Model-free Reinforcement Learning: Q-Learning

Course assignment: Machine Learning
University of Piraeus, Demokritos

Sarafidis, Tasos
sar.tasos@gmail.com

Papadopoulos, Vasileios
vassilispapadop@gmail.com

February 27, 2021

1 Reinforcement Learning

Reinforcement Learning is an area of Machine Learning, where the main purpose is to find a way for an agent to learn a policy by interacting with its environment. An agent lives within an environment and interacts with it by examining the state in which it is in, for every time step. After examining its state, the agent chooses an action and the environment returns the corresponding reward, as well as the agent's new state. The agent's goal is to find the actions that return the most reward by trying them, without knowing which actions to choose beforehand. This means, the agent must discover the best action that will allow it to learn the optimal policy towards achieving its goals.

In Reinforcement Learning problems, the agent is always in interaction with the environment. In every time step t, the agent receives a state of the environment, st, and chooses an action, at. In the next time step, the agent gets a reward, r_{t+1}, from the environment and moves to the next state, s_{t+1}. In every step, the agent assigns a state to an action thus forming the agent's policy, π_t.

Problems in the area of Reinforcement Learning are usually modeled as Markov Decision Processes (MDPs). A MDP is a time-discrete stochastic control process, where agents' actions affect immediate and future rewards, as well as next states. A Markov Decision Process can be described by a tuple S,A,R,T, where:

- S is the set of all possible states.
- A is the set of actions.
- R(s) is the reward function, which returns a numerical value as a result of an agent's action.
- $T(s, \alpha, s')$ is the state transition function, which given a state s and an action α returns the next state s'.

2 Multi-Agent Systems

A Multi-Agent System is a set of agents who interact with each other and have a common or contradictory goal. In multi-agent systems, each agent is a part of the other agents' environment, meaning that as an agent learns and looks for the best policy, the results of its actions depend, not only on its state, but also on the other agents' actions. Agents either cooperate with each other to achieve a common goal, or come into conflict with each other in order to succeed their personal goals.

The Markov Decision Process that was described in section 1, can be also expanded on multi-agent systems and can be denoted by a tuple Ag, S,A,R,T:

- Ag is set of agents.
- S is set of states.
- A is set of possible actions.
- R is the reward function.
- T is the transition function.

3 Q-Learning

Q-Learning is an off-policy reinforcement learning algorithm that allows an agent to choose the best action in a given state. When the agent is in a state

and performs an action, the environment returns the reward and the agent's next state. Then, the agent estimates the value of the new state. Each time an agent chooses an action and learns about its new state, the overall value of the executed action in the corresponding state is updated by the *update rule*:

There are two ways for an agent to choose its action in each state. Firstly, the agent chooses randomly its next action without looking up the Q-table. This method is called *exploration*. The second way of choosing its action is called *exploitation*. According to this, the agent looks up the Q-table to find out which action has the best Q-value and execute it.

But, how does the agent decide whether to explore or exploit? This is done by using the $\epsilon - greedy$ algorithm. This algorithm states that with a probability of $0 \leq \epsilon \leq 1$ an agent chooses to explore and with a probability of the agent chooses to exploit the Q-table.

References

[1] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Fourth Edition 2020.

[2] Sutton, R. and Barto, A. *Reinforcement Learning: An Introduction*. MIT Press, 2017.

[3] Junling Hu and Michael P. Wellman. *Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm*. University of Michigan, Ann Arbor, MI 48109-2110, USA.