

CS394R
Reinforcement Learning:
Theory and Practice
Fall 2007

Peter Stone

Department of Computer Sciences
The University of Texas at Austin

Good Afternoon Colleagues

- Are there any questions?

This Chapter

- Solution methods **given a model**

This Chapter

- Solution methods **given a model**
 - So no exploration vs. exploitation
- Why is it called dynamic programming?

Policy Evaluation

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π . (p. 90)

Policy Evaluation

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π . (p. 90)
- Policy evaluation converges under the same conditions (p. 91)

Policy Evaluation

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π . (p. 90)
- Policy evaluation converges under the same conditions (p. 91)
- Policy evaluation on the week 0 problem
 - undiscounted, episodic

Policy Evaluation

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π . (p. 90)
- Policy evaluation converges under the same conditions (p. 91)
- Policy evaluation on the week 0 problem
 - undiscounted, episodic
 - Are the conditions met?

Policy Evaluation

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π . (p. 90)
- Policy evaluation converges under the same conditions (p. 91)
- Policy evaluation on the week 0 problem
 - undiscounted, episodic
 - Are the conditions met?
 - (book slides)

Policy Evaluation

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π . (p. 90)
- Policy evaluation converges under the same conditions (p. 91)
- Policy evaluation on the week 0 problem
 - undiscounted, episodic
 - Are the conditions met?
 - (book slides)
- Exercises 4.1, 4.2

Policy Improvement

- Policy improvement theorem:

$$\forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^{\pi'}(s) \geq V^\pi(s)$$

Policy Improvement

- Policy improvement theorem:

$$\forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^{\pi'}(s) \geq V^\pi(s)$$

- (book slides)

Policy Improvement

- Policy improvement theorem:

$$\forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^{\pi'}(s) \geq V^\pi(s)$$

- (book slides)
- Polynomial time convergence (in number of states and actions) even though m^n policies.
 - Ignoring effect of γ and bits to represent rewards/transitions

Policy Improvement

- Policy improvement theorem:

$$\forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^{\pi'}(s) \geq V^\pi(s)$$

- (book slides)
- Polynomial time convergence (in number of states and actions) even though m^n policies.
 - Ignoring effect of γ and bits to represent rewards/transitions
 - p. 107: Is LP still inefficient?

Student Discussion

- Jeremy on the Gambler's Problem (p. 101)

Student Discussion

- Jeremy on the Gambler's Problem (p. 101)
- Email discussion linked to the book web page

Value Iteration on Week 0 problem

- Show the new policy at each step
 - Not actually to compute policy

Value Iteration on Week 0 problem

- Show the new policy at each step
 - Not actually to compute policy
 - Break policy ties with equiprobable actions

Value Iteration on Week 0 problem

- Show the new policy at each step
 - Not actually to compute policy
 - Break policy ties with equiprobable actions
 - No stochastic transitions

Value Iteration on Week 0 problem

- Show the new policy at each step
 - Not actually to compute policy
 - Break policy ties with equiprobable actions
 - No stochastic transitions
- How would policy iteration proceed in comparison?
 - More or fewer policy updates?

Value Iteration on Week 0 problem

- Show the new policy at each step
 - Not actually to compute policy
 - Break policy ties with equiprobable actions
 - No stochastic transitions
- How would policy iteration proceed in comparison?
 - More or fewer policy updates?
 - True in general?

Summary

- p. 109: This chapter treats **bootstrapping** with a model

Summary

- p. 109: This chapter treats **bootstrapping** with a model
 - Next: no model and no bootstrapping

Summary

- p. 109: This chapter treats **bootstrapping** with a model
 - Next: no model and no bootstrapping
 - Then: no model, but bootstrapping