Parallel Reinforcement Learning (Q-Learning)

Yumeng Pan, Qianqian Guo

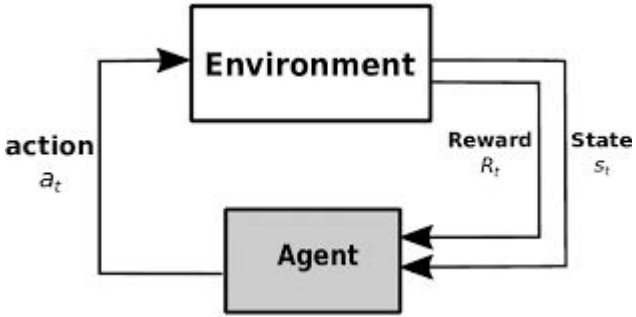
Q-Learning

Q-learning is a **Reinforcement Learning** (RL) method that deals with the problem of learning to **control autonomous agents**. The learning process works based on interactions by **trial and error** with a dynamic environment which provides **reward** signals for each **action** the agent executes.

Bellman equation in Q-value form:

$$Q^{\pi}(s,a) = \mathbb{E}[r_{t+1} + \lambda r_{t+2} + \lambda^2 r_{t+3} + \ldots | s, a]$$

$$= \mathbb{E}_{s'}[r + \lambda Q^{\pi}(s', a') | s, a]$$



Q-learning Algorithm

1 Initialize (with 0's or random values) $Q(s,a)$ for all $s \in S$ and for all $a \in A(s)$
2 Repeat (for each episode)
3 Initialize s
4 Repeat (for each step episode):
5 Choose a from s using a policy derived from Q (e.g., $\in$-greedy)
6 Take action a, observe resultant state s' and the reward r.
7 $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$ **Greedy algorithm**
8 $s \leftarrow s'$;
9 until s is terminal

s: Sate.
a: Action.
r: Reward.
alpha: Learning rate parameter.
gamma: Decay rate (future reward discount) parameter.

Q-table

- A table with the Q value for every <S, A> pair.

Initialized

Q-Table		Actions					
		South (0)	North (1)	East (2)	West (3)	Pickup (4)	Dropoff (5)
States	0	0	0	0	0	0	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	327	0	0	0	0	0	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	499	0	0	0	0	0	0

Training

Q-Table		Actions					
		South (0)	North (1)	East (2)	West (3)	Pickup (4)	Dropoff (5)
States	0	0	0	0	0	0	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	328	-2.30108105	-1.970920[illegible]96	-2.30357004	-2.20[illegible]839	-10.3607344	-8.5583017
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	499	9.96984239	4.02706992	12.96022777	29	3.32877873	3.38230603

OpenMPI
(multi processors)

Objectives

- Select the RL problem for performance comparison.
- Parallelize Q-learning by splitting the Q-table.
- Find an efficient way of information exchange.
- Compare its performance with serial code.
- Compare the performance for a different number of processors.