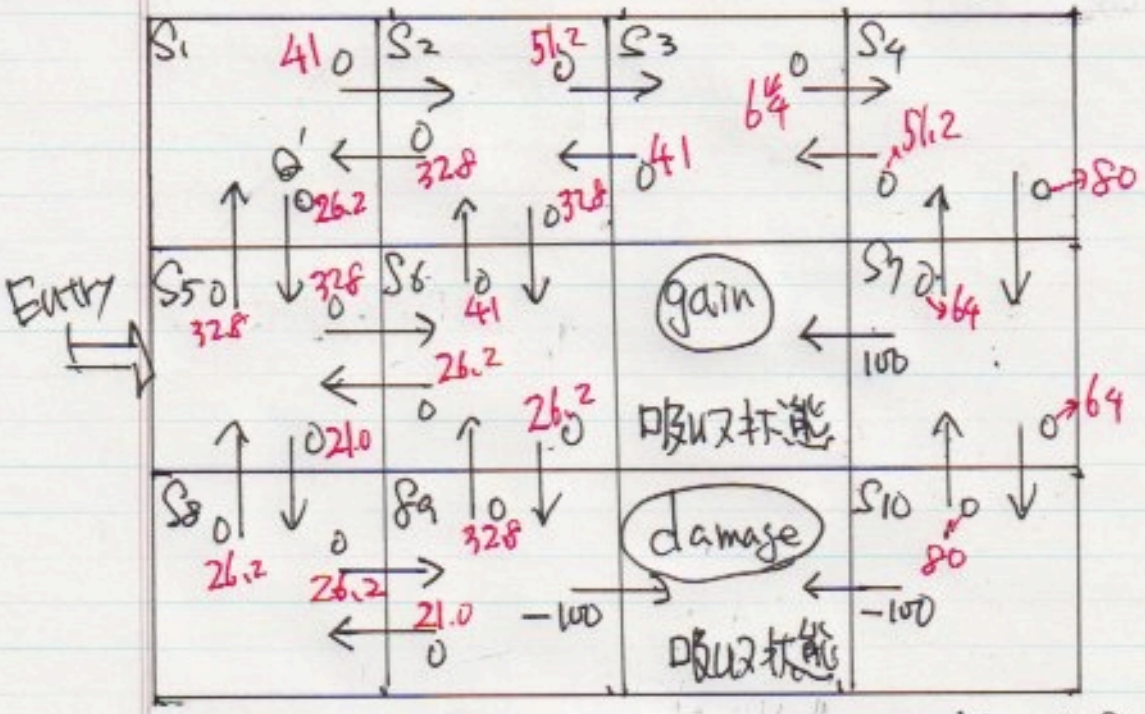
AI2. 12th week

$$Q(s,a) \equiv r(s,a) + \gamma \max_{a'} Q(\delta(s,a)a')$$



$\gamma = 0.8$

Qテーブル

状態	行動	Q'
S_1	→	0
S_1	↓	0
S_2	→	0
S_2	←	0
S_2	↓	⋮
⋮		
S_{10}	↑	0

Q'
41
26.2
51.2
32.8
⋮
80.

$\pi(S_t) = a_t$

学習の結果

$\pi(S_5) = \uparrow$

$\pi(S_8) = \uparrow$

⋮

最適な行動ポリシー(π)

(↑, →, →, →, ↓, ←)

(→, ↑, →, →, ↓, ←)