Relevance Determination in Reinforcement Learning

Katharina Tluk v. Toschanowitz[1], Barbara Hammer[2], Helge Ritter[1]
[1]University of Bielefeld, [2]Clausthal University of Technology

1 Reinforcement Learning

- Learn optimal policy: states S $\rightarrow$ actions A based on a reward r
- Based on Q-Learning (*Watkins*):

$$Q_{s_t,a_t} \leftarrow Q_{s_t,a_t} + \alpha \left[r_{t+1} + \gamma \max_a Q_{s_{t+1},a} - Q_{s_t,a_t} \right]$$

with learning rate α and discount factor γ

2 Relevance Determination

- Reduce size of state-action space by determining equivalence classes $\rightarrow$ faster learning
- theoretical basis: equivalence in MDPs
- application to Q-learning: clusters of Q-values

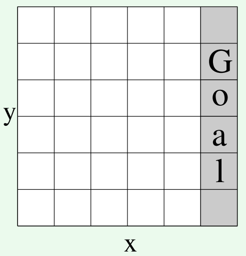
8	7	6	5	4	5	6	7	8
7	6	5	4	3	4	5	6	7
6	5	4	3	2	3	4	5	6
5	4	3	2	1	2	3	4	5
4	3	2	1	0	1	2	3	4
5	4	3	2	1	2	3	4	5
6	5	4	3	2	3	4	5	6
7	6	5	4	3	4	5	6	7
8	7	6	5	4	5	6	7	8

equivalent states in grid world with one goal

3 Approach 1: Variance-based

- Compute variances along dimensions over all other states and actions
  $\Rightarrow$ low variance indicates low relevance

state–space with redundant dimension



Q–learning $\rightarrow$

Q–values for optimal actions

0.1	1.0	2.1	3.0	4.0	5.1
0.1	0.9	2.0	2.8	3.9	4.9
0.0	0.8	1.9	2.9	4.0	5.0
0.2	1.1	2.0	3.1	4.1	5.2
0.0	0.9	1.8	2.9	3.8	4.9
0.0	0.8	1.8	2.8	3.9	4.8

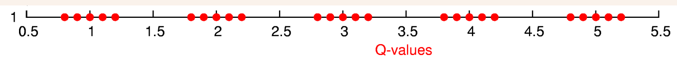
y (rows), x (columns)

4 Approach 2: Neural Gas-based

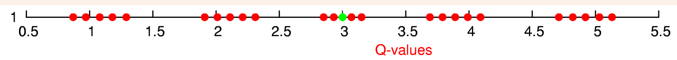
- standard update of visited (s_t, a_t)-pair
- update for all other (s, a):

$$\Delta Q_{s,a} = \eta \cdot (Q_{s_t,a_t} - Q_{s,a}) \cdot exp\left(-\frac{|Q_{s_t,a_t} - Q_{s,a}|}{2\sigma^2}\right)$$

with learning rate $\eta < \alpha$ and standard deviation σ.



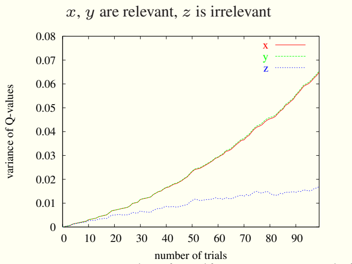
$\downarrow$ update with $Q(s_t, a_t) = 3$ $\downarrow$

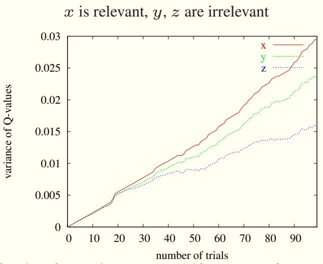


5 Test scenarios

- two dimensional grid-world with one goal
- add irrelevant dimensions

6 Results - Variance

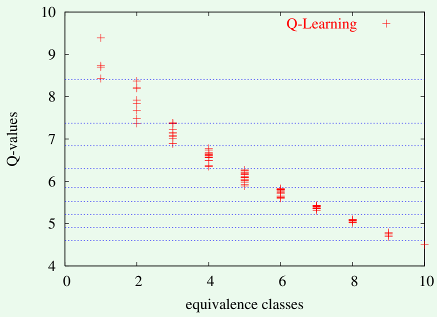




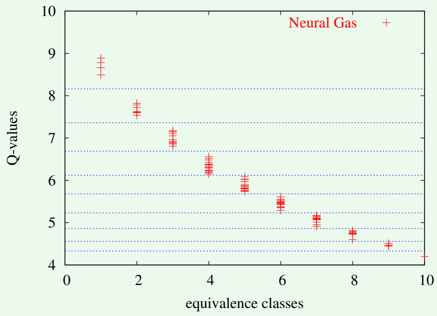
- correctly indicates multiple irrelevant dimensions

7 Results - Neural Gas

Q-values from regular Q-learning



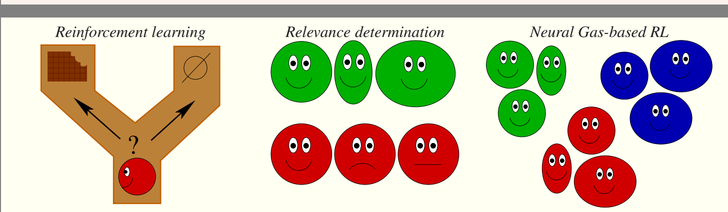
Q-values from Neural Gas-based learning



- facilitates correct clustering of Q-values

8 Further work

- more complex applications including cooperation (blockers task in progress)
- enhance algorithm, e.g. local variances
- integrate equivalence classes while learning



Contact: ktoschan@techfak.uni-bielefeld.de