Graduate Artificial Intelligence
CS 640
Markov Decision Processes

Jeffrey Considine

jconsidi@bu.edu

Announcements

Shared Compute Cluster (SCC) Tutorial next class (9/25)

- Bring your laptop!
- Will walk through account setup and ways to access the SCC.

AI Research seminar (AIR)

- Tuesdays 2-3pm @ CDS 1101

Updated Weekly Lecture Schedule

- What is AI?
- Responsible AI and Rule-based Systems
- Searching and Planning
- Markov Decision Processes
- Computing Optimal Policies
- Hidden Markov Models
- Midterm
- Neural Networks
- Neural Networks and Polices
- Computer Vision
- Face Recognition and Pose Estimation
- Natural Language Processing
- Large Language Models
- Game Playing
- Logic and Planning

Plan for Today

- Markov property and Markov processes
- Markov reward processes
- Markov decision processes
- Policies (may spill over to next time)

Markov Property

- Only the current state matters to predict the future.

- “memoryless”

Formalizing the Markov Property

If we know , then do not matter.

Implications of the Markov Property

- History does not matter once you fully know the current state.

- However, more information may need to be included in the state to get a Markov process.
  - Physical simulations: velocity
  - Chess: past positions where 3-fold repetition rule might trigger.

Markov Processes

A system or process is a Markov process if all transitions in the system have the Markov property.

- For any pair of states , is fixed and independent of states before time .
- If the Markov process has an infinite number of states, express as a probability density instead…

Examples of Markov Processes

- Physics
- Weather
- Users engaging with a web site?

- Autocorrect typing analysis
- Credit card fraud detection

General State Representations

- First pass at identifying state –
  - Write down everything that you care to observe later.
  - Write down everything that might affect the future.

- Often looks like a tuple of many features.
  - If each of these features has a finite number of values, then the Markov process will have a finite number of states.
  - Rounding to make features discrete and finite may compromise prediction power.
    - Precision hidden by rounding may affect probabilities.

Variations on Markov Processes

- Finite vs infinite
- Discrete vs continuous time
- Fully observable vs hidden state (covered in 2 weeks)

These apply to all the Markov variations covered later, but we will focus on finite, discrete and fully observable cases.

Finite Markov Processes

If a Markov process has a finite number of states, then we can order them as and use the shorthand

Sometimes you will see this further abbreviated just using the state numbers.

Transition Matrices

The transitions of a finite Markov process with states can be represented in an matrix.

Tricks with Transition Matrices

- There is a lot of analysis that can be done with transition matrices…
  - is a step transition matrix.
  - Eigenvectors and eigenvalues tell you about steady state distributions.

Any Questions?

???

Markov Reward Processes

- A Markov Reward Process = Markov Process + rewards
  - Like transitions in a Markov process, the next reward only depends on the current state.
  - At time , receive reward based on .

- Let denote the average reward after being in state . Then

Examples of Markov Reward Processes

- Weather – rewards are enjoying sun vs getting soaked…
- User on web site – did they do something that made us money?

Evaluating Markov Reward Processes (take 1)

- What is the total expected reward if you know the next current state?

Call the value function of this process.

* denotes optimality.

What about Loops?

???

Evaluating Markov Reward Processes (take 2)

- What is the total expected reward if you know the next current state?

- is a discount factor where .
  - Only use when loops are impossible.

Monte Carlo Evaluation of Markov Reward Process

- How should we evaluate ?

- Easy way is Monte Carlo simulation.
  - Given the transition matrix and expected state rewards , simulate the process many times and calculate the average…
  - How many simulations are needed?

Evaluating Markov Reward Processes (take 3)

Rewrite

to

Solving Markov Reward Processes (part 1)

Representing and as vectors in the same order as states…

Rewrite

to

Solving Markov Reward Processes (part 2)

Then (linear) algebra as follows.

Any Questions?

???

Markov Decision Processes (MDPs)

Markov decision processes add actions to the process.

- Transition probabilities and rewards depend on current state and action.

Examples of Markov Decision Processes

- Weather + take an umbrella decision
- Driving and delays (negative rewards)

This is our super generic model for decision making.

- The catch is it is too generic.
- Hidden state makes this really hard!

Markov Decision Processes vs Search

- Nodes are states.
- Edges are transitions.
- Choices of edges are actions.
  - Usually a lot fewer actions than nodes/states.

- Probabilistic component of MDPs allows probabilistic next nodes.
  - Deterministic transitions give “easy” search problem.
  - Probabilistic transitions break search algorithms.

Evaluating Markov Decision Processes (take 1)

- How do we evaluate a Markov decision process?
  - What is the goal of our actions?

Maximize the value function of Markov reward processes?

Maximize ?

Need to stick actions in here.

Evaluating Markov Decision Processes (take 2)

Now with explicit actions,

But not explicit enough – need to pick actions at multiple times!

Note: stands for optimal value from the best possible actions.

Evaluating Markov Decision Processes (take 3)

Rewrite

to

Note: this is a Bellman equation expressing optimal value as a recursive function of optimal actions and values.

Evaluating Markov Decision Processes (take 4)

Rewrite

to

Still open: how do we do this maximization?

Any Questions?

???

What is a Policy?

A policy is a function mapping states to actions.

- A deterministic policy returns a single action.

- A probabilistic policy returns a probability distribution of actions.

Usually denoted as or with subscripts for context...

Probabilistic vs Deterministic Policies

Deterministic policy advantages

- Usually sufficient for optimal performance.
  - Exceptions with simultaneous adversarial choices.
- Much smaller to represent.
  - actions for states

Probabilistic policy advantages

- Can always represent deterministic policies.
  - All probability on one action.
- Often convenient for numerical optimizations.

Representing Policies for Finite MDPs

Represent a deterministic policy as a table of actions.

Represent a probabilistic policy as an matrix for states and actions.

Evaluating Policies for MDPs (take 1)

Previously,

For a specific policy (optimal or not),

Evaluating Policies for MDPs (take 2)

Rewrite

to

Evaluating Policies for MDPs (take 3a)

Rewrite

with

Evaluating Policies for MDPs (take 3b)

Rewrite

to

Rewriting with and is a rewrite as a Markov reward process.

And we already know how to solve for their values.

What is an Optimal Policy?

So, what is ?

A policy is optimal if and only if

Optimal policies always exist. Is that surprising?

Conclusions

- MRPs are easy to evaluate.
- MDP + policy = MRP

- But how do we pick an optimal policy?

Any Questions?

???