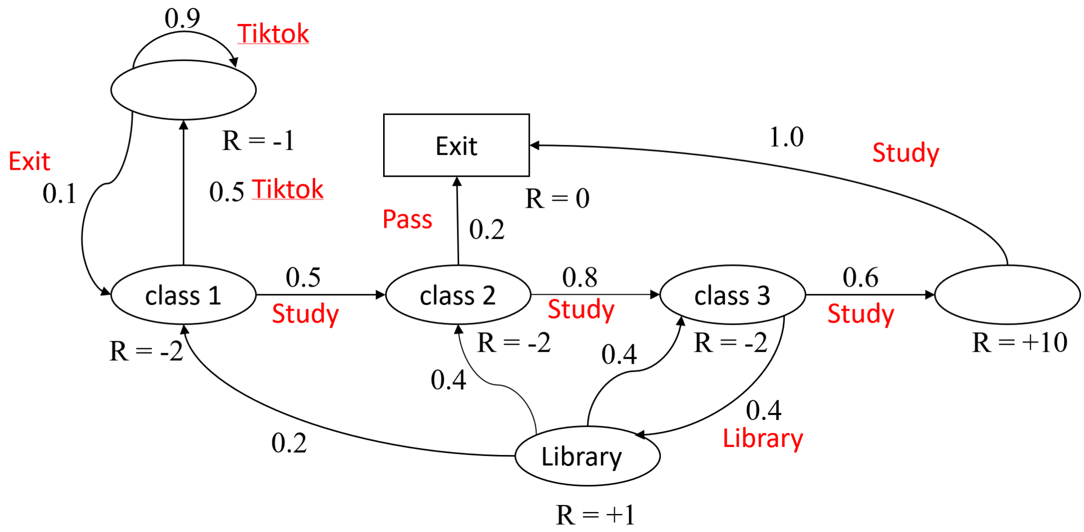

图 13

策略：采取动作的概率分布。

$$\pi(a \mid s) = \mathbb{P}[A_t = a \mid S_t = s]$$

给定一个 MDP $M = \langle S, A, P, R, \gamma \rangle$ 和一个策略 π。 序列 $S_1, S_2, \cdots$ 是一个 Markov 过程 $\langle S, p^\pi \rangle$。

状态和奖励序列 $S_1, R_2, S_2, \cdots$ 是一个 Markov 过程 $\langle S, P^\pi, R^\pi, \gamma \rangle$。

在策略 π 下，从状态 s 转移到 s' 的概率是：

$$P^\pi_{ss'} = \sum_{a \in A} \pi(a \mid s) P^a_{ss'}$$

在策略 π 下，状态 s 的奖励是：

$$R^\pi_s = \sum_{a \in A} \pi(a \mid s) R^a_s$$

价值函数：

$$V_\pi(s)=\mathbb{E}_\pi[G_t \mid S_t=s]$$

策略价值函数：

$$q_\pi(s,a)=\mathbb{E}_\pi[G_t \mid S_t=s, A_t=a]$$

Bellman 方程：

$$\begin{aligned}
V_\pi(s) &= \mathbb{E}_\pi\left[R_{t+1}+\gamma v_\pi\left(s_{t+1}\right) \mid s_t=s\right] \\
q_\pi(s,a) &= \mathbb{E}_\pi\left[R_{t+1}+\gamma q_\pi\left(s_{t+1}, A_{t+1}\right) \mid s_t=s, A_t=a\right] \\
\text{因此} \quad V_\pi(s) &= \sum_{a\in\mathcal{A}} \pi(a \mid s) q_\pi(s,a)
\end{aligned}$$

且 $q_\pi(s)=R_s^a+\gamma \sum_{s'\in S} P_{ss'}^a V_\pi(s')$

因此，

$$q_\pi(s)=R_s^a+\gamma \sum_{s'\in S} P_{ss'}^a \sum_{a'\in A} \pi(a' \mid s') q_\pi(s', a')$$

最优价值函数

$$V_*(s)=\max_\pi V_\pi(s)$$

最优动作-价值函数

$$q_*(s,a)=\max_\pi q_\pi(s,a)$$

我们可以使用 $q_*(s,a)$ 来得到最优策略 π：

$$\pi_*(a \mid s)=\begin{cases} 1 \text{ if } a=\arg\max_{a\in A} q_*(s,a) \\ 0 \text{ otherwise}\end{cases}$$