MDP Wyprowadzenie

18 kwietnia 2020

Stan opisany jako krotka: (D, PDO, N)

1. D - czy pole, na którym jest agent jest brudne,
2. PDO - prawdopodobieństwo, że drugie pole jest brudne,
3. N - ilość kroków która została.

Warunki końcowe - nie ma więcej kroków do zrobienia.

$$V(\cdot, \cdot, 0) = 0$$

Szukane: $V(N) = (1 - p_0) \cdot V(0, p_0, N) + p_0 \cdot V(1, p_0, N)$.
Zakładamy $p_0 = 0.5$ oraz $p = 0.05$.
Zdefiniujmy pomocniczą funkcję $n(PDO)$, która dla prawdopodobieństwa, że na drugim polu jest brudno (PDO) zwraca jakie będzie to prawdopodobieństwo po zrobieniu kroku:
$n(PDO) \equiv PDO + (1 - PDO) * p$

N = 1

$$\begin{aligned}
V(0, p_0, 1) &= max\{\overbrace{0 + ((1-p) \cdot V(0, n(p_0), 0) + p \cdot V(1, n(p_0), 0))}^{SUCK}, \\
&\quad \overbrace{-1 + (n(p_0)V(1, n(0), 0) + (1 - n(p0))V(0, n(0), 0))}^{GO}\} \\
&= max\{0, -1\} = 0 \\
V(1, p_0, 1) &= max\{\overbrace{10 + ((1-p) \cdot V(0, n(p_0), 0) + p \cdot V(1, n(p_0), 0))}^{SUCK}, \\
&\quad \overbrace{-1 + (n(p_0)V(1, n(1), 0) + (1 - n(p0))V(0, n(1), 0))}^{GO}\} \\
&= max\{10, -1\} = 10
\end{aligned}$$

$$\begin{aligned}
V(N = 1) &= (1 - p_0) \cdot V(0, p_0, 1) + p_0 \cdot V(1, p_0, 1) \\
&\quad \text{podstawiamy } p_0 \equiv 0.5 \\
&= 0 * 0.5 + 10 * 0.5 = 5
\end{aligned}$$

N = 2

$$
\begin{aligned}
V(0,p_0,2) &= max\{\overbrace{0+((1-p)\cdot V(0,n(p_0),1)+p\cdot V(1,n(p_0),0))}^{SUCK},\\
&\quad \overbrace{-1+(n(p_0)V(1,n(0),0)+(1-n(p0))V(0,n(0),1))}^{GO}\}\\
&= max\{\overbrace{0+(0.95\cdot 0+0.05\cdot 10)}^{SUCK},\overbrace{-1+(0.525\cdot 10+0.475\cdot 0)}^{GO}\}\\
&= max\{0+0.5,-1+5.25\}=4.25\\
V(1,p_0,2) &= max\{\overbrace{10+((1-p)\cdot V(0,n(p_0),1)+p\cdot V(1,n(p_0),0))}^{SUCK},\\
&\quad \overbrace{-1+(n(p_0)V(1,n(1),1)+(1-n(p0))V(0,n(1),1))}^{GO}\}\\
&= max\{\overbrace{10+(0.95\cdot 0+0.05\cdot 10)}^{SUCK},\overbrace{-1+(0.525\cdot 10+0.475\cdot 0}^{GO}\}\\
&= max\{10+0.5,-1+5.25\}=10.5
\end{aligned}
$$

Użyte wyżej wartości:

$$
\begin{aligned}
V(0,n(p_0),1) &= max\{\overbrace{0+((1-p)\cdot V(0,n(n(p_0)),0)+p\cdot V(1,n(n(p_0)),0))}^{SUCK},\\
&\overbrace{-1+(n(n(p_0))V(0,n(0),0)+(1-n(n(p0)))V(1,n(0),0))}^{GO}\}\\
&= max\{0,-1\}=0\\
V(1,n(p_0),1) &= max\{\overbrace{10+((1-p)\cdot V(0,n(n(p_0)),0)+p\cdot V(1,n(n(p_0)),0))}^{SUCK},\\
&\overbrace{-1+(n(n(p_0))V(0,n(1),0)+(1-n(n(p0)))V(1,n(1),0))}^{GO}\}\\
&= max\{10,-1\}=10\\
V(0,\overbrace{n(1)}^{=1},1) &= max\{\overbrace{0+((1-p)\cdot V(0,1,0)+p\cdot V(1,1,0))}^{SUCK},\overbrace{-1+(1\cdot V(0,n(0),0)}^{GO}\}\\
&= max\{0,-1\}=0\\
V(1,\overbrace{n(1)}^{=1},1) &= max\{\overbrace{10+((1-p)\cdot V(0,n(p_0),0)+p\cdot V(1,n(p_0),0))}^{SUCK},\overbrace{-1+(1\cdot V(0,n(1),0)}^{GO}\}\\
&= max\{10,-1\}=10\\
V(1,n(0),1) &= V(1,p_0,1)=10\\
V(0,n(0),1) &= V(0,p_0,1)=0
\end{aligned}
$$

$$
\begin{aligned}
V(N=2) &= (1-p_0)\cdot V(0,p_0,2)+p_0\cdot V(1,p_0,2)\\
&\quad \text{podstawiamy } p_0\equiv 0.5\\
&= 4.25*0.5+10.5*0.5=7.375
\end{aligned}
$$