Online Policy Training vs Heuristic Search Using Reinforcement Learning to Avoid Dynamic Obstacles

cs980 project

Tianyi Gu, Mostafa Hussein, Yishan Luo, Yuncong Zhou

University of New Hampshire

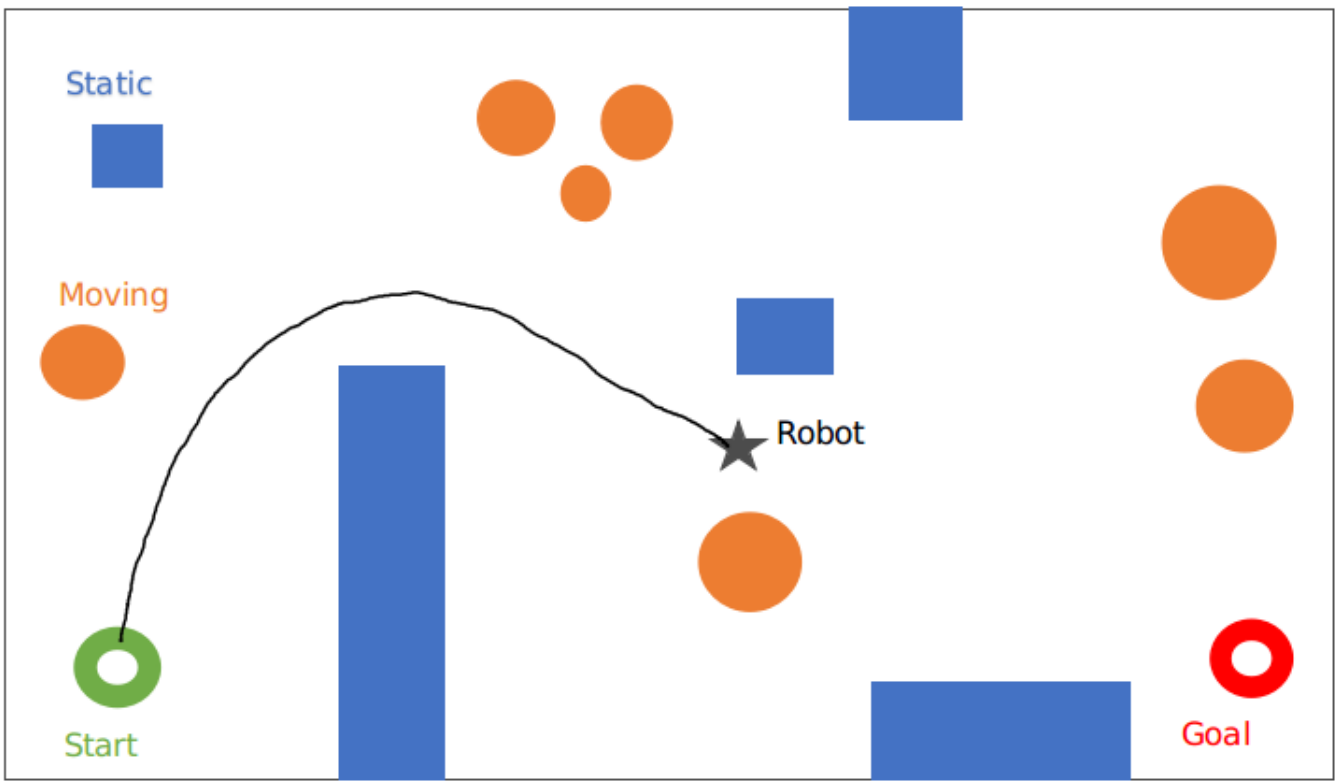
October 21, 2017

MOTIVATION

- Avoid dynamic obstacles is crucial
  - Autonomous vehicle navigation
  - Indoor robot navigation

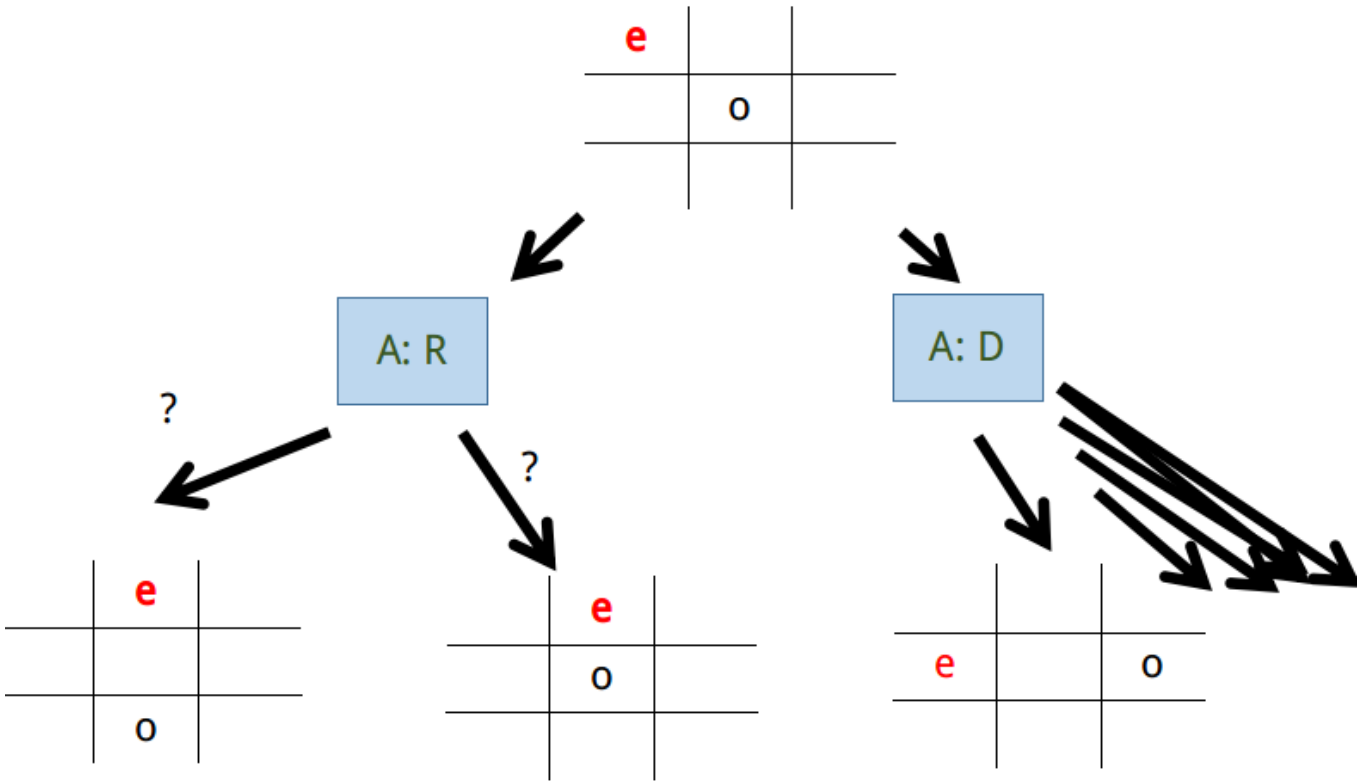
ONLINE POLICY TRAINING

- Given start state, goal state, and static obstacles
- Compute online policy based on observation of dynamic obstacles



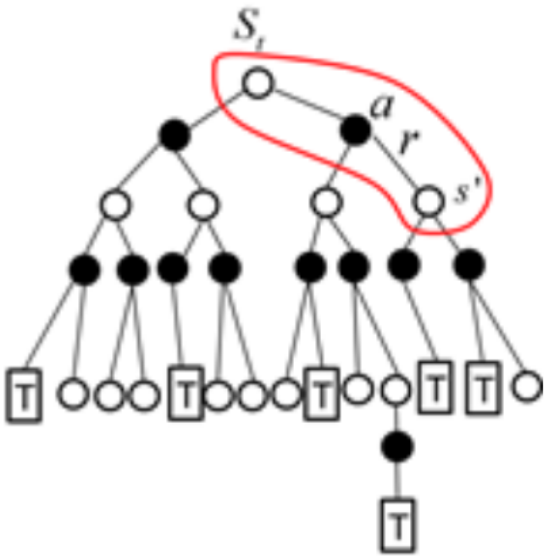
ONLINE POLICY TRAINING: MDP

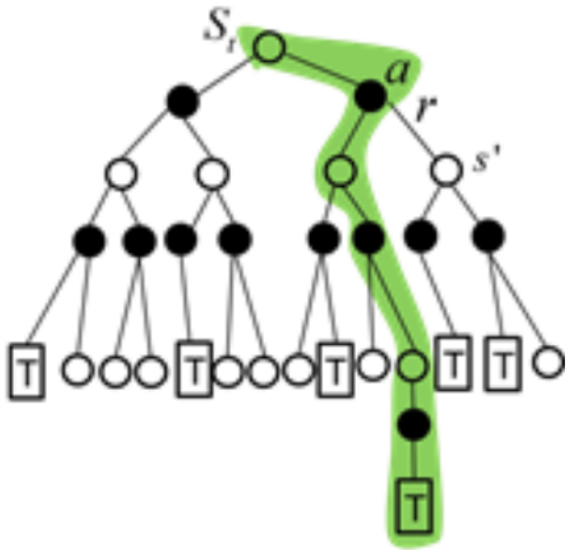
- MDP:
  - States: grid world states
  - Actions: up, down, left, right
  - Rewards: goal: 1, collision: -1000
  - Transition: Unknown



ONLINE POLICY TRAINING: MDP

- MDP:
  - States: grid world states
  - Actions: up, down, left, right
  - Rewards: goal: 1, collision: -1000
  - Transition: Unknown
- Approach:
  - TD(λ), Q learning, Sarsa
  - Monte Carlo Tree Search



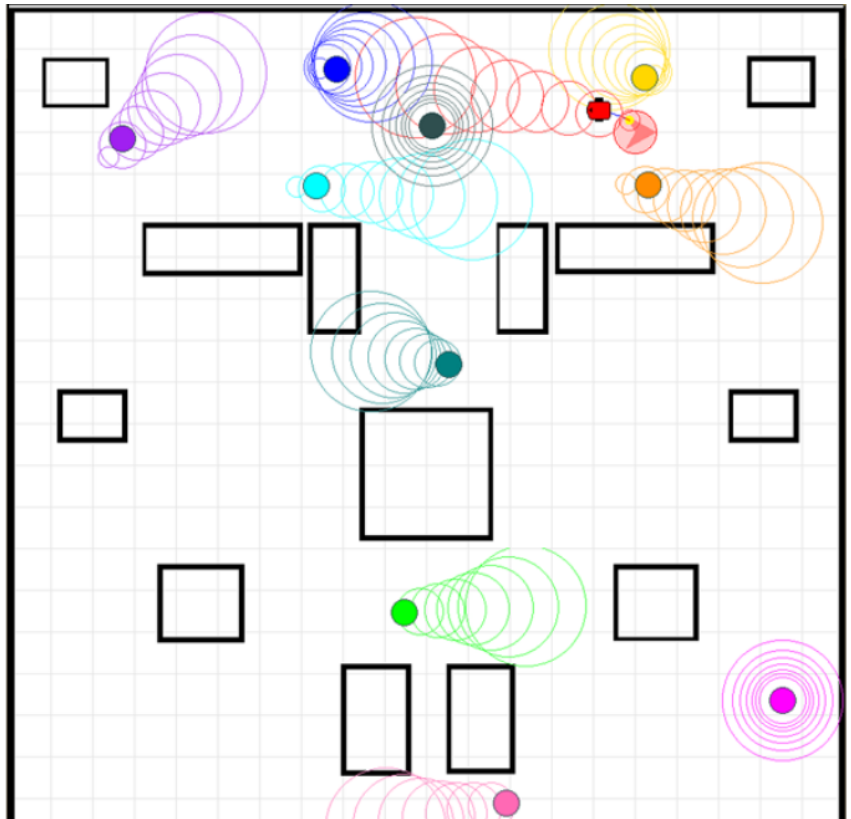


ONLINE PATH PLANNING

- Real Time Heuristic Search
  - estimate obstacles' movement
  - deterministic problem with different cost function

ONLINE PATH PLANNING

- Real Time Heuristic Search
  - estimate obstacles' movement
  - deterministic problem with different cost function
- Approach:
  - Greedy, A*
  - LSS-LRTA*



EXPERIMENT

- Compare Policy Training vs Heuristic Search
- Theoretically prove the value functions are the same