Value Iteration and Q-Learning Algorithms

Kajetan Frąckowiak

26.03.2025

1 Value Iteration Algorithm

Valute Iteration is a reinforcment learning method that updates state values iteratively using the Bellman equation:

$$V^(s) = \max_a \sum_{s'} P(s'|s,a) \left[R(s,a,s') + \gamma V^(s') \right] \tag{1}$$

where:

- $V^*(s)$ is the optimal value function.
- $P(s'|s,a)$ is the transition probability.
- $R(s,a,s')$ is the reward.
- γ is the discount factor.

The algorithm iterates by computing action values and updating state values accordingly.

2 Q-Learning Algorithm

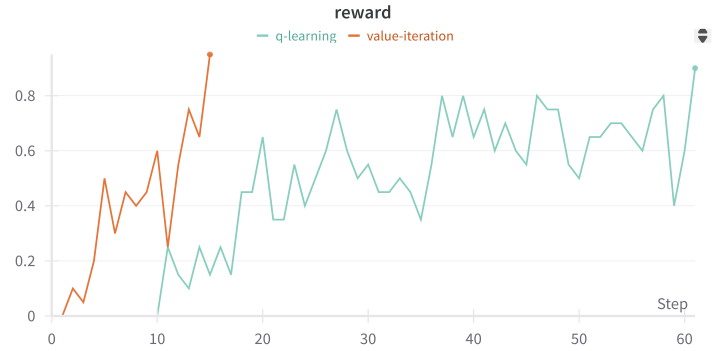
Q-Learning is a model-free reinforcment learning algorithm that updates the state-action value function using:

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

where:

- $Q(s,a)$ is the action-value function.
- $R(s,a,s')$ is the immediate reward.
- γ is the discount factor.
- $max_{a'} Q(s',a')$ estimates future rewards.

The following chart presents the history of rewards obtained during training with Value Iteration and Q-Learning.



3 Summary

As we can see, both Value Iteration like Q-Learning work well without gradient on FrozenLake-v1 with very few iterations. We plan to test it on other environments else in future.