CS394R
Reinforcement Learning: Theory and Practice

Peter Stone

Department of Computer Science
The University of Texas at Austin

Good Afternoon Colleagues

- Are there any questions?

Helicopter Control

- State: position, orientation, velocity, angular vels
- Actions: Settings of the 4 or 5 controls
- Goal: Hover

Helicopter Control

- State: position, orientation, velocity, angular vels
- Actions: Settings of the 4 or 5 controls
- Goal: Hover
- How would you formulate the problem "by the book"?

Helicopter Control

- State: position, orientation, velocity, angular vels
- Actions: Settings of the 4 or 5 controls
- Goal: Hover
- How would you formulate the problem "by the book"?
- Could you implement that? Why or why not?

Helicopter Control

- State: position, orientation, velocity, angular vels
- Actions: Settings of the 4 or 5 controls
- Goal: Hover
- How would you formulate the problem "by the book"?
- Could you implement that? Why or why not?
- At a high level, what do they do instead?

Helicopter Control

- State: position, orientation, velocity, angular vels
- Actions: Settings of the 4 or 5 controls
- Goal: Hover
- How would you formulate the problem "by the book"?
- Could you implement that? Why or why not?
- At a high level, what do they do instead?
  - Collect a small amount of human expert data
  - Use that to train a **1-step** model (simulator)
  - Determine the optimal policy in the simulator
  - Fly it!

Helicopter Control

- State: position, orientation, velocity, angular vels
- Actions: Settings of the 4 or 5 controls
- Goal: Hover
- How would you formulate the problem "by the book"?
- Could you implement that? Why or why not?
- At a high level, what do they do instead?
  - Collect a small amount of human expert data
  - Use that to train a **1-step** model (simulator)
  - Determine the optimal policy in the simulator
  - Fly it!
- Would this approach work on the Aibo walking task?

Ng paper

- Why quadratic reward (p. 5)?

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy
- How does he do policy optimization?

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy
- How does he do policy optimization?
  - Represent policy as a Neural net (Fig 2c)

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy
- How does he do policy optimization?
  - Represent policy as a Neural net (Fig 2c)
  - greedy hillclimbing over few parameters (the weights)!

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy
- How does he do policy optimization?
  - Represent policy as a Neural net (Fig 2c)
  - greedy hillclimbing over few parameters (the weights)!
- Shaping rewards

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy
- How does he do policy optimization?
  - Represent policy as a Neural net (Fig 2c)
  - greedy hillclimbing over few parameters (the weights)!
- Shaping rewards
- Topics for other courses:
  - Kalman filter (robotics)
  - Cross-validation/hold-out testing (supervised learning)

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy
- How does he do policy optimization?
  - Represent policy as a Neural net (Fig 2c)
  - greedy hillclimbing over few parameters (the weights)!
- Shaping rewards
- Topics for other courses:
  - Kalman filter (robotics)
  - Cross-validation/hold-out testing (supervised learning)
- Can it generalize to adverse conditions?

Ng paper

- Why quadratic reward (p. 5)?
- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples to evaluate each policy
- How does he do policy optimization?
  - Represent policy as a Neural net (Fig 2c)
  - greedy hillclimbing over few parameters (the weights)!
- Shaping rewards
- Topics for other courses:
  - Kalman filter (robotics)
  - Cross-validation/hold-out testing (supervised learning)
- Can it generalize to adverse conditions?
- Easy problem or a powerful approach?