Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Deep W-Networks: Solving Multi-Objective Optimisation Problems With Deep Reinforcement Learning

15th International Conference on Agents and Artificial Intelligence, Lisbon, Portugal

Jernej Hribar, Luke Hackett, and Ivana Dusparic
School of Computer Science and Statistics, Trinity College Dublin, Ireland
23/02/2023

Multi-Objective Optimisation Problems

Background

- **These problems require balancing trade-offs between objectives to find a compromise solution that satisfies all constraints.**
- **Many real world problems can be formulated as a multi-objective optimisation problem:**
  - Radio resource management;
  - infectious disease control;
  - marketing optimization in advertising;
  - energy management of sensor networks.

Pareto Front

Background

- **Is defined as the set of non-dominated solutions;**
- **Each objective is considered as equally good;**
- **Provides a way to visualize the trade-offs.**

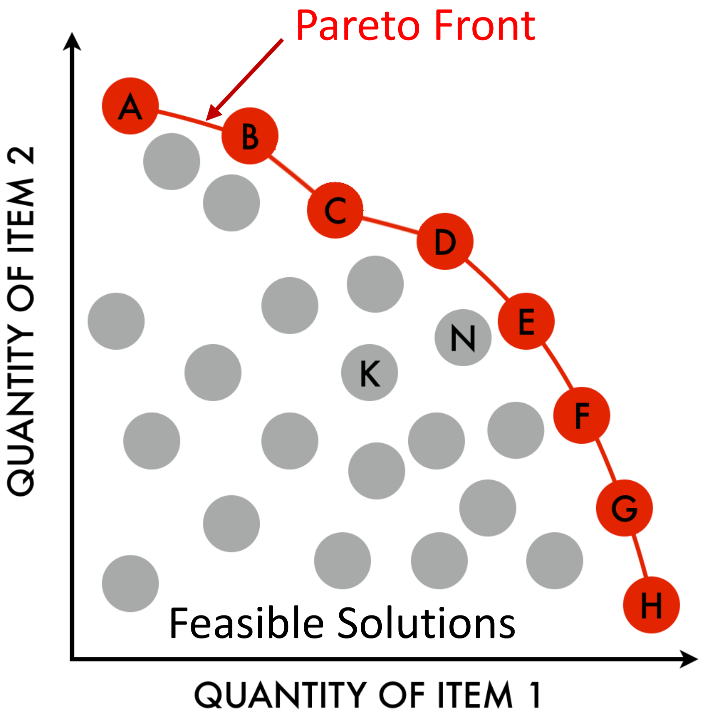


Image taken from: Wikipedia

W-Learning

Solution to Resolve Multi-Objective Problems

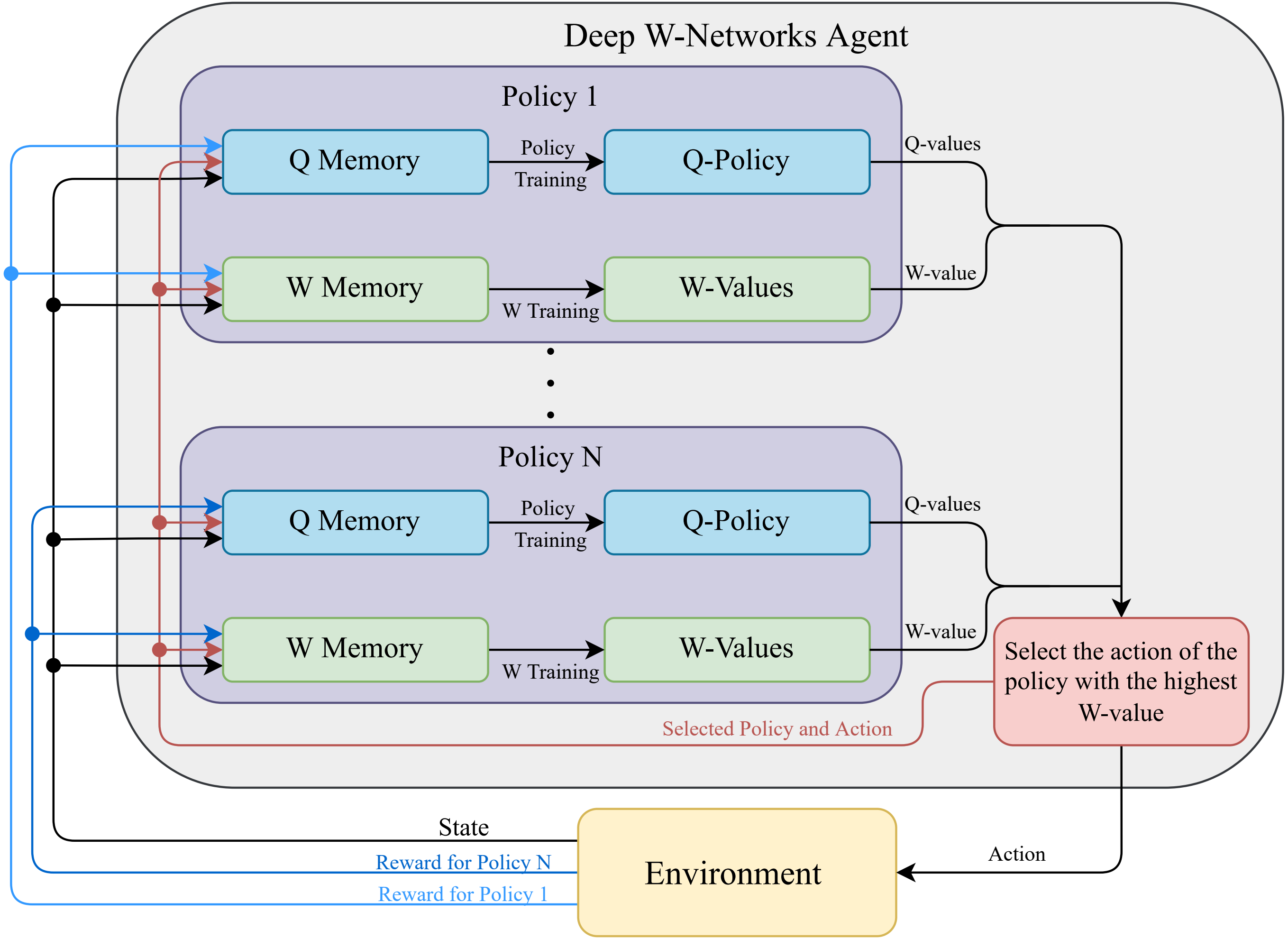
- **Developed in the 1990s [1] to find the optimal policy for a multi-objective problems.**
- **Resolves competition between different policies, with the winner policy being the one that is most likely to suffer the most if it does not win.**
- **Each policy is implemented with a tabular Q-learning, and W-values representing the W weight.**
- **Computationally efficient, intuitive, versatile applicability, etc.**

[1] Humphrys, M. (1995). W-learning: Competition among selfish Q-learners.

Aims and Objectives of Our Work

Contributions

- **We propose a Deep W-Networks (DWN), a deep learning extension to W learning algorithm;**
- **DWN resolves the competition between greedy single-objective policies by relying on W-values;**
- **We show the modularity of DWN.**


Deep W-Networks Agent
Policy 1
Q Memory
Policy
Training
Q-Policy
Q-values
W Memory
W Training
W-Values
W-value
Policy N
Q Memory
Policy
Training
Q-Policy
Q-values
W Memory
W Training
W-Values
W-value
Select the action of the policy with the highest W-value
Selected Policy and Action
State
Reward for Policy N
Reward for Policy 1
Environment
Action

Deep W-Networks

Algorithm

- **We employ two DQNs for each objective.**
- **The agent takes the action suggested by the policy associated with the highest W-value:**

$$W_j(t) = \max\left(\{W_1(t), \ldots, W_N(t)\}\right).$$

Deep W-Networks

Training W-values

- **W-values are updated similarly to Q-values:**

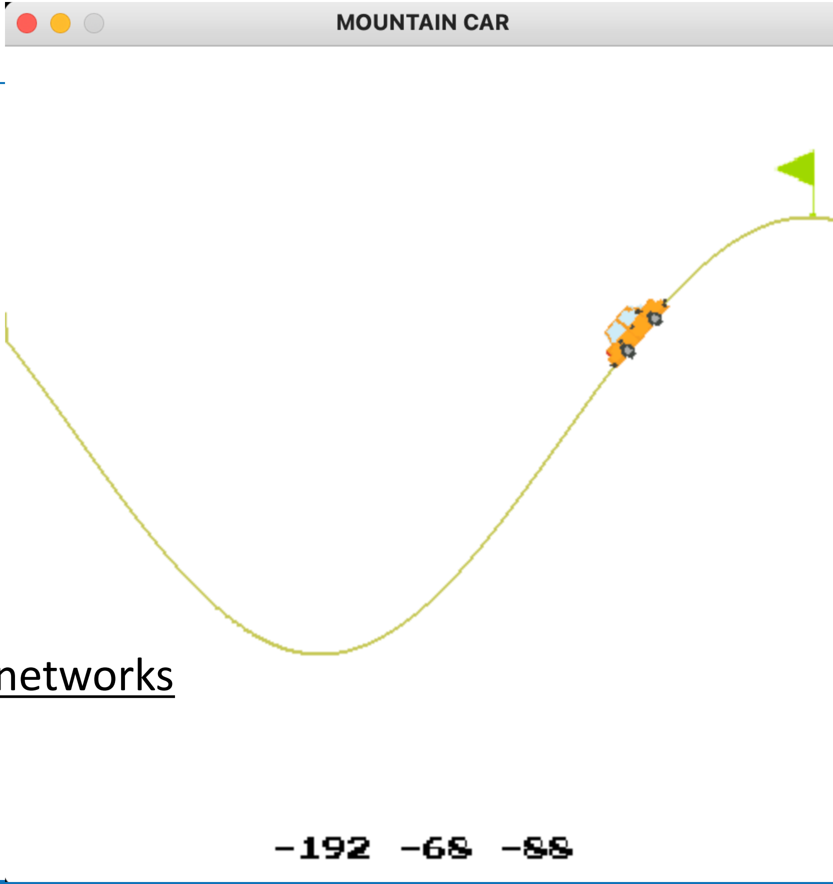
$$W_i(t) \leftarrow (1-\alpha)W_i(t) + \alpha\left[Q(s(t), a_j(t)) - (R_i(t) + \gamma \max_{a_i(t+1)\in\mathcal{A}} Q(s(t+1), a_i(t+1))\right].$$

- **W policy saves the experience only when it was not selected.**

Mountain Car

Environment

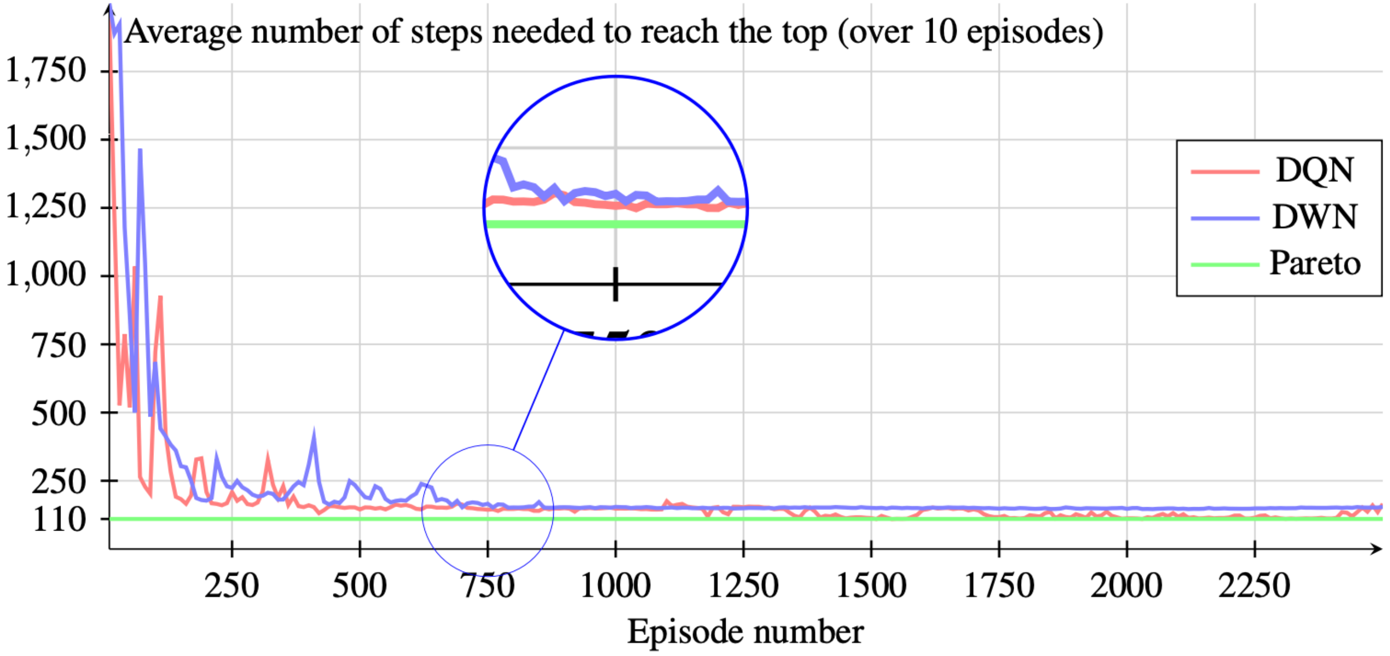
- **The environment has three different objectives:**
    - time penalty;
    - backward acceleration penalty;
    - forward acceleration penalty.



- **Github code:** https://github.com/deepwlearning/deepwnetworks

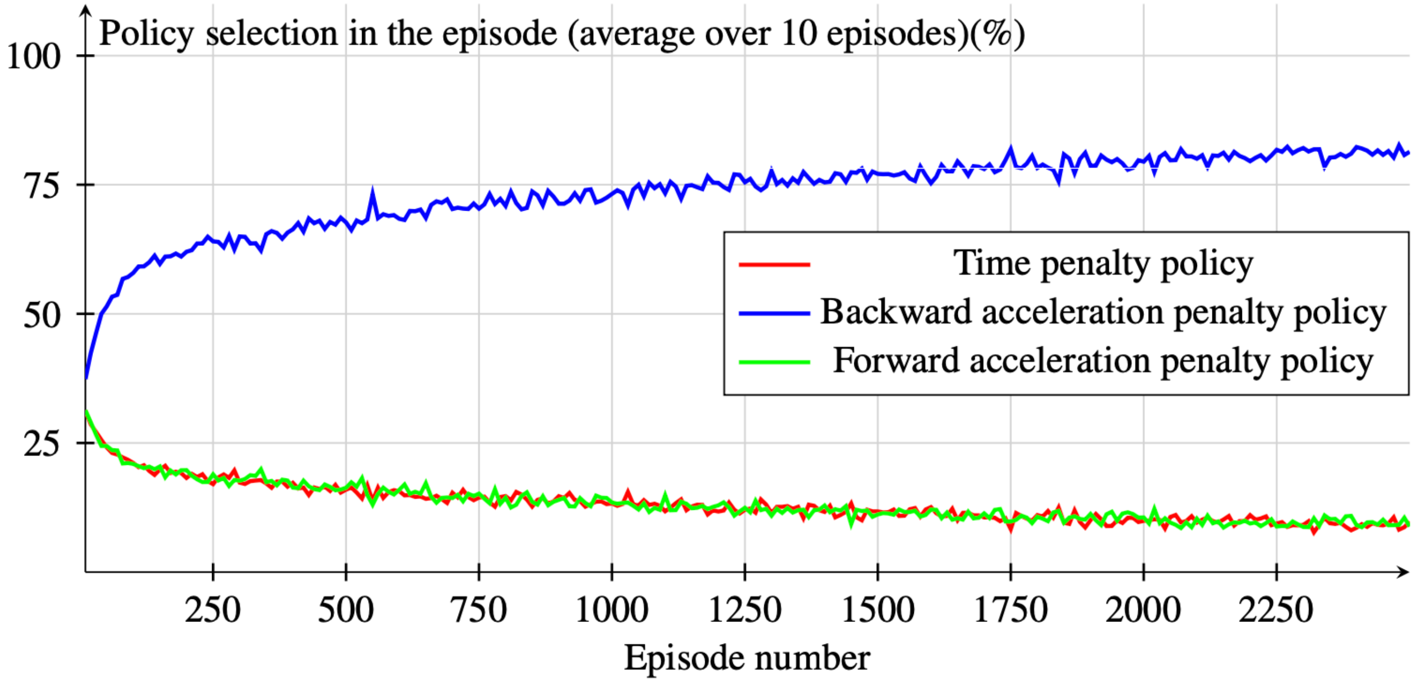
Mountain Car

The number of steps to finish the episode



Mountain Car

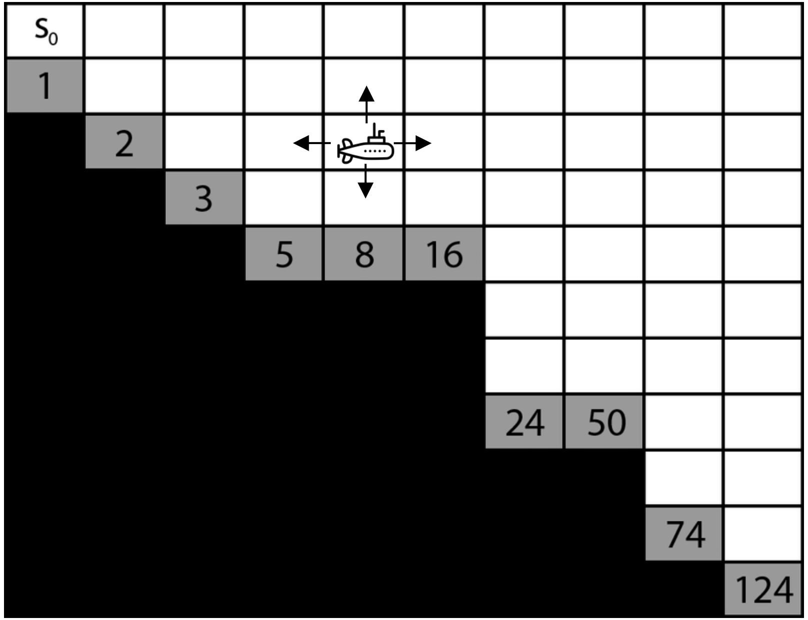
The percentage of each policy DWN agent selects in an episode,



Deep Sea Treasure

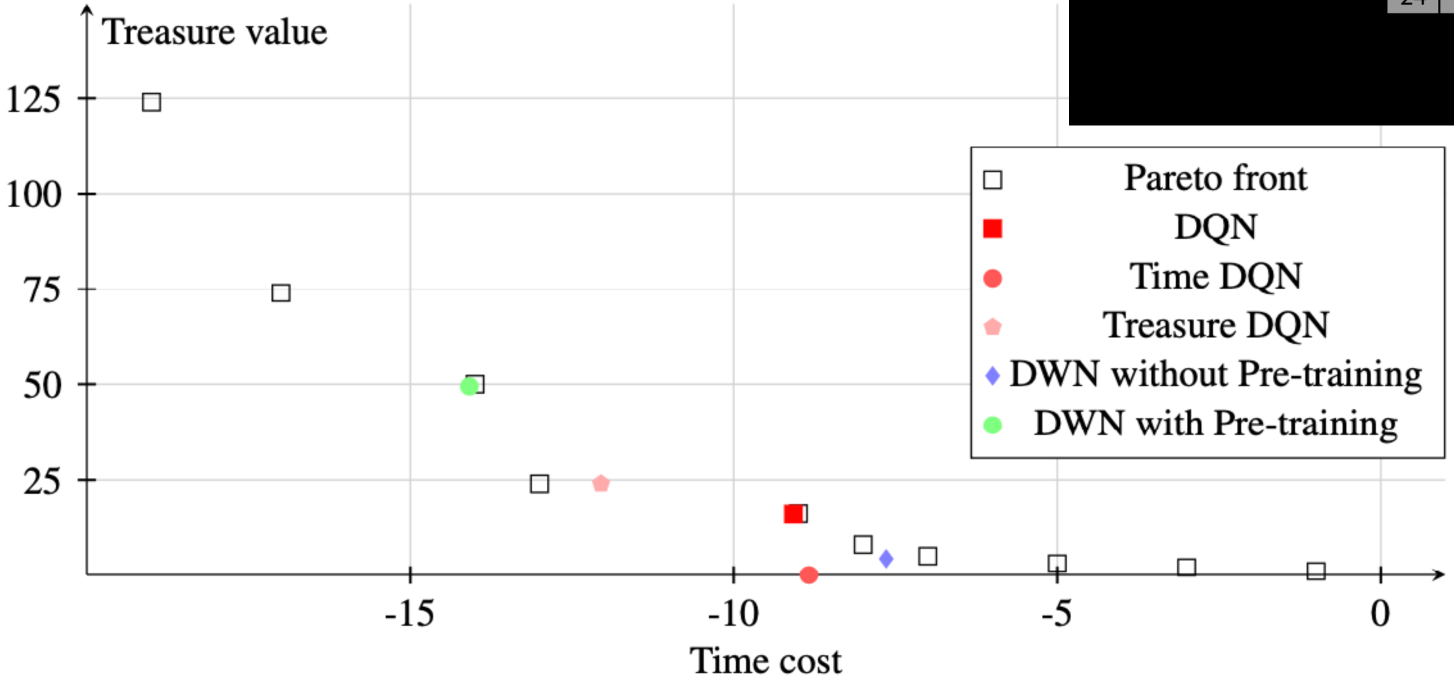
Environment

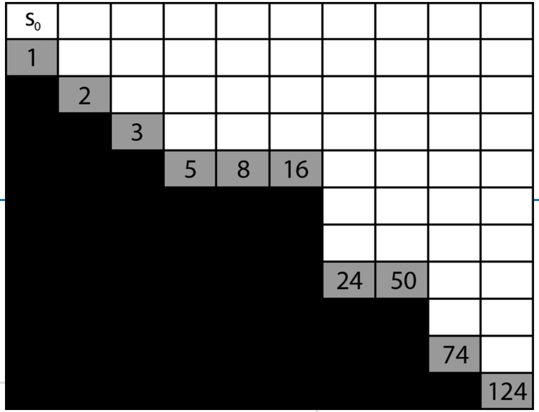
- **The environment has two objectives:**
  - Time penalty;
  - Collected treasure.
- **We use Convolutional Neural Network (CNN) structure in DWN Agent**



Deep Sea Treasure

Pareto Front





Conclusion
and Future Work

- **The proposed DWN is capable of finding the Pareto front.**
- **The main advantage of DWN is its ability to train multiple policies simultaneously.**
- **Future work:**
    - Improving the computational performance;
    - Evaluating in more complex environments.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Thank you for your attention!

Email: **jhribar@tcd.ie** or **jernej.hribar@ijs.si**