



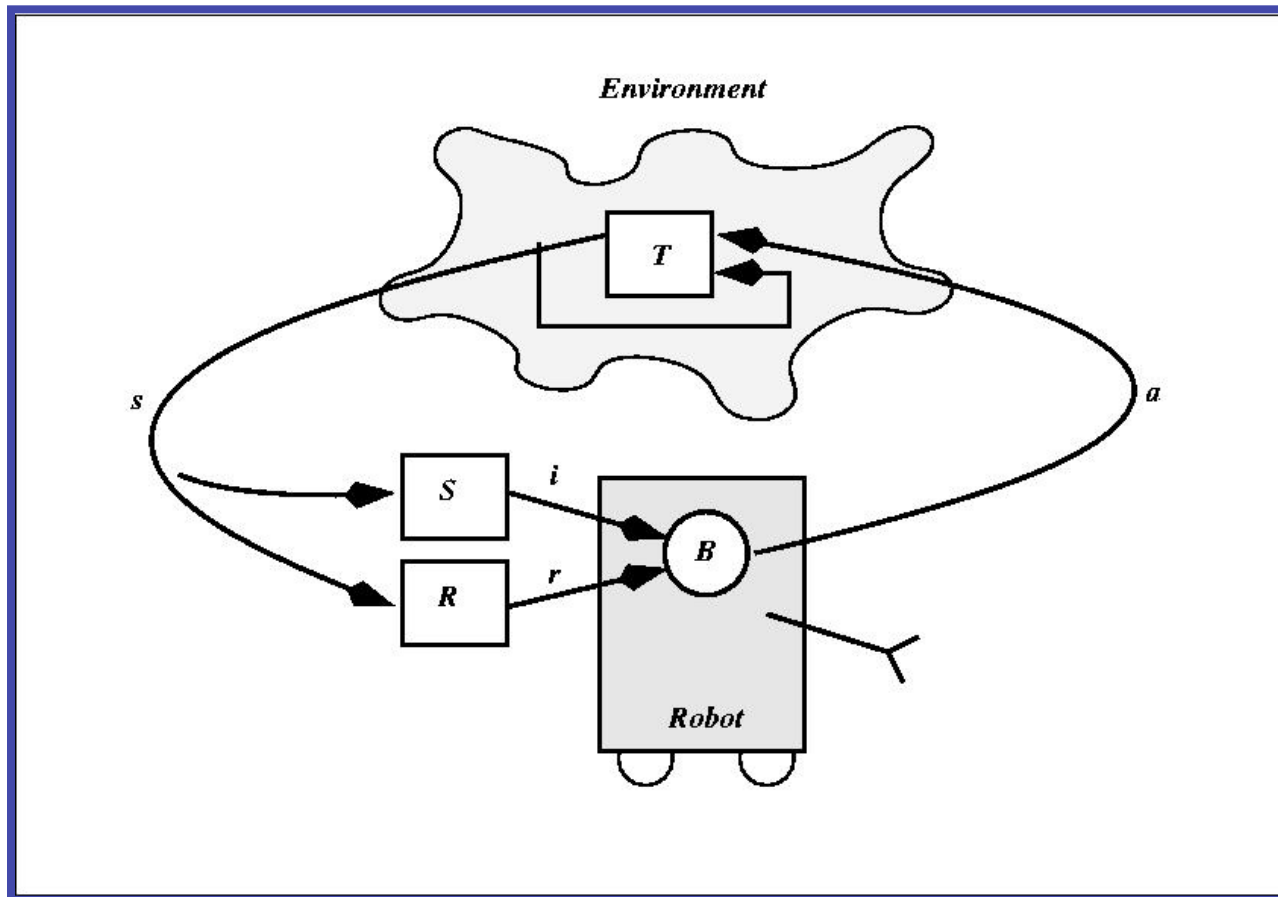
Behavior-Based Programming of Robots and Multi-Robot Teams

Part 5: Learning

IJCAI tutorial SA-4
Presented by Tucker Balch

Reinforcement Learning

The Standard RL Model

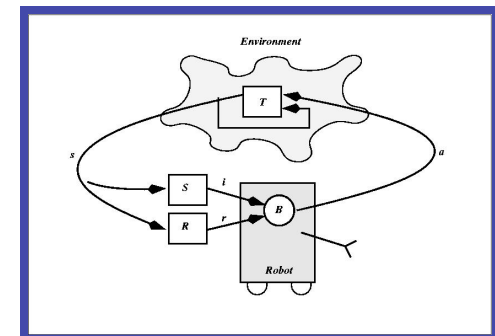


SA4-2

Reinforcement Learning

Markov Decision Problems

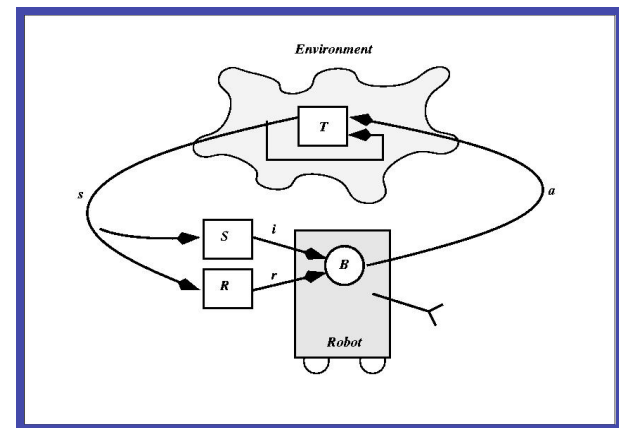
- Set of states S
- Set of actions A
- Transition function $T(s,a,s') =$
 - Probability, if in state s , take action a , end up in s'
- Reinforcement function $R(s,a) =$
 - Reward, if in state s and take action a



Reinforcement Learning

Markov Decision Problems

- “Experience tuple” provided at each step = $\langle s, a, s', r \rangle$
- Reward r is for LAST action
- Objective is to find a POLICY that maps state to action to maximize sum of r over time



Reinforcement Learning

Example

Environment: "you are in state 3, you have two choices"

Agent: "I pick action number 1"

Environment: "you receive a reward of -1, you are still in state number 3, you have two choices"

Agent: "I pick action number 2"

Environment: "you receive a reward of +1, you are now in state 5, you have three choices"

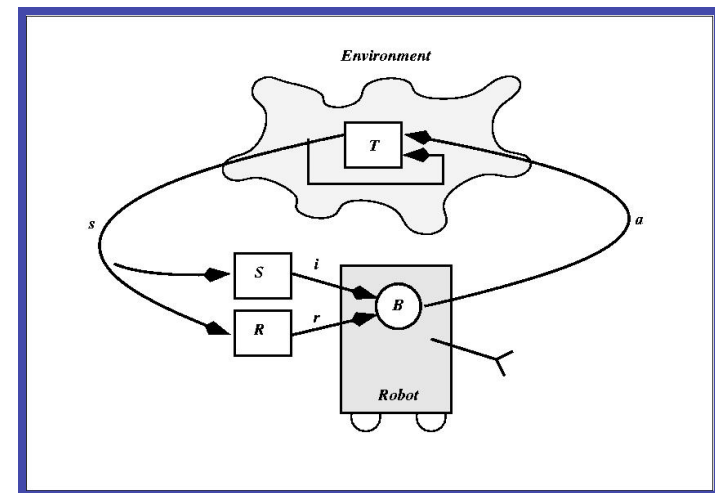
Agent: "I pick action number 1"

...

Reinforcement Learning

How Map to Behavior Based Control?

- Environment T is the world
- State s = current perceptual features
- Action a = selected behavior
- Reward r = function selected by programmer



Reinforcement Learning

What is Optimal?

- Finite horizon
- Infinite horizon
- Average reward

Reinforcement Learning

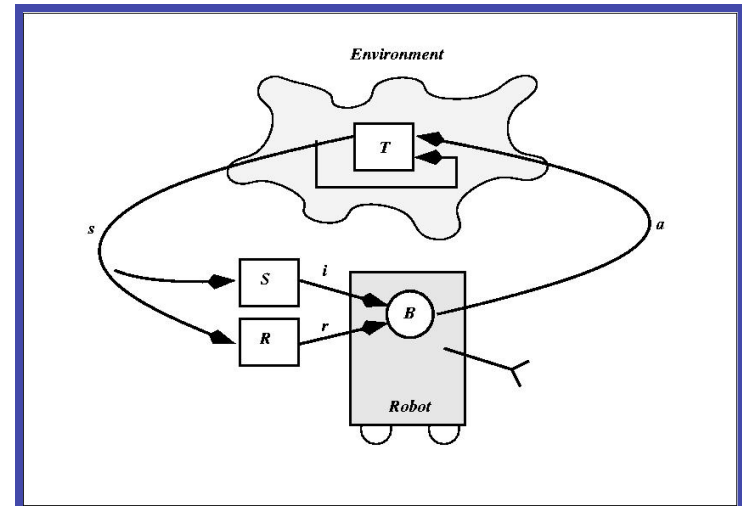
How to Evaluate ?

- Does it converge to optimal?
- How quickly does it converge to optimal?
- Regret

Reinforcement Learning

Algorithms

- Assume you have T and R
 - Value iteration
 - Policy iteration
- Assume you don't have T and R
 - Model-based learning
 - Model-free learning



Reinforcement Learning

Algorithms: Q-Learning

- $Q[s,a]$ is value of taking action a in state s
- Update Q as follows:

Given experience tuple $\langle s,a,s',r \rangle$:

$$Q[s,a] = (1-\alpha) * Q[s,a] + \alpha * (r + \gamma * \operatorname{argmax}_{a'} Q[s',a'])$$

(Watkins)

Reinforcement Learning

Algorithms: Dyna

- Use $\langle s, a, s', r \rangle$ to learn statistical models of T and R
- On each execution cycle:
 - Perform regular Q update
 - Choose k additional $\langle s, a \rangle$ pairs at random
 - Use models of T and R to find s' and r
 - Perform additional Q update using artificial experience $\langle s, a, s', r \rangle$

Reinforcement Learning

Sources

- *Reinforcement Learning, a Survey*, by Kaelbling, Littman, Moore
- *Reinforcement Learning* (book) by Sutton and Barto
- *Machine Learning*, by Mitchell

Reinforcement Learning

How to Integrate RL and BB

- Select perceptual features == State
- Select behavioral assemblages == Actions
- Define utility of current situation == Reward

- Examples