Learning An Unknown Domain

Ni Lao 2011.11.11

ACRL presentation

Motivation

We want a general method that assumes no domain knowledge

- Unknown states and transitions to be discovered from interactions, e.g. SLAM

We want to study the role of forget in reinforcement learning

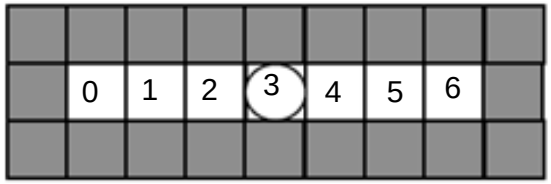
Task 1

The hall way maze

- -1 reward for each step
- +5 reward for reaching the goal
- -5 reward for not finishing after 6 steps

Evaluation

- the total reward for traces starting from each of the 6 non-terminal states (highest score possible is 12)



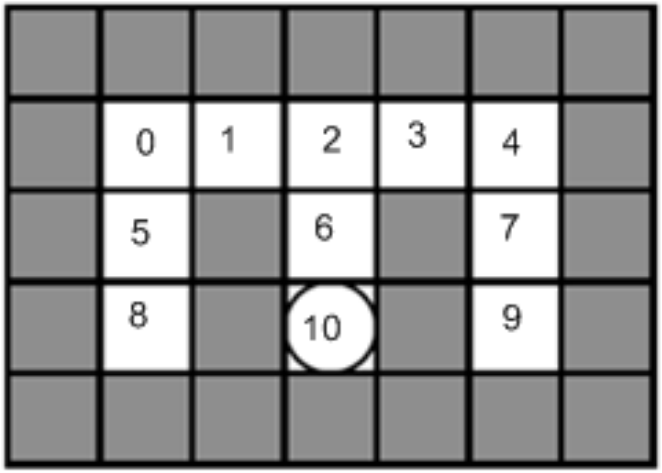
Task 2

The Cheese Maze

- -1 reward for each step
- +5 reward for reaching the goal
- -5 reward for not finishing after 10 steps

Evaluation

- the total reward for traces starting from each of the 10 non-terminal states (57 is the highest score possible)



Recurrent neural network

Value function

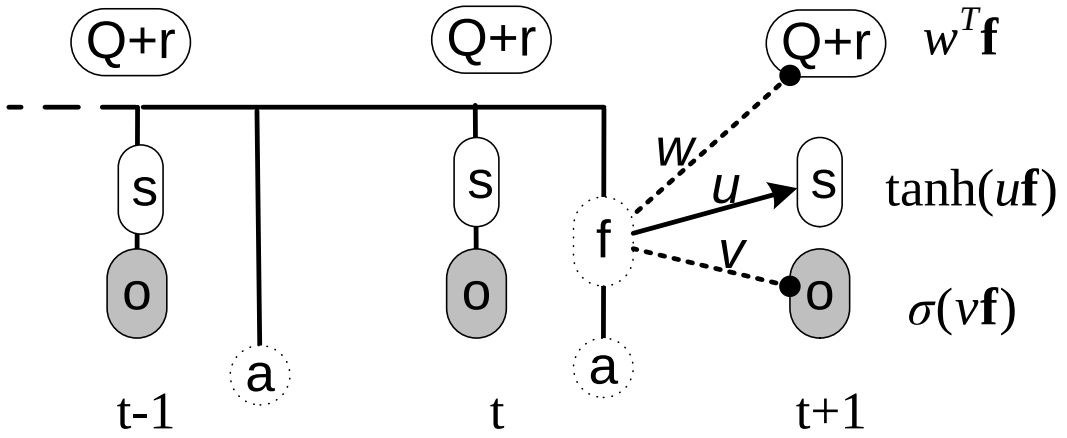
definition $Q^{\pi}(h^t, a^t) = E[r^{t+1} + \max_a Q^{\pi}(h^{t+1}, a)]$

implementation $Q^{\pi}(h, a) = w^T \mathbf{f}(h, a)$

Decision $a^t = \arg\max_a Q^{\pi}(h^t, a)$

Each feature has form [s|o]a+

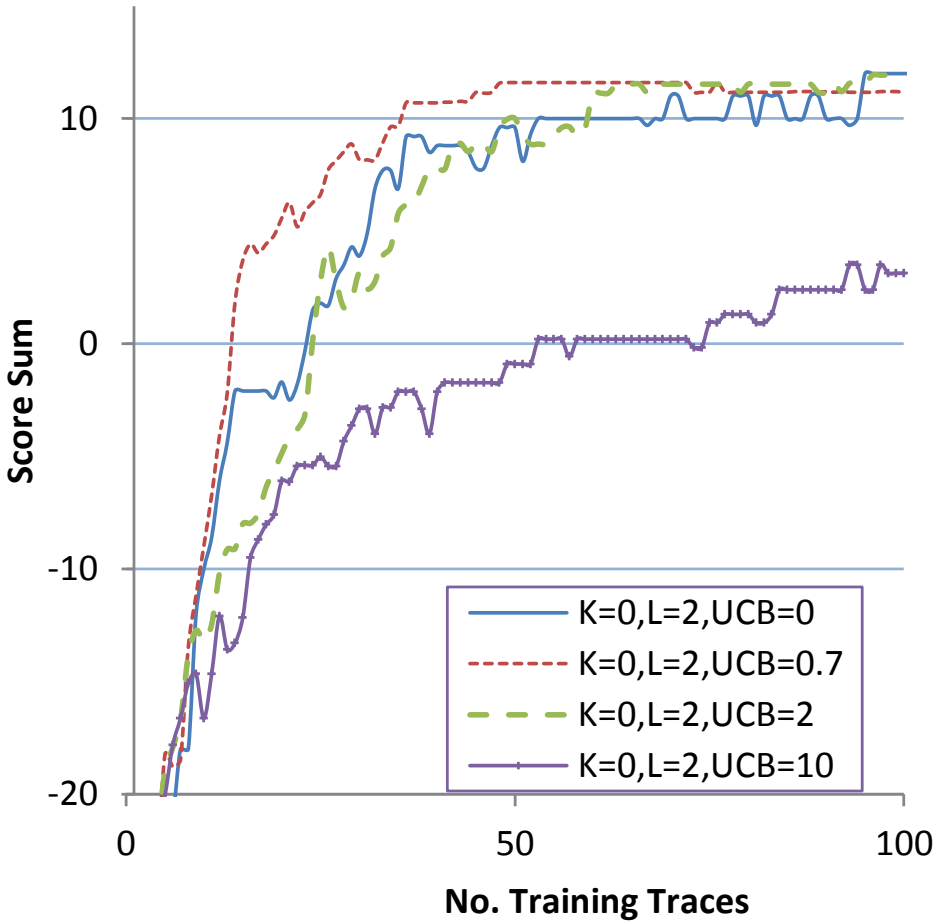
s is a state, o is an observation, a is an action



Upper Confidence Bound (UCB)

$$a^t = \arg\max_a \left[(w + std(w))^T \mathbf{f}(h,a) \right]$$

Right amount of exploration is important

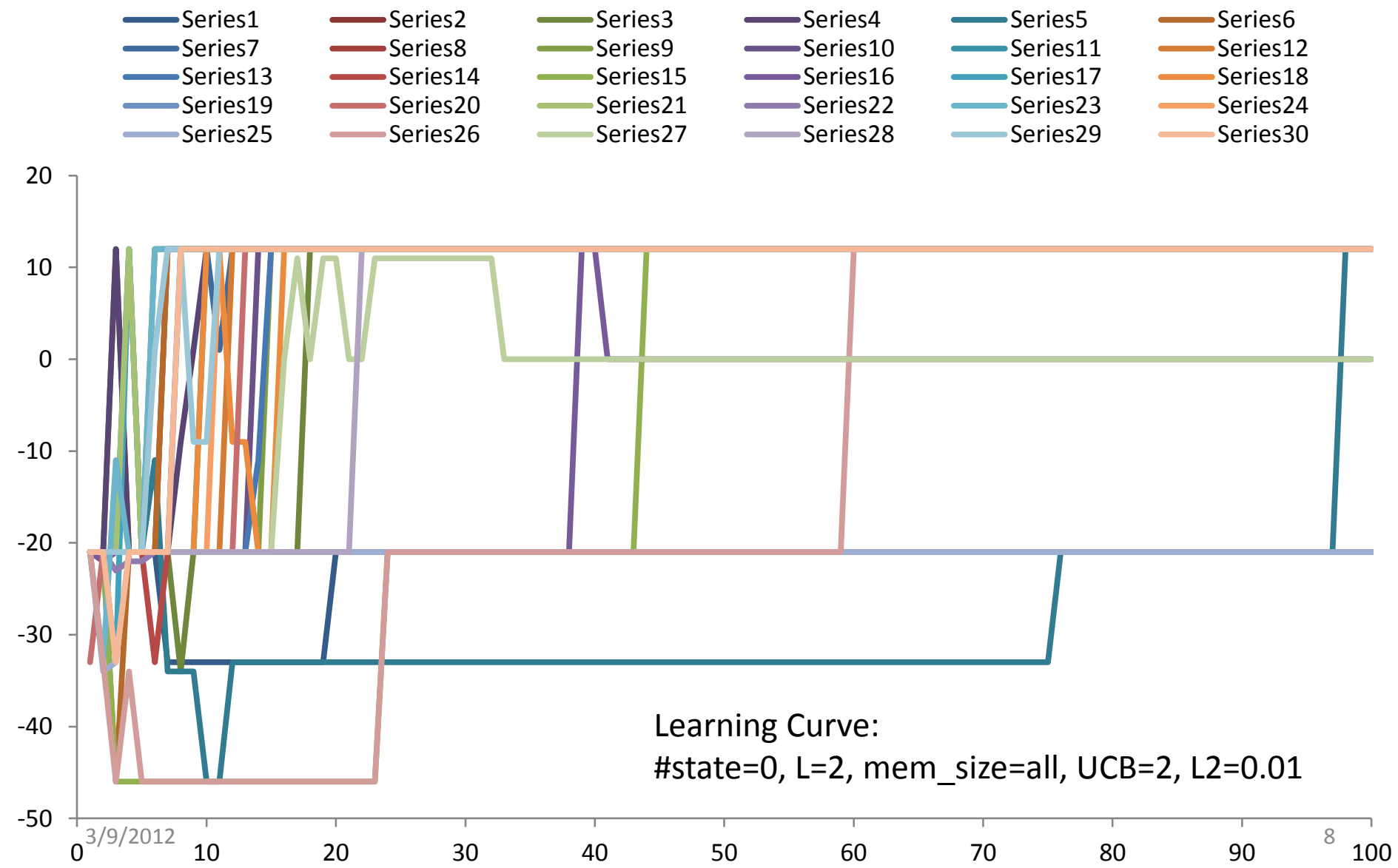


Training

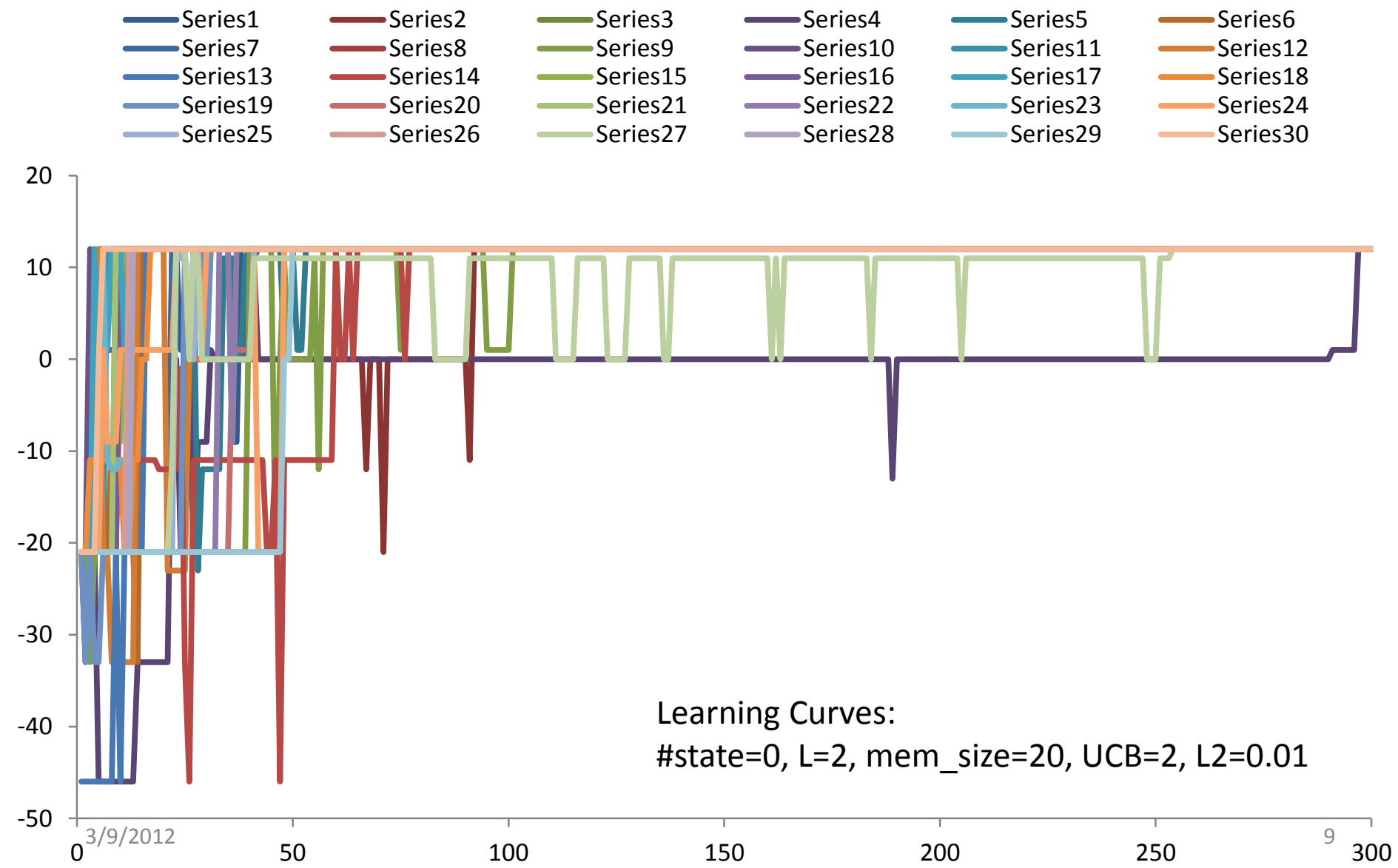
Value iteration:

```
history={}
for (i=0..300){
  Trace t=generateTrace();
  history.addTrace(t);              // keep only the latest
                                       K=30 traces
  addAllFeatures(t);                // observed features
                                       up to length L
  minimize_loss_θ(history, θ);      // l-bfgs
  updateConfidence();               // UCB
}
```

Remember Everything

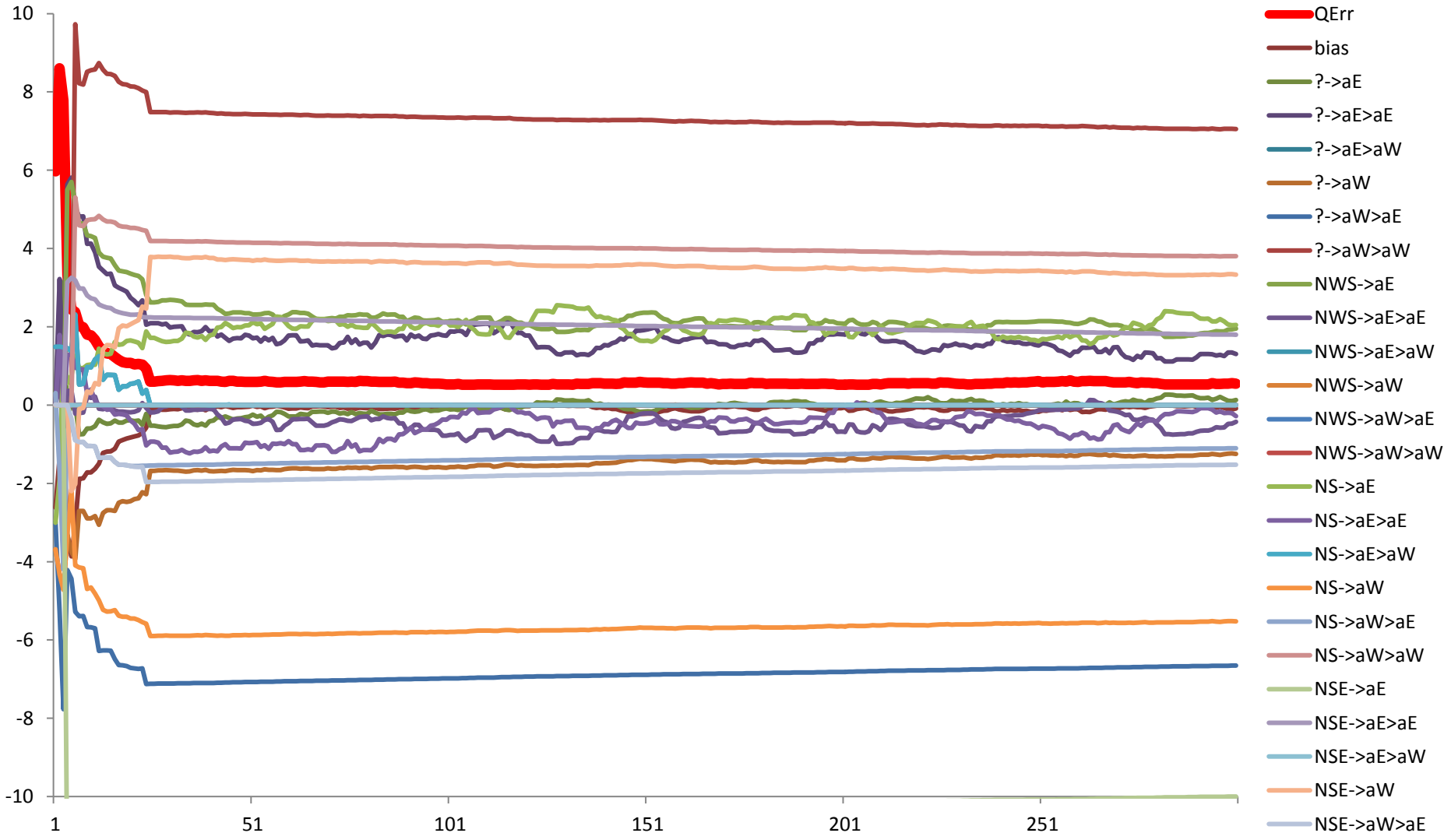


Remember Recent Events Only



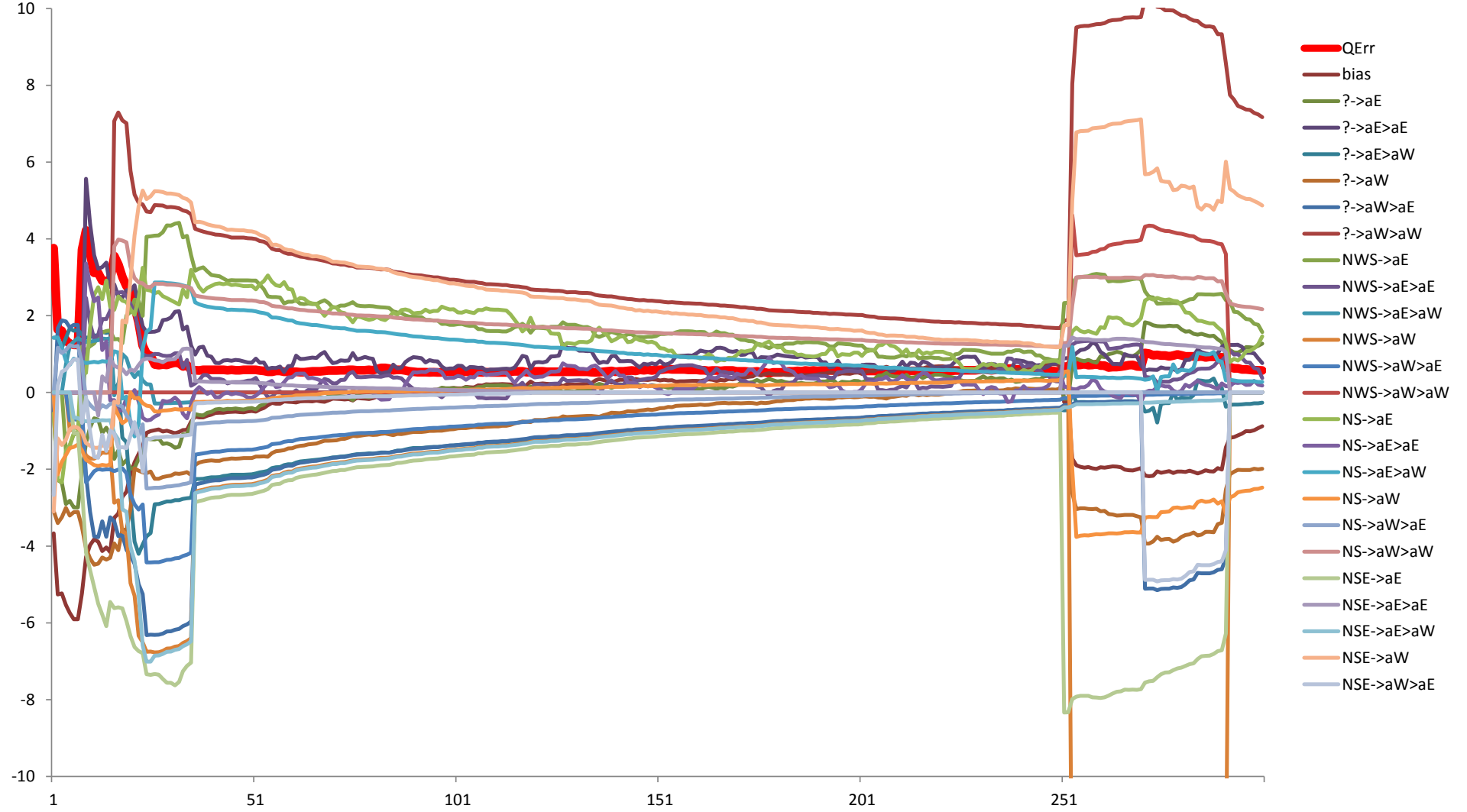
The Error Quickly Drops

Model weights over time: #state=0, L=2, mem_size=20, UCB=2, L2=0.01



Things Get Forgotten Overtime

Model Weights over time: #state=0, L=2, mem_size=20, UCB=2, L2=0.03



Lack of Long Term Memory

Model weights over time: #state=0, L=2, mem_size=20, UCB=2, L2=0.1