
Playing games with QLearning

— Florentina Bratiloveanu —

What is Reinforcement Learning?

- area from Machine Learning having roots in behaviourist psychology
- used in multi-agent systems, swarm intelligence, game theory
- environment specified as Markov Decision Process(MDP)

$$MDP = (S, A, P(s, s'), R(s, s'), \gamma)$$

S - finite set of states

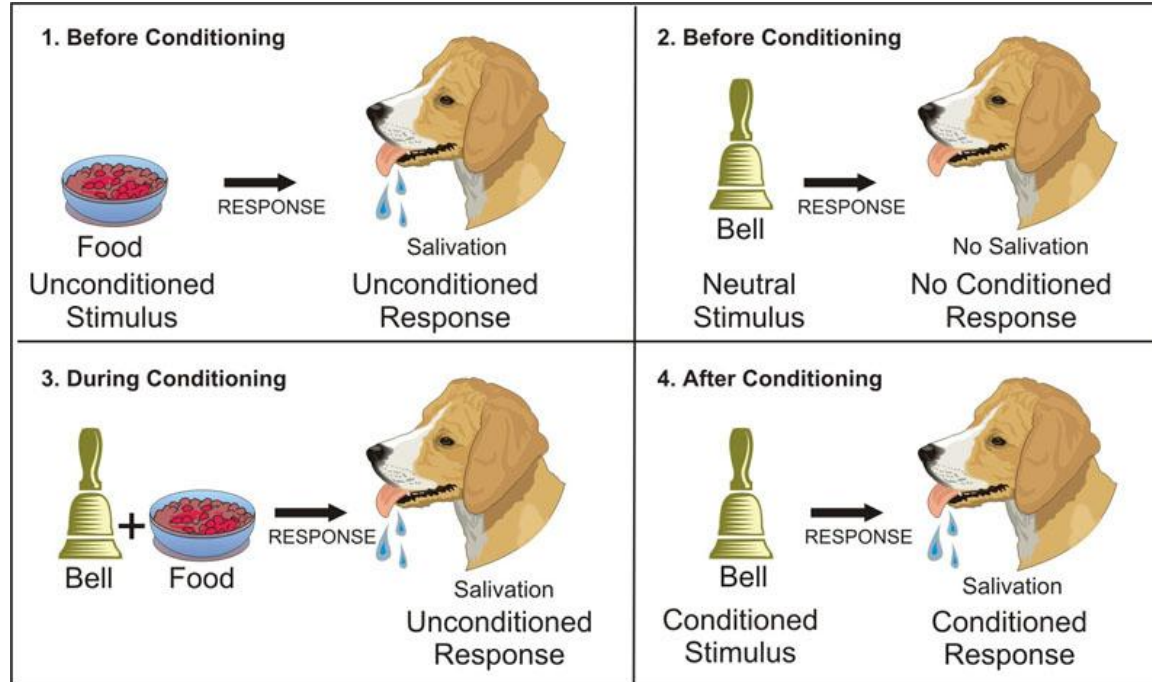
A - finite set of actions

$P_a(s, s')$ - probability of getting state s' after applying action a on state s

$R_a(s, s')$ - the reward after getting state s' from state s with action a applied

γ - discount factor

Classical conditioning



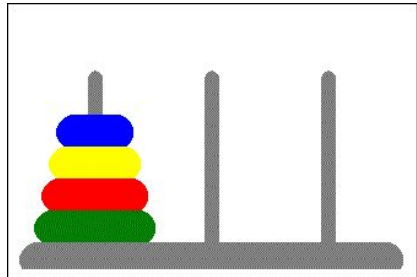
Classical Conditioning [1]

Q-Learning

- maintains a table of $Q[s,a]$ where S is the set of states and A is the set of actions
- long term learning
- exploration vs exploitation
- we don't know how a move will affect the reward
- update table Q using the next formulas

$$Q(s, a) := Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad [2]$$

Game state codification



Hanoi state

```
[picker]: disk=2 pointing=2  
[0]: 3  
[1]:  
[2]: 1
```

Map example

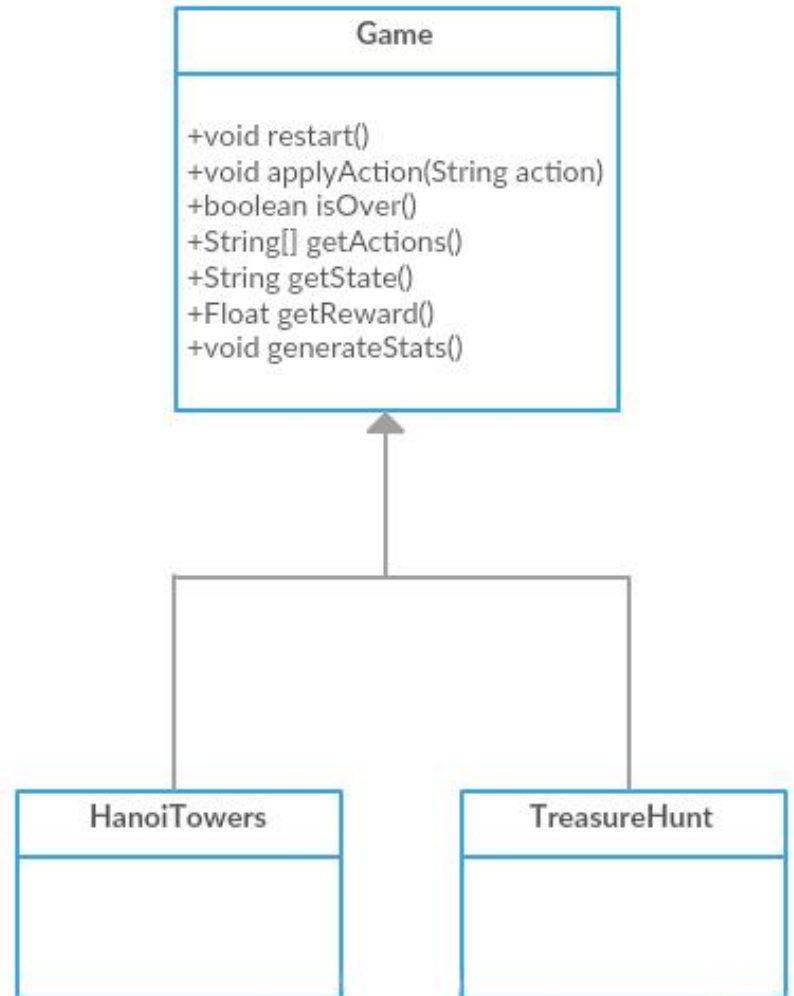
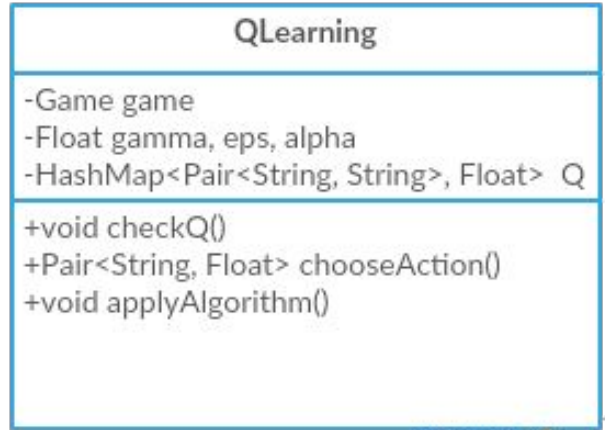
```
4 7  
#####  
#H    M#  
#  T  #  
#####
```

Treasure Hunt state

```
Hero[position=(row=1, col=1)]  
Monster[position=(row=1, col=5)]
```



Architecture



Results(I) - Hanoi Towers

- 3 stacks, 3 disks with $\gamma=0.95$ and $\alpha=0.1$ (trained 100 episodes for each α)

Eps	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	off(0.0)
Moves	5417	256	186	89	52	62	52	35	41	31	27

7

- 4 stacks, 4 disks with $\gamma=0.95$ and $\alpha=0.1$ (trained 100 episodes for each α)

Eps	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	off(0.0)
Moves	24050	646	265	208	162	112	76	50	53	45	41

15

Results (II) - Treasure Hunt

- 1 monster, 1 treasure with $\gamma=0.95$ and $\alpha=0.1$ (trained 1000 episodes for each α)

Eps	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	off(0.0)
Moves	3	6	7	3	3	3	17	3	16	9	7
Score (Hero-Monster)	1-0	453-547	491-509	500-500	533-467	574-426	573-427	511-489	516-484	451-549	474-526

Questions?

Thank you
for your attention!