CS394R
Reinforcement Learning: Theory and Practice

Peter Stone
Department of Computer Science
The University of Texas at Austin
Good Morning Colleagues

• Are there any questions?
Logistics

- Make progress on final projects!
Logistics

- Make progress on final projects!
- This week’s readings: robot applications
Logistics

- Make progress on final projects!
- This week’s readings: robot applications
- Next week’s readings: Practical RL
Logistics

- Make progress on final projects!
- This week’s readings: robot applications
- Next week’s readings: Practical RL
  - A general study, motivated by applications like the ones we’ve read
Logistics

- Make progress on final projects!
- This week’s readings: robot applications
- Next week’s readings: Practical RL
  - A general study, motivated by applications like the ones we’ve read
  - A survey of transfer learning
Quadruped locomotion

- Why update A in that way?

- Used on other tasks? (chin pinch, humanoid walk)
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 vel
- 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!
• PEGASUS - how does it help policy evaluation?
PEGASUS - how does it help policy evaluation?

- General question: is policy good or lucky?
Ng paper

- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples for evaluation of each policy
Ng paper

- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples for evaluation of each policy
- How does he do policy optimization?
Ng paper

- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples for evaluation of each policy
- How does he do policy optimization?
  - greedy hillclimbing over few parameters (the NNs)!
- Can it generalize to adverse conditions?
Ng paper

- PEGASUS - how does it help policy evaluation?
  - General question: is policy good or lucky?
  - Use same random samples for evaluation of each policy
- How does he do policy optimization?
  - greedy hillclimbing over few parameters (the NNs)!
- Can it generalize to adverse conditions?
- Where’s the power? Is it an easy problem or a powerful approach?
Class Discussion

• Vic