University of the West of England BRISTOL

BAE SYSTEMS

Learning Hierarchical Fuzzy Rule-based Systems for a Mobile Robot Controller

Antony Waldock

antony.waldock@baesystems.com

antony.waldock@uwe.ac.uk

Outline

- Reinforcement Learning
- RL for Robot Control
- Variable Resolution Discretization
- Hierarchical Fuzzy Q-Learning
- Results
- Discussion
- Conclusions and Future Work

University of the West of England BRISTOL

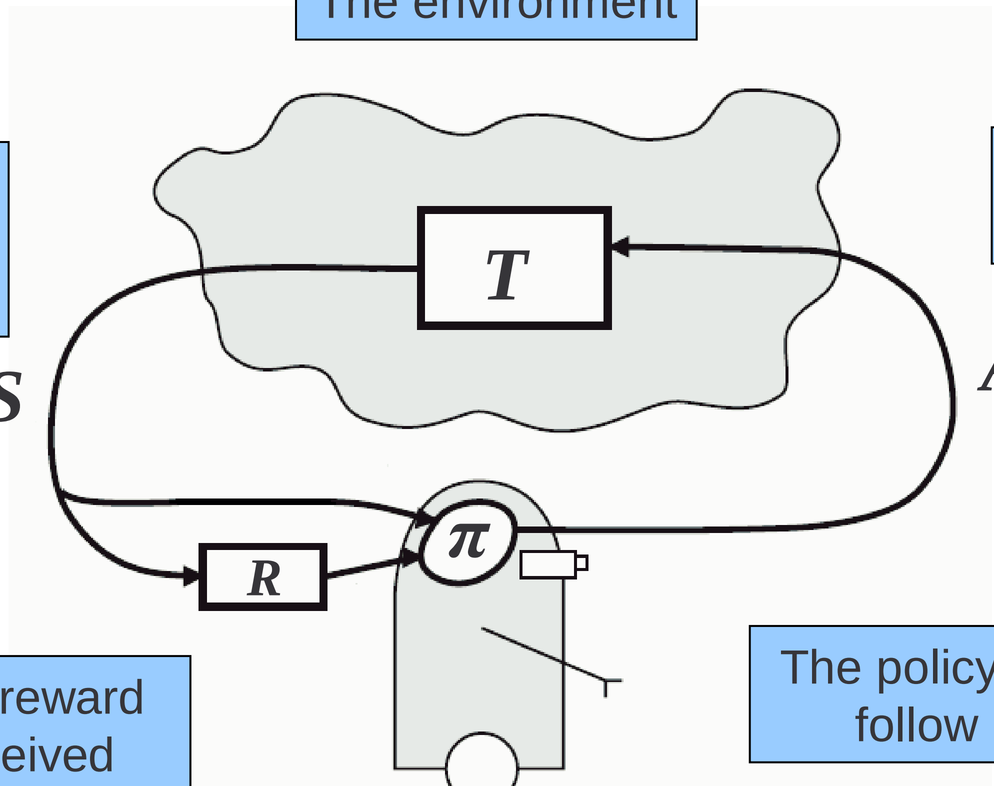
BAE SYSTEMS

Reinforcement Learning (RL)

The environment

The current state of the environment

The action to perform



Kaelbling, Litman and Moore 1996

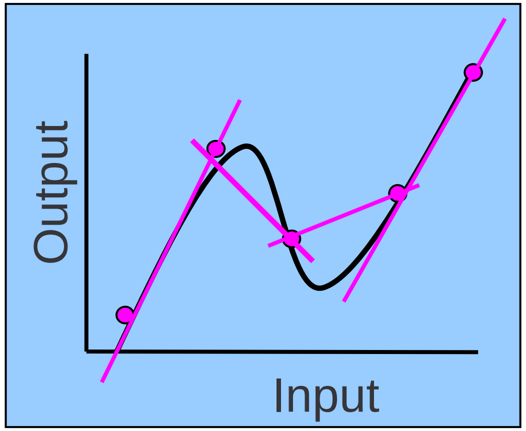
The reward received

The policy to follow

RL attempts to learn a policy that maximises the expected long-term reward received

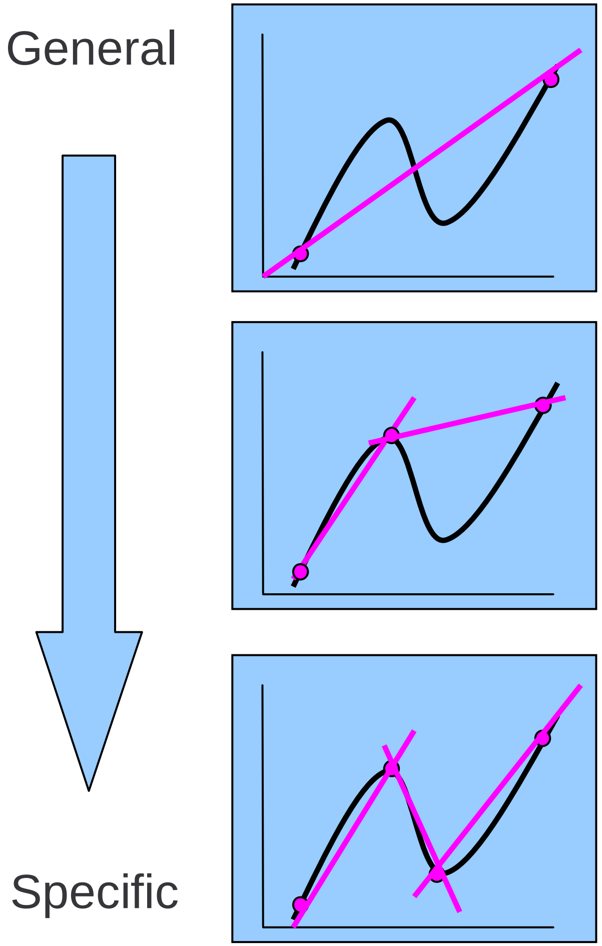
This model assumes the world is Markovian

Reinforcement Learning for Robot Control



- Robots operate in continuous environments using continuous sensors and actuators.
- Generalisation is essential!
- Function Approximation techniques can be used to cope with continuous states and actions.
  - Neural Networks
  - Linear Weighted Regression
  - Fuzzy Rule Based Systems
- The majority of techniques generalise uniformly across the state and action space.

Variable Resolution Discretization

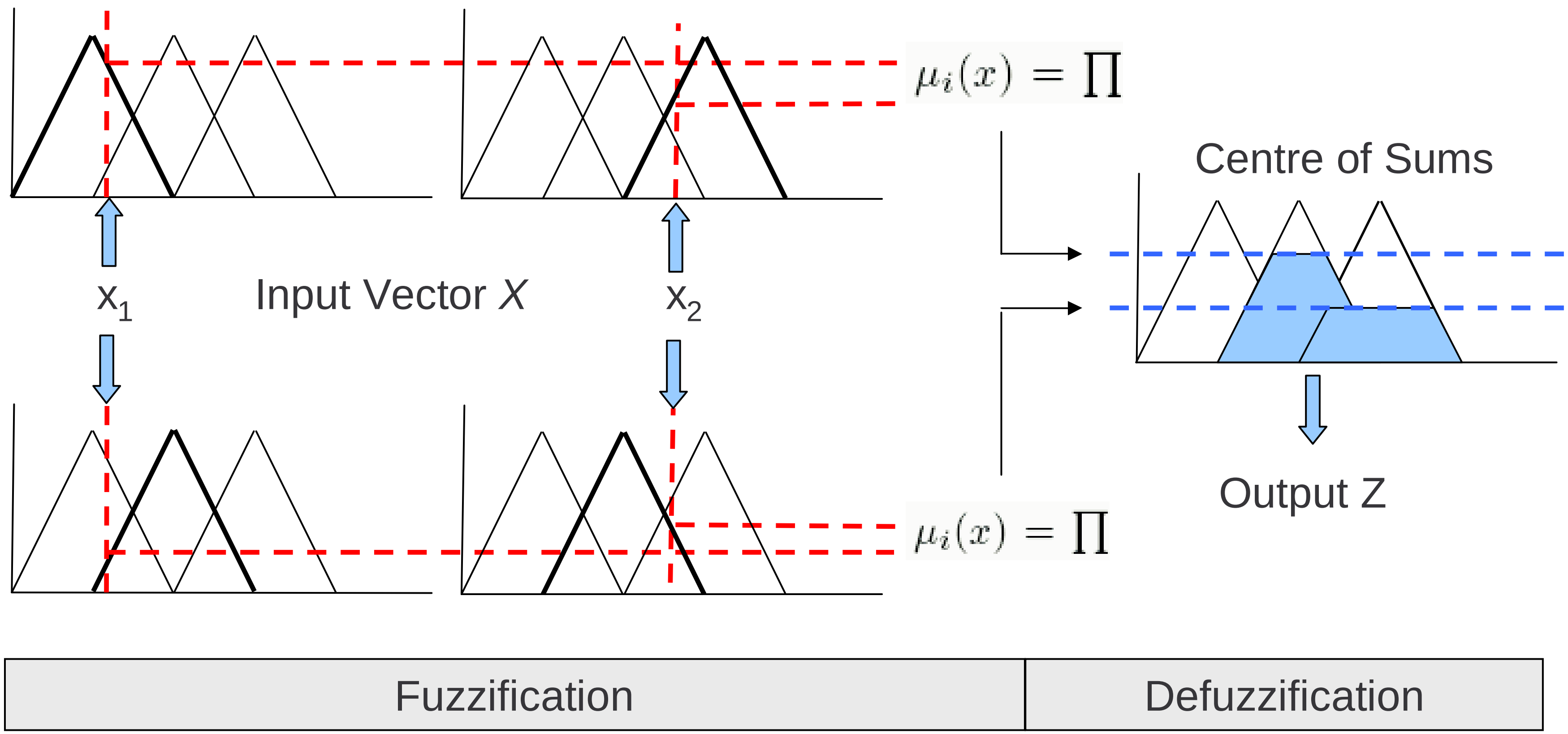


- Variable Resolution Discretization varies the generalisation in an attempt to minimise the approximation error.
- The technique moves from a ‘*general to specific*’ representation by successfully refining areas of the state space.
- Any technique must define:
  - Splitting Criterion (What to split)
  - Stopping Criterion (When to stop)
- Related Work
  - Optimal Control [Munos and Moore]
  - Decision Trees [Pyeatt]
- The focus of this work is to use variable resolution discretization with reinforcement learning for robot control

Fuzzy Q-Learning

Fuzzy Q-Learning provides a framework for Q-Learning in a continuous state and action space.

IF x_1 **AND** x_2 **THEN** Z

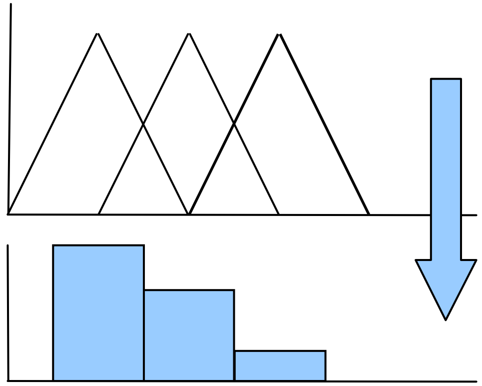


Fuzzy Q-Learning Algorithm

Output Membership Functions

Q Values

$q[i, j]$



Algorithm 1 Fuzzy Q-Learning algorithm

repeat
    Observe the input vector x
    Select actions using an Exploration/Exploitation Policy (EEP)
    Compute the global consequence $a(x)$ and the Q-value $Q(x, a)$
    Apply the action $a(x)$
    Receive the reinforcement r
    Observe the input vector y
    Update the Q-Value using (6)
until End

- Each rule *i* has a *Q* value associated with each action *j*

- The *Q* value for an input vector *x* (sensor values) can be calculated using:

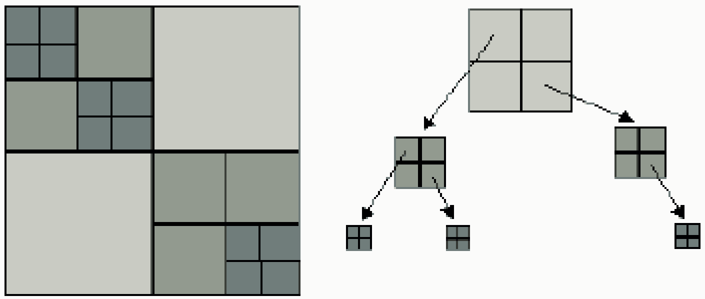
$$Q(x, a) = \frac{\sum_{i=0}^{N} \mu_i(x) \times q[i, i']}{\sum_{i=0}^{N} \mu_i(x)}$$

- The *Q* values are updated using an extension of the standard *Q*-Learning algorithm

$$q[i, i'] = q[i, i'] + \alpha \frac{\mu_i(x)}{\sum_{i=0}^{N} \mu_i(x)} \Delta Q$$

ΔQ is $(r_{t+1} + \gamma V(y) - Q(x, a))$

Hierarchical Fuzzy Q-Learning

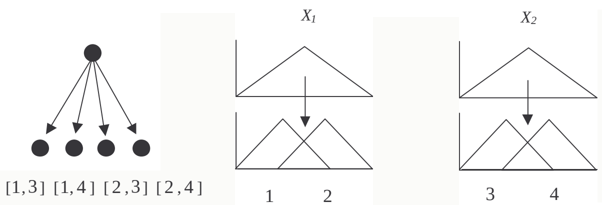


(i) (ii)

Identification

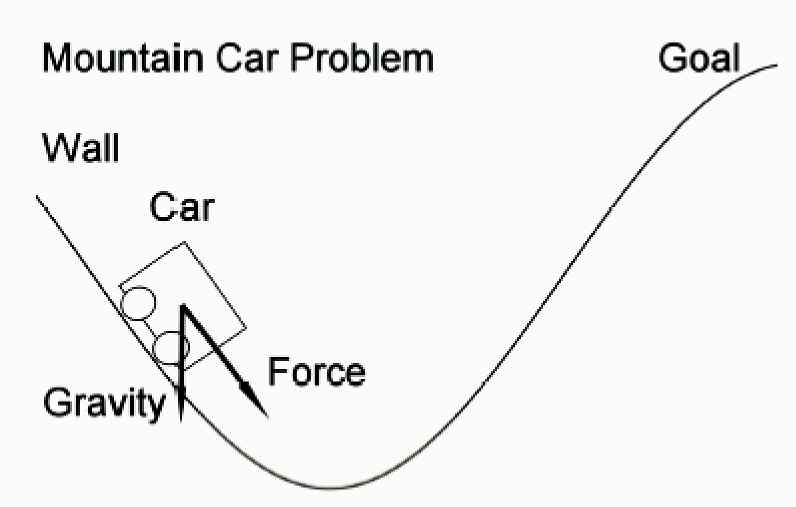
$$SD_{i,j} = \sqrt{\frac{\sum_{t=0}^{n} q_t[i,j]^2 - \frac{(\sum_{t=0}^{n} q_t[i,j])^2}{n}}{n}}$$

Specialisation



- Hierarchical Fuzzy Rule-Based Systems (HFRBS) provides a framework to incorporate variable resolution discretization with Fuzzy Q-Learning.
- A HFRBS uses an expansion policy to identify and specialise inaccurate rules (splitting and stopping criterion).
- Fuzzy Q-Learning can be used to train a HFRBS
- In this work, we will use the variance of the *Q* value to identify inaccurate rules.

Mountain Car Problem



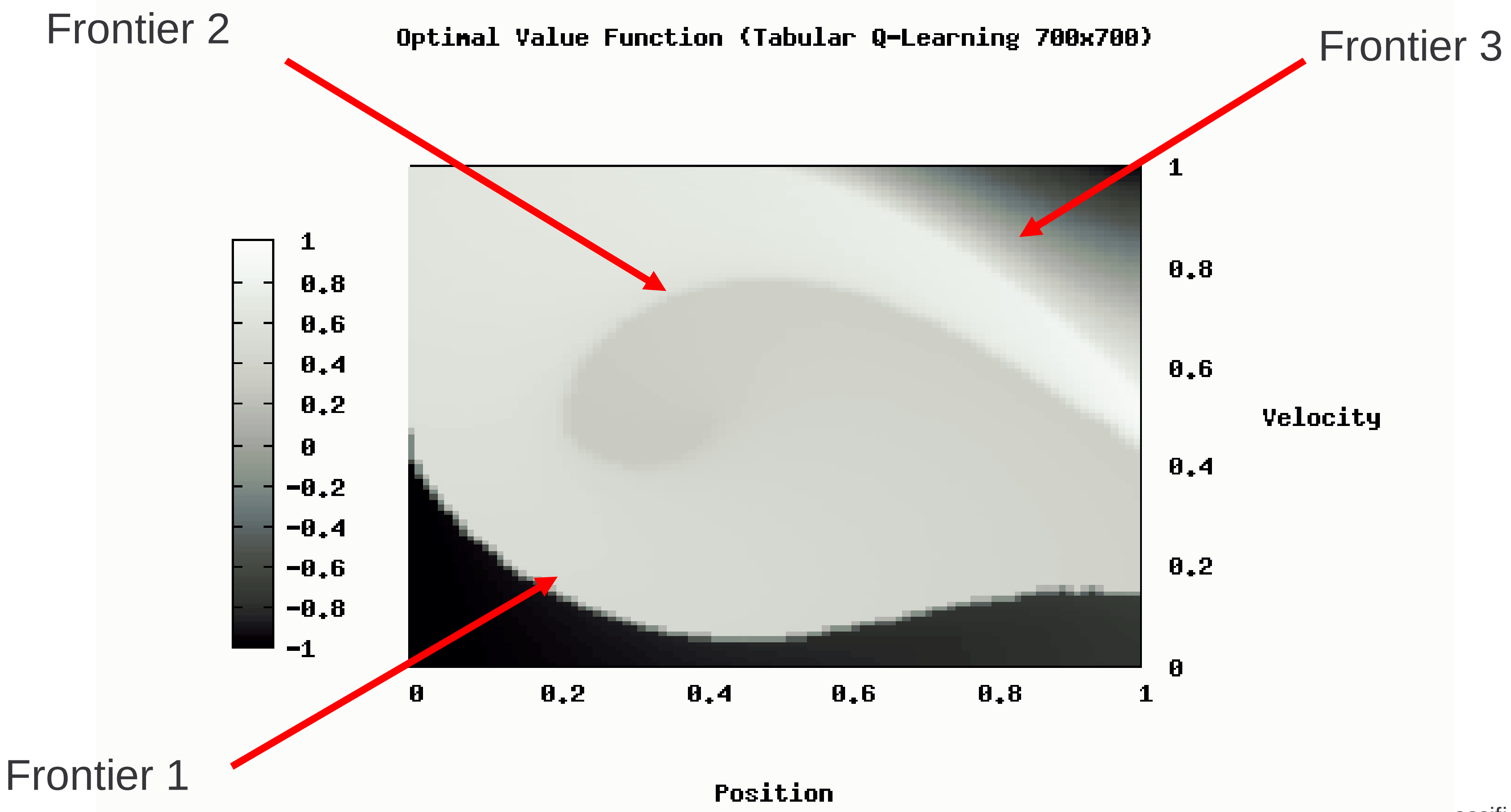
The goal of the mountain car problem is to park the car at the top of the hill (with zero velocity) without hitting the wall on the opposite side.

$$R_t = \begin{cases} -1 & \text{if } x_t \leq \text{-1.2} \\ 0 & \text{if -1.2} > \; x_t \; < 0.5 \\ 1 - \frac{|dx_t|}{0.07} & \text{if } x_t \geq 0.5 \end{cases}$$

Table 4.1: Benchmark Results for the Mountain Car Problem

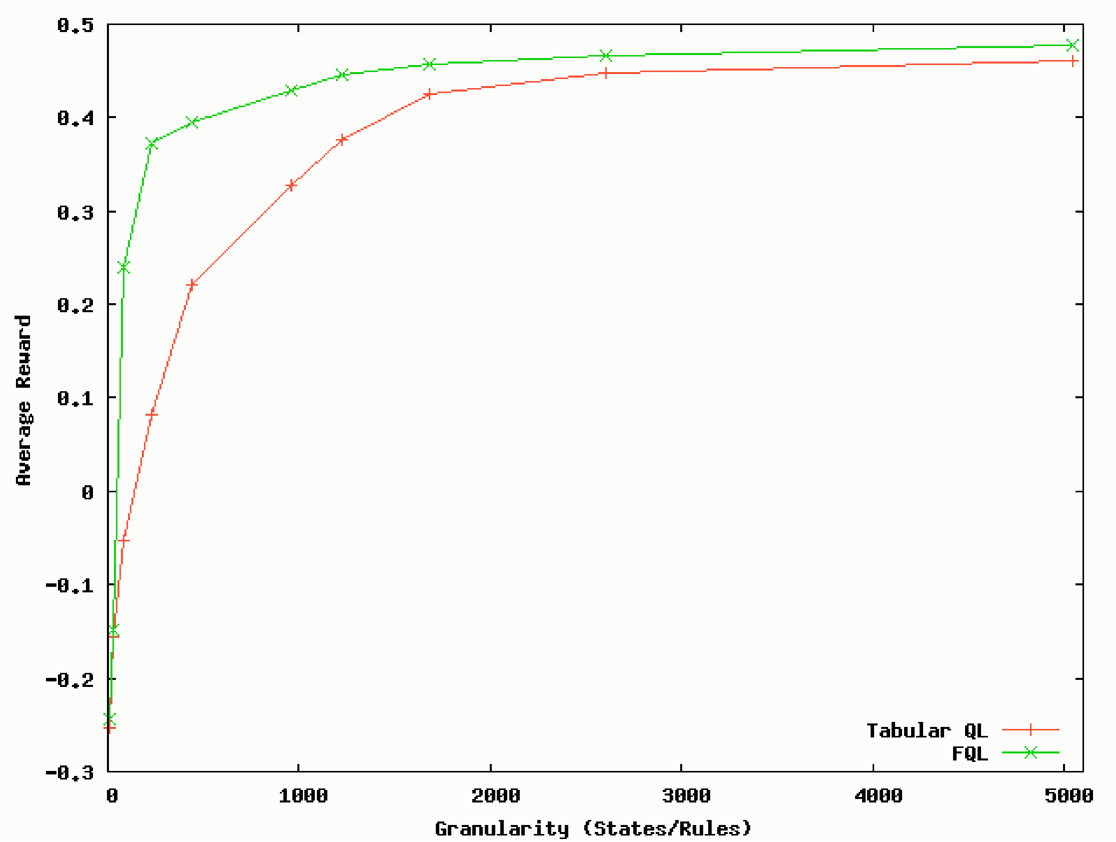
Method	States	Steps	Completed	Reward
Left	-	189.55	0.054	-0.701
Zero	-	185.31	0.081	-0.476
Right	-	162.29	0.254	-0.213
Tabular Q-Learning (257x257)	66049	104.17	0.736	0.403
Tabular Q-Learning (500x500)	250000	99.72	0.763	0.474
Tabular Q-Learning (700x700)	490000	99.95	0.763	0.484

Optimal Value Function

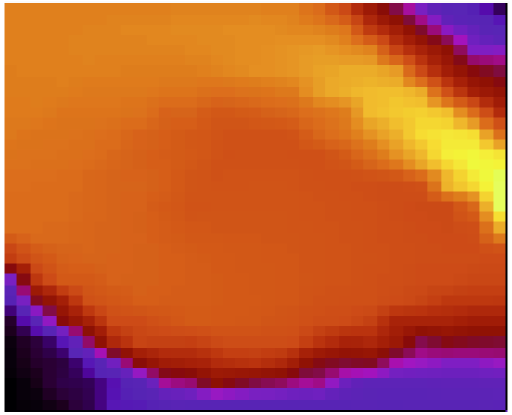


assified

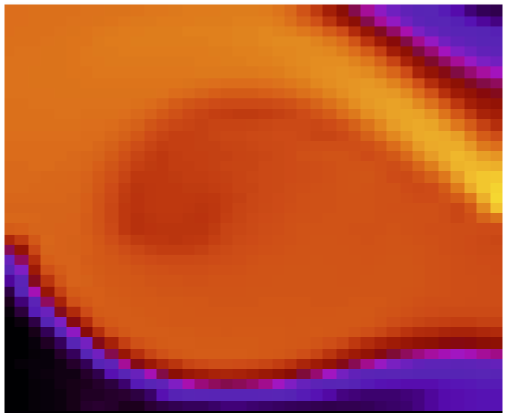
Tabular vs Fuzzy Q-Learning



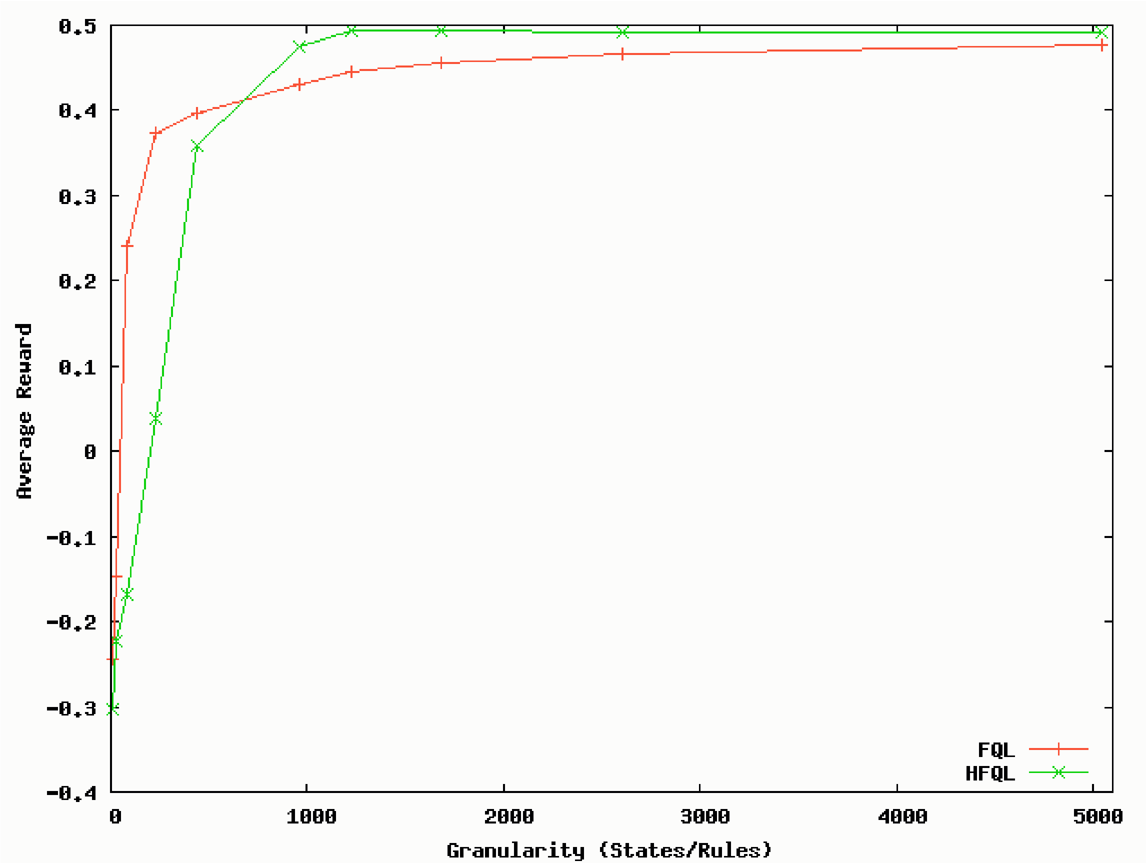
Tabular Q-Learning (441)



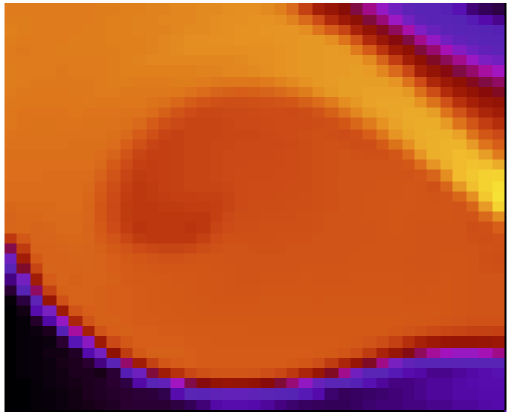
Fuzzy Q-Learning (441)



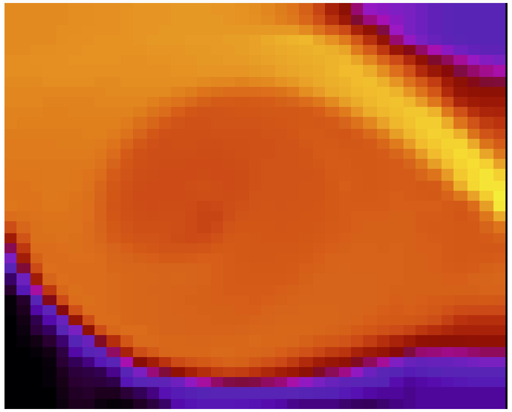
Uniform vs Variable Resolution



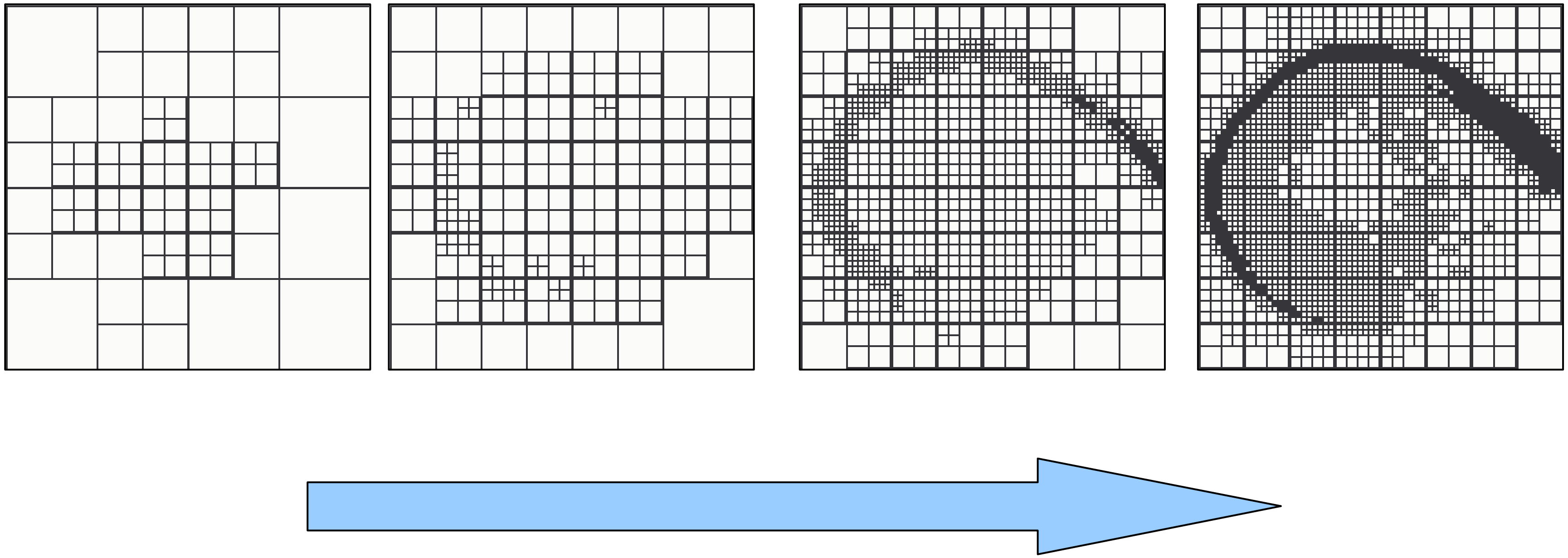
Fuzzy Q-Learning (1225)



Hierarchical FQL (1225)

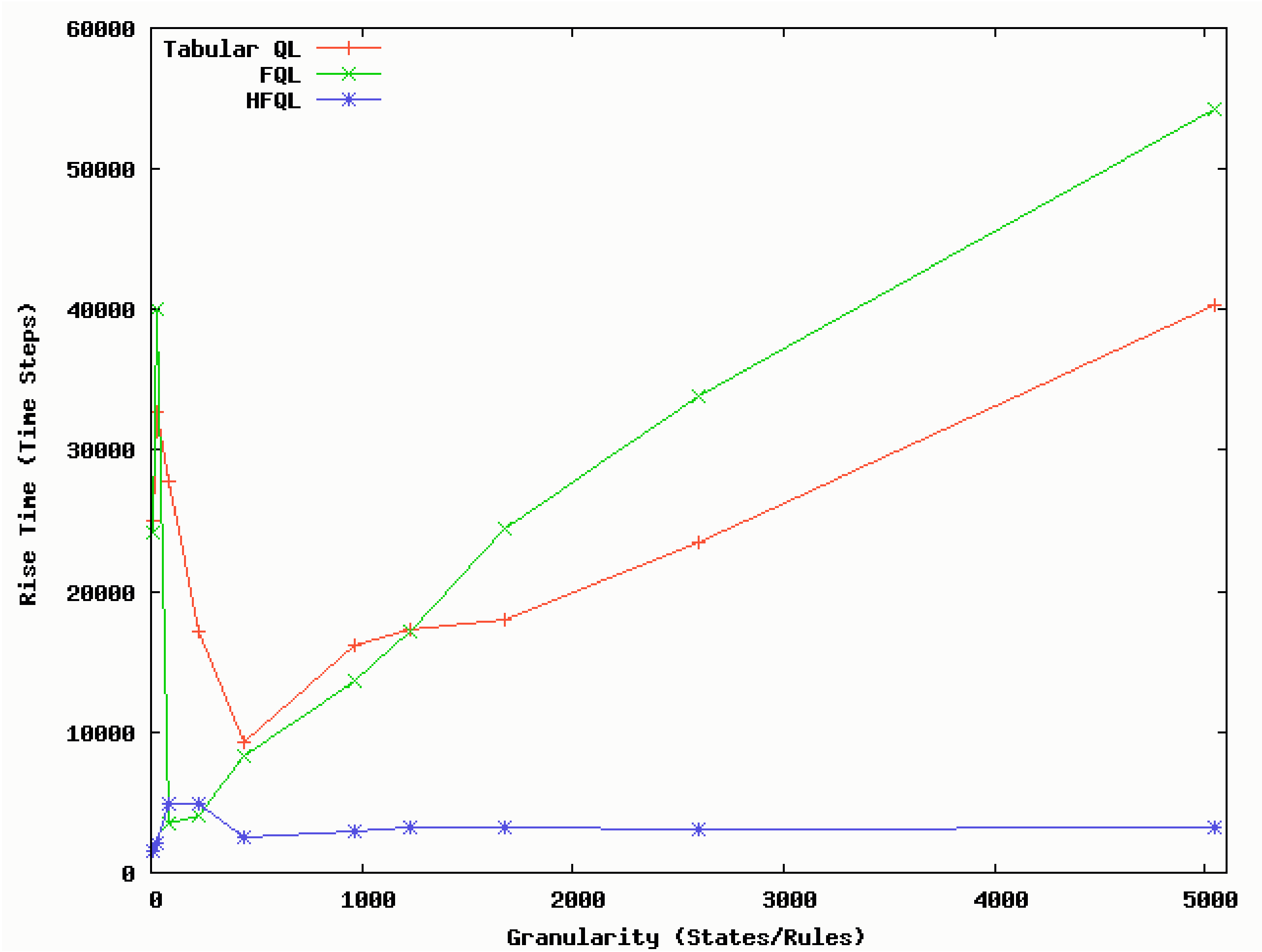


How does Variable Resolution improve performance?



Increasing Granularity

Learning Rate



Discussion

- The variance measure (raw score) must allow for convergence and sampling.
  - Convergence Time: The *Q* value requires a period of time to converge (or not!)
  - Sampling Time: The time taken to establish an consistent estimate of the variance of the *Q* value
- Magic Numbers!
- Specialisation is one shot.
- Trade-off between improved specialisation and memory requirements.

Conclusion and Future Work

- Introduced the concept of Variable Resolution Discretization
- Introduced Hierarchical Fuzzy Q-Learning as a framework for investigating Variable Resolution Discretization for RL in continuous domains.
- Demonstrated the performance benefits of an expansion policy based on the variance in the value function approximation.

- More RL problems with more expansion policies
- Eliminate (or understand) the magic numbers

Questions?