

AI(人工知能)

機械学習

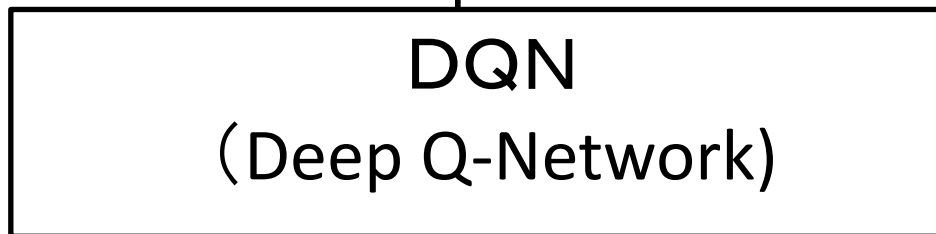
強化学習

Q学習

ディープラーニング

DQN

(Deep Q-Network)



巣

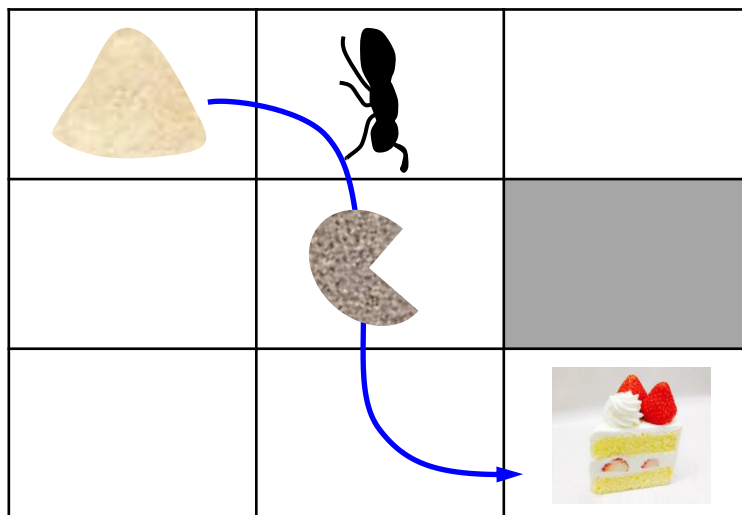
弱いフェロモン

ケーキ

クッキー

蟻

強いフェロモン



状態s=1	s=2  7	s=3
s=4	s=5	s=6
s=7	s=8	s=9



Q値: $Q(s, a)$

ここでs: 状態 (state)

a: アクション (action)

Q値		アクション			
		右	上	左	下
状態	...	...	...	...	...
	2	①	欄外	③	⑦
	...	...	...	...	...

Q値は、行動の「価値」、「報酬」

ここではフェロモン量やクッキー

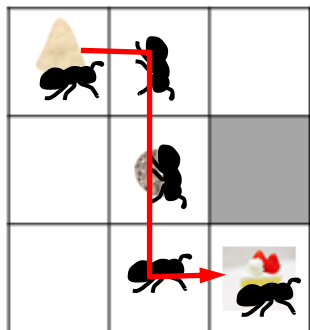
$$Q(2, \text{右}) = 1$$

$$Q(2, \text{上}) = \text{欄外}$$

$$Q(2, \text{左}) = 3$$

$$Q(2, \text{下}) = 7$$

エピソード1

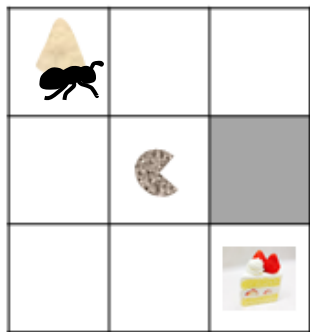


状態

 $s_1=1$	 $s_1=2$	 $s_1=3$
 $s_1=4$	 $s_1=5$	 $s_1=6$
 $s_1=7$	 $s_1=8$	 $s_1=9$

時間 報酬

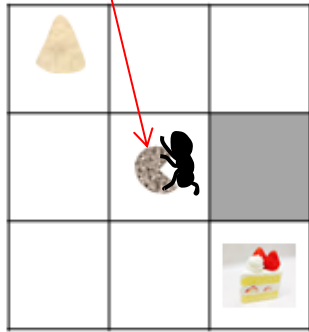
t=1
 $r_1=0$



t=2
 $r_2=0$



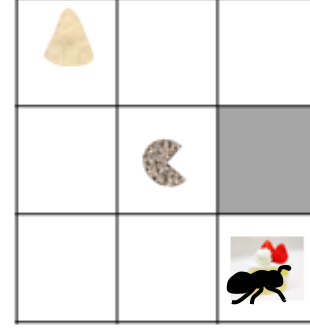
t=3
 $r_3=2.71$



t=4
 $r_4=0$



t=5
 $r_5=100$



アクション

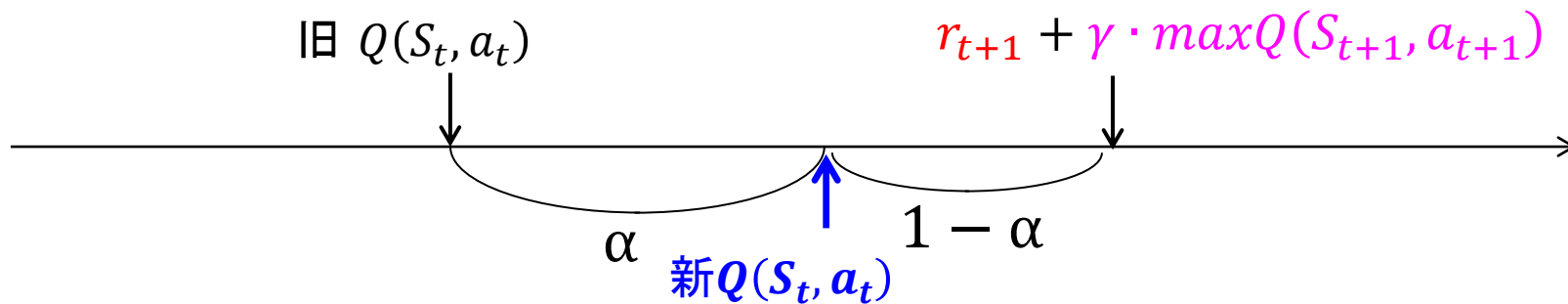
$a_1=右$

$a_2=下$

$a_3=下$

$a_4=右$

γ : 割引率 …… フェロモンが時間と共に薄れたりする分を割り引きます
 α : 学習率 …… 何度も往復して学習すると信頼度が上がってきます



遅延報酬

新 $Q(S_t, a_t) \leftarrow Q(S_t, a_t) + \alpha \{ r_{t+1} + \gamma \cdot \max Q(S_{t+1}, a_{t+1}) - Q(S_t, a_t) \}$

期待報酬

