

Hierarchical Clustering and Reinforcement learning in Soar

Yongjia Wang
John E. Laird

Outline

- Motivations
- Introduction to Soar-RL
- Hierarchical Clustering
- Using Clustering with Soar-RL
- Simulation Results

Motivations

- Explore and extend learning capabilities in Soar
 - Statistical learning
 - Clustering (category learning)
 - Reinforcement learning
- Study the interaction among architectural learning mechanisms
 - RL + Clustering

Outline

- Motivations
- Introduction to Soar-RL
- Hierarchical Clustering
- Using Clustering with Soar-RL
- Simulation Results

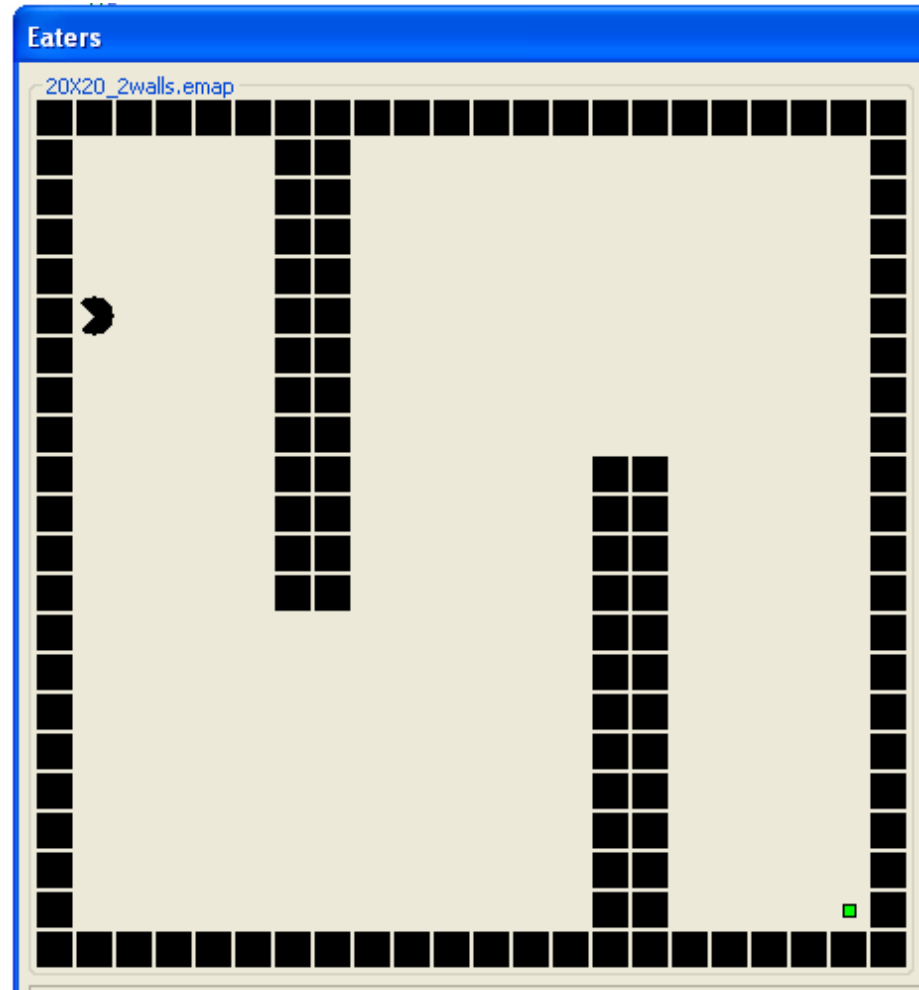
Reinforcement Learning Task

20X20 grid world

The agent starts from the left-most column

The goal is to reach the lower-right corner

The agent knows the (x,y) coordinates of current location

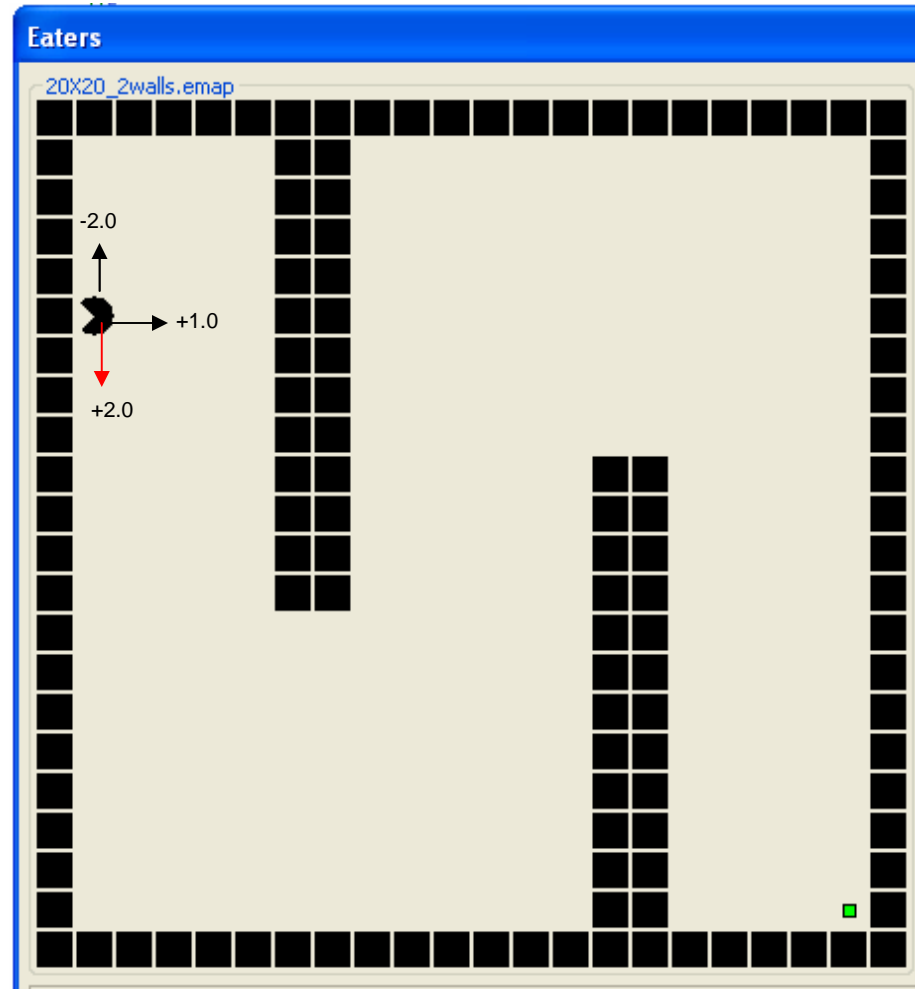


Soar-RL Mechanism

```
sp {Soar-RL-1
  (state <s> ^agent-position <ap>
    ^operator <o> +)
  (<ap> ^x 1
    ^y 5)
  (<o> ^name move
    ^direction south)
-->
(<s> ^operator <o> = 2.0)
}

sp {Soar-RL-2
  (state <s> ^agent-position <ap>
    ^operator <o> +)
  (<ap> ^x 1
    ^y 5)
  (<o> ^name move
    ^direction north)
-->
(<s> ^operator <o> = -2.0)
}

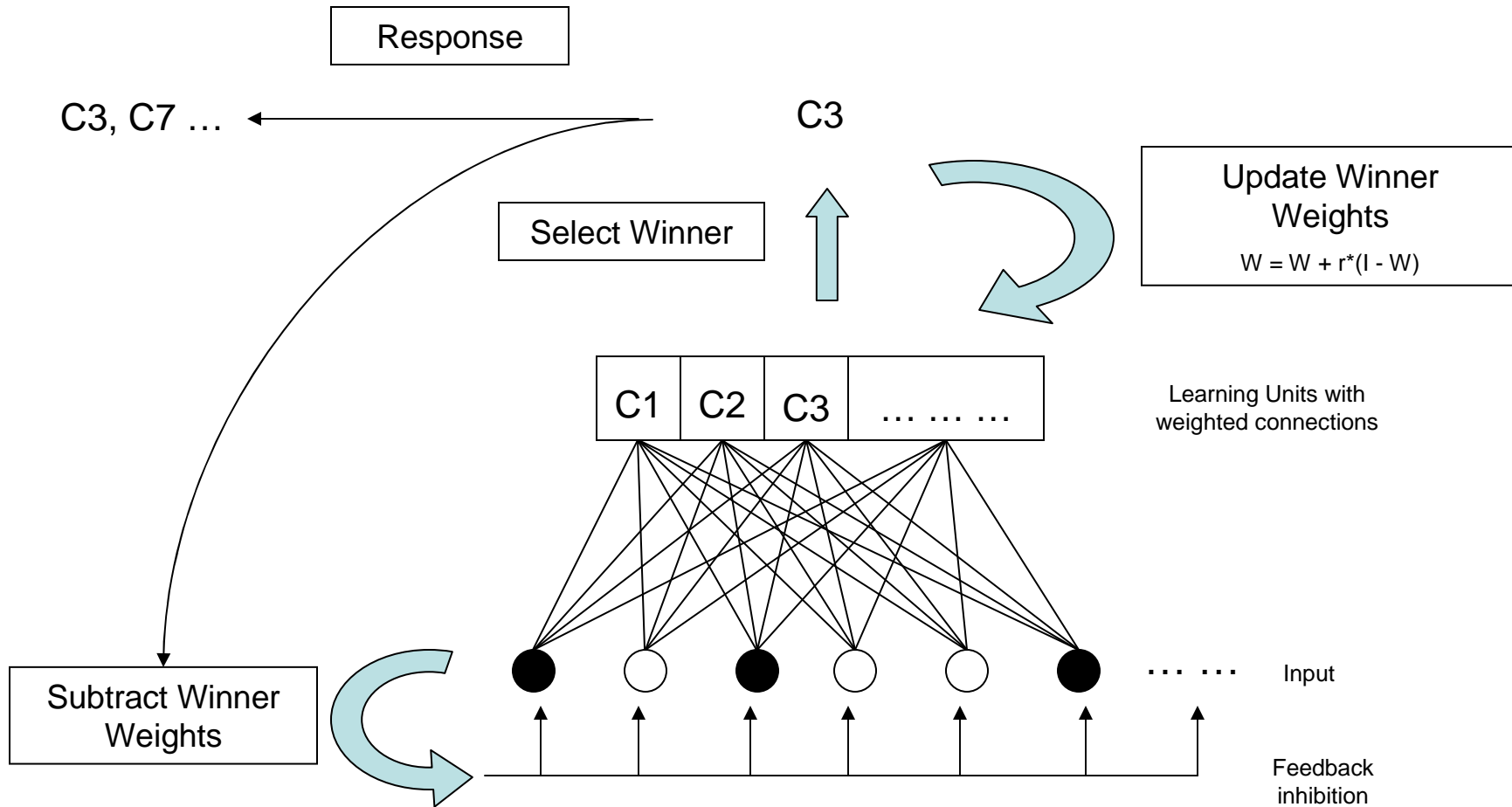
sp {Soar-RL-3
  (state <s> ^agent-position <ap>
    ^operator <o> +)
  (<ap> ^x 1
    ^y 5)
  (<o> ^name move
    ^direction east)
-->
(<s> ^operator <o> = 1.0)
}
```



Outline

- Motivations
- Introduction to Soar-RL
- Hierarchical Clustering
- Using Clustering with Soar-RL
- Simulation Results

Hierarchical Clustering



Outline

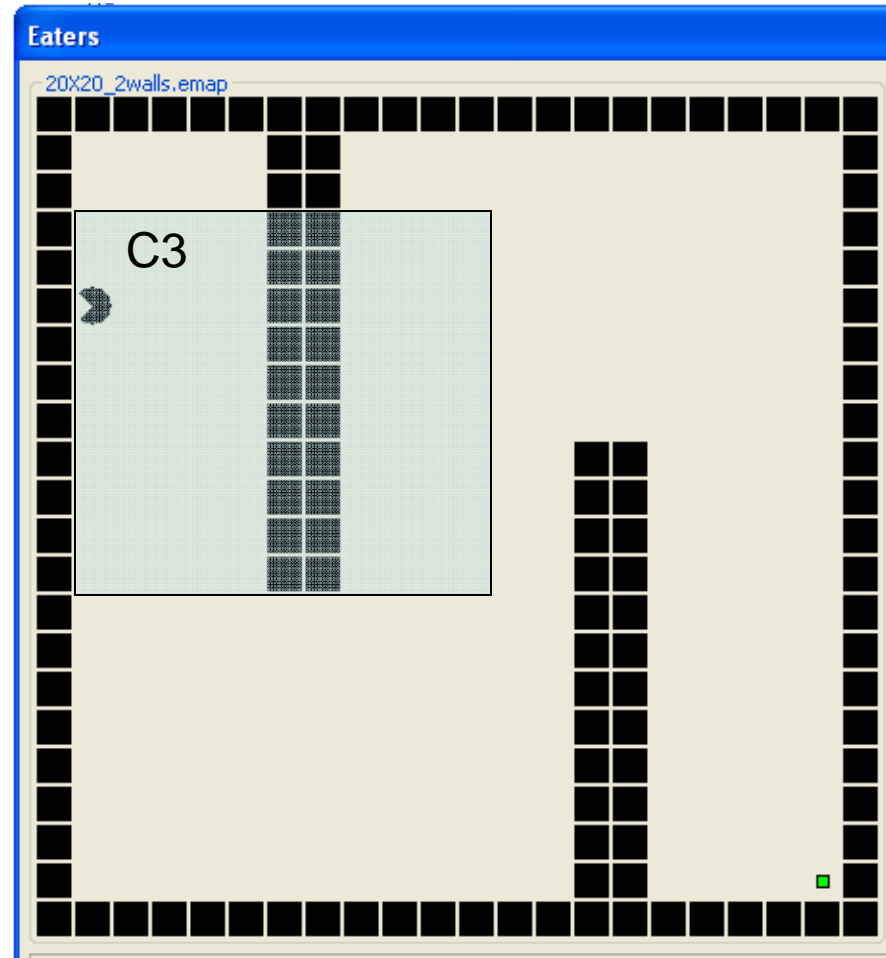
- Motivations
- Introduction to Soar-RL
- Hierarchical Clustering
- Using Clustering with Soar-RL
- Simulation Results

Using Clustering with RL

- Hierarchical clustering + RL
 - RL learns the utility of state-action pairs
 - Clustering mechanism groups similar states into clusters
 - RL can use cluster labels as state representation
 - Result: Improve learning performance

Soar-Cluster-RL Example

```
sp {Soar-Cluster-level-1-RL-1
  (state <s> ^agent-perception <ap>)
  (<ap> ^cluster-level-1 C3)
  (<o> ^name move
    ^direction south)
-->
  (<s> ^operator <o> = 1.0)
}
```

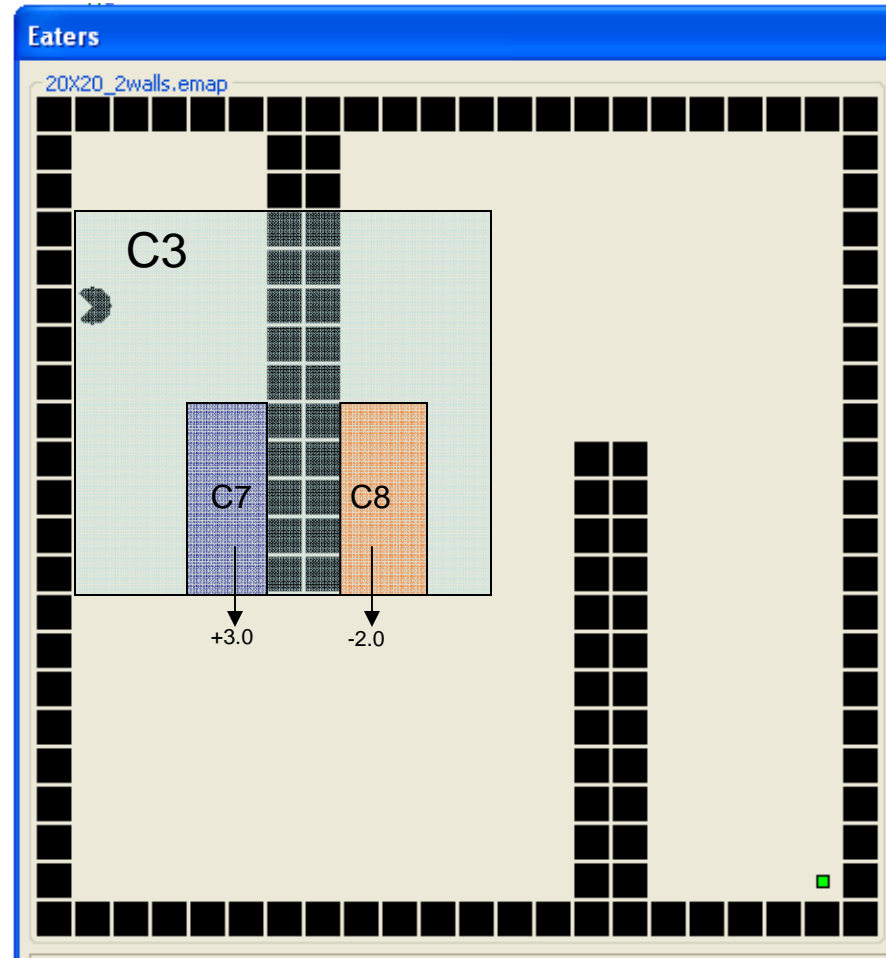


Soar-Cluster-RL Example

```
sp {Soar-Cluster-level-1-RL-1
  (state <s> ^agent-perception <ap>)
  (<ap> ^cluster-level-1 C3)
  (<o> ^name move
    ^direction south)
-->
(<s> ^operator <o> = 1.0)
}

sp {Soar-Cluster-level-2-RL-1
  (state <s> ^agent-perception <ap>)
  (<ap> ^cluster-level-1 C3
    ^cluster-level-2 C7)
  (<o> ^name move
    ^direction south)
-->
(<s> ^operator <o> = 2.0)
}

sp {Soar-Cluster-level-2-RL-2
  (state <s> ^agent-perception <ap>)
  (<ap> ^cluster-level-1 C3
    ^cluster-level-2 C8)
  (<o> ^name move
    ^direction south)
-->
(<s> ^operator <o> = -3.0)
}
```

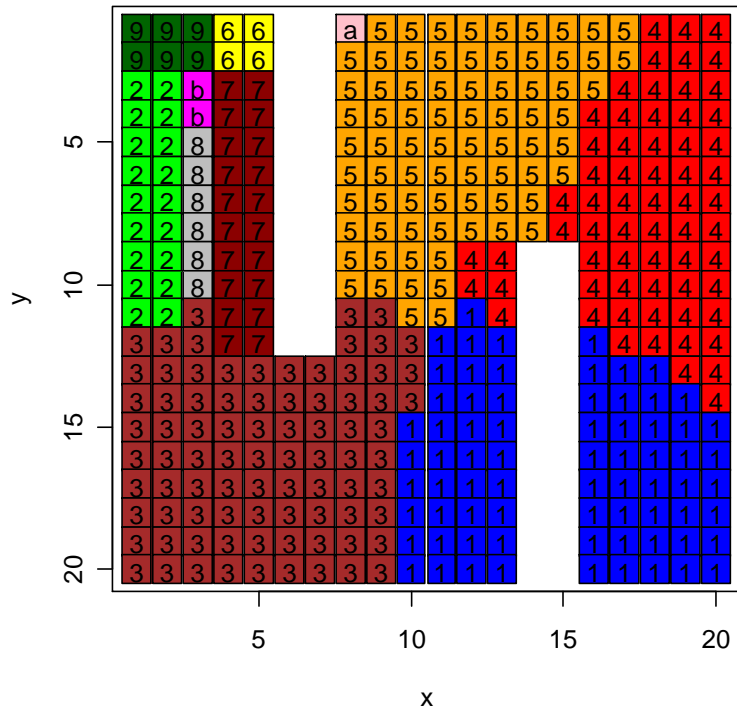


Outline

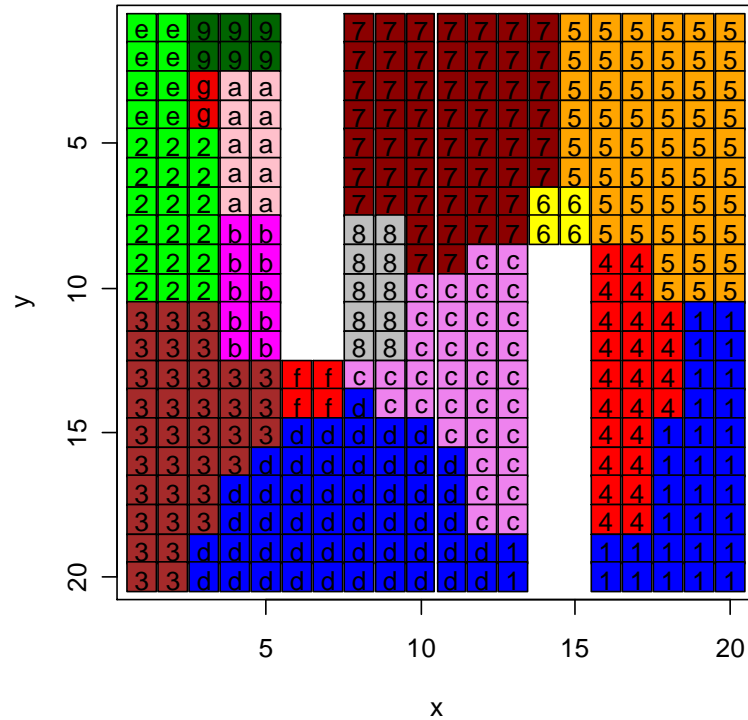
- Motivations
- Introduction to Soar-RL
- Hierarchical Clustering
- Using Clustering with Soar-RL
- **Simulation Results**

Result of Clustering

Input vector: $(x_{\{1\sim 20\}}, y_{\{1\sim 20\}}, \text{east-wall}\{0,1\}, \text{west-wall}\{0,1\}, \text{south-wall}\{0,1\}, \text{north-wall}\{0,1\})$



Level 1 Cluster



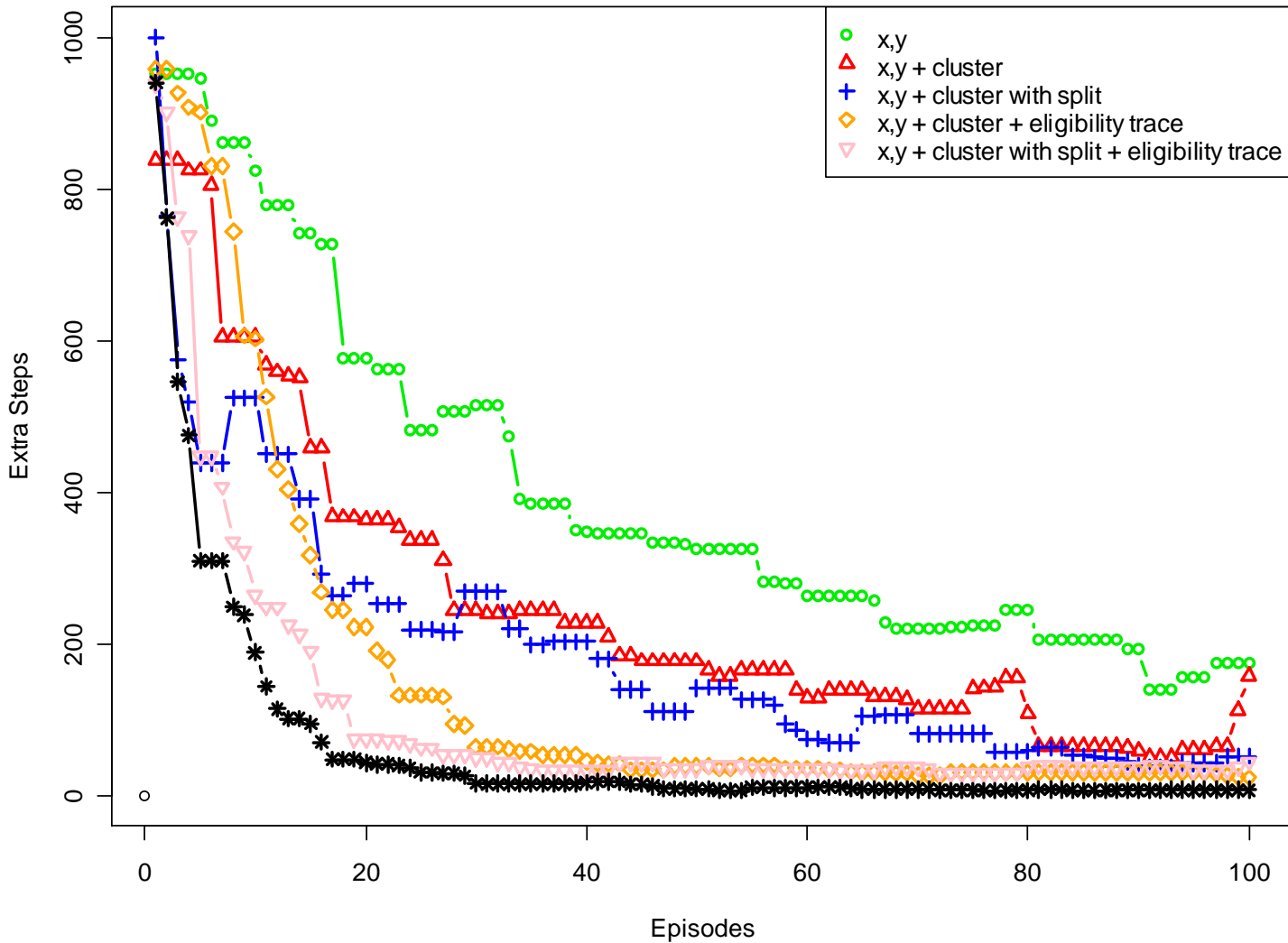
Level 1 Cluster with heuristic splitting₁₄

Eligibility Trace

- Soar-RL can be configured to use eligibility trace
- RL with eligibility trace can update multiple steps - speeds up learning when reward horizon is long

Results of Learning

cluster: 3 levels of cluster-RL rules



Conclusions

- Soar-RL automatically learns from general to specific situations with hierarchically clustered state representation
- Better clusters can be generated with proper heuristics, and results in better learning performance

Nuggets and Coal

- Nuggets
 - Studied the interaction between hierarchical clustering and RL
 - Demonstrated that Soar-RL with parallel rule learning can automatically discover regularities from general to specific
- Coal
 - Simple task: simple perception and action; discrete environment