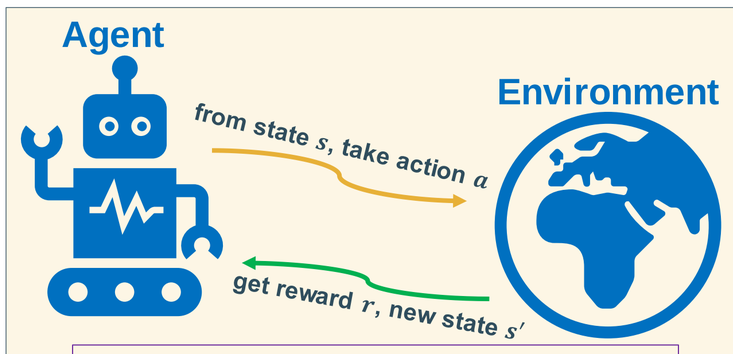
A generalisation of the Bellman Equation in Epistemic Reinforcement Learning

Keivan Shariatmadar, Jacob Golub, Adam Faza, David Moens
Mecha(tro)nic System Dynamics (LMSD), KU Leuven, Belgium
Workshop on Symbolic XAI, May 23rd, 2024 - Auditorium 6, TU Eindhoven
A joint workshop by EAISI, Leuven.AI, and RWTH AI Center
{*keivan.shariatmadar,jacob.golub,adam.faza,david.moens*}*@kuleuven.be*

Abstract

In reinforcement learning, when a state value is unknown, we use probabilistic/stochastic MDP. Lots of work has been done regarding this problem. However, if the probability is not unique and changing, i.e., the uncertainty is epistemic, or the agent still needs to meet the state, we have less information. In this case, a novel idea is to use an epistemic uncertainty model and solve the MDP via an approach by the generalisation of the Bellman equation. In this poster, we will present this idea and show numerical results on a simple toy example.

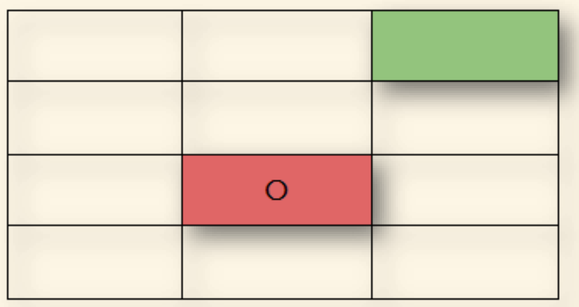
Classical Reinforcement Learning



An **environment** with an **agent**

We pursue our dreams through various **actions**, some **rewarding** and some not. By exploring different **paths** and using **feedback** to refine our **strategies**, we gauge our progress. Our mental states continuously adjust to reflect our proximity to our goals.

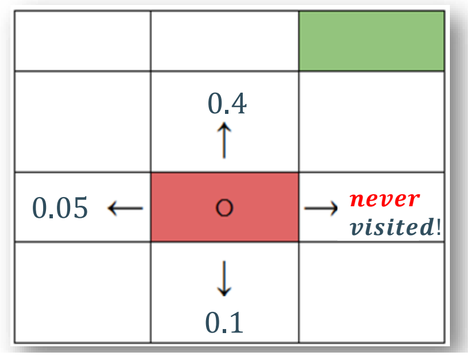
Let's consider, that a robot has a slight chance of dysfunctioning and might take the left right or bottom turn instead of taking the upper turn in order to get to the green room from where it is now (red room). Now, the question is how do we enable the robot to handle this when it is out there in the given environment?



Consider the robot is currently in the **red** room and it needs to go to the **green** room

Epistemic Uncertainty

Scenarios:
- A state which is less visited
- A new state
- A state with changing probability
- A moving obstacle in the environment
- Non-probabilistic value/reward for a state
- Non-stationary Reinforcement Learning
- The first episode



An Environment with an **Epistemic** Uncertainty

Aleatoric vs Epistemic

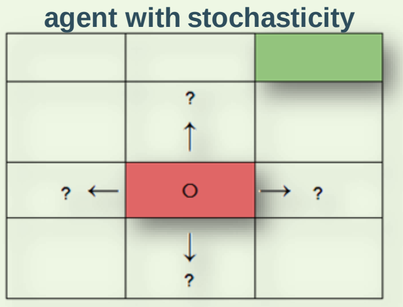
Aleatoric (statistical) uncertainty refers to randomness or variability: What a random sample drawn from a probability distribution will be.

Epistemic (systematic) uncertainty refers to the lack of knowledge: What the relevant probability distribution is.

Bellman Equation (Deterministic case)

$$V(s) = \max_a \left(R(s,a) + \gamma V(s')\right)$$

An **environment** with an **agent with stochasticity**



- s = a particular state (room)
- a = action (moving between the rooms)
- s′ = state to which the robot goes from s
- γ = discount factor (we will get to it in a moment)
- R(s, a) = a reward function which takes a state s and action a and outputs a reward value
- V(s) = value of being in a particular state (the footprint)

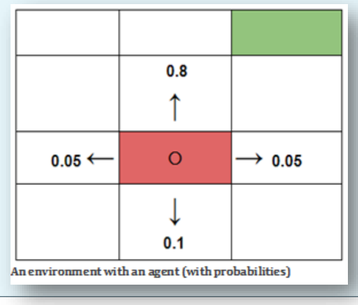
Bellman Equation (non-deterministic case)

$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s')\right)$$

- $P(s,a,s')$ is the probability of moving from room s to room $s'$with action a
- $\sum_{\acute{s}} \boldsymbol{P(s,a,s')V(s')}$ is the expectation of the situation that the robot incurs randomness

When we associate probabilities to each of these turns, we essentially mean that there is an 80% chance that the robot will take the upper turn. If we put all the required values in our equation, we get:

$$V(s) = \max_a(R(s,a) + \gamma((0.8V(room_{up})) + (0.1V(room_{down})) + \ldots))$$



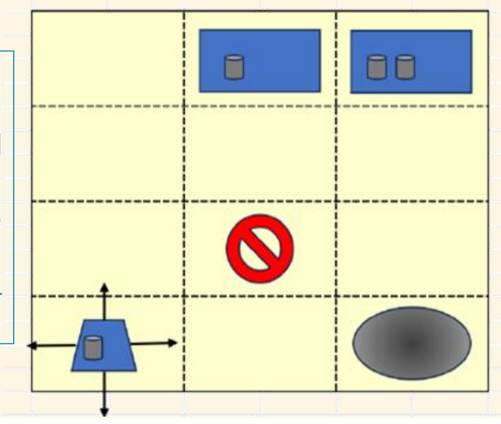
An environment with an agent (with probabilities)

Generalisation of Bellman Equation

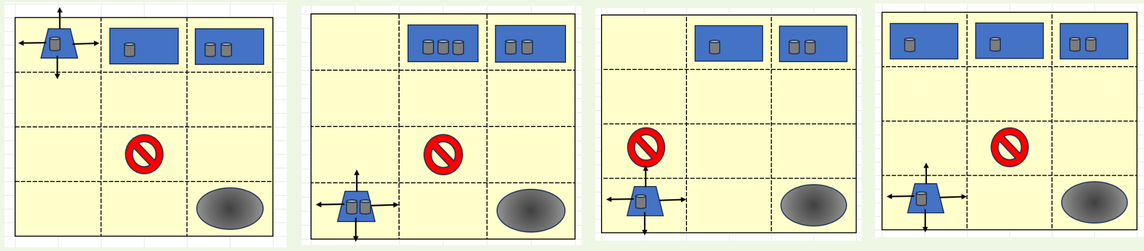
$$V(s) = \max_a \left(R(s,a) + \gamma \underline{E}_{s'} V(s,a,s')\right)$$

- $\underline{E}_{s'} V(s,a,s')$ is the Epistemic Uncertainty model over the value function moving from room s to room $s' \in S$ with action a and S is the epistemic uncertainty set.

Simple toy problem
- A robot starts from the left bottom corner.
- There is a barrier in the middle with a red sign.
- The upper blue areas are unloading positions (which are occupied).
- The right bottom corner is the state with a punishment

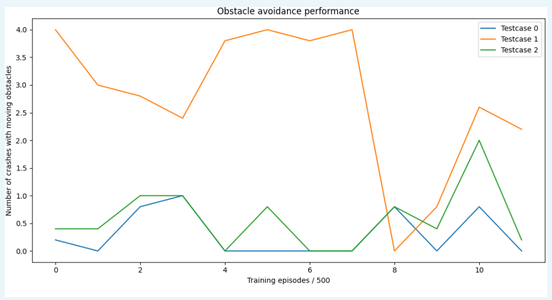


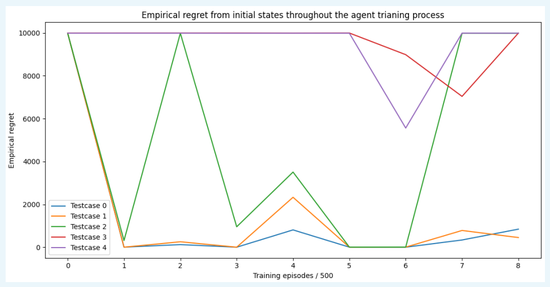
Test cases



Results – Obstacle Avoidance

Deterministic – Rainbow DQN





Uncertainty Aware DQN - Bootstrapped

Results – Rainbow vs Bootstrapped DQN

The uncertainty-aware approach is computationally more expensive and requires an adjusted epsilon-decay schedule but is more stable and demonstrates better feature learning.

Both approaches learn a strategy for the reachable states very well.

The rainbow shows signs of over-training, and less uncertainty aware.

The uncertainty-aware approach demonstrates that it is better at learning strategies for unreachable states, particularly when the environment is the same.

Neither strategy could learn to handle every **unseen state**. Our current research focuses on **Epistemic Reinforcement Learning** to solve and deal with the epistemic uncertainty in the problem.

This work is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 964505