Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 2

Sep, 11, 2017
Lecture Overview

Value of Information and Value of Control

Recap Markov Chain

Markov Decision Processes (MDPs)

- Formal Specification and example
422 big picture

Deterministic

- Logics
  - First Order Logics
- Ontologies
  - Full Resolution
  - SAT

Stochastic

- Belief Nets
  - Approx. : Gibbs
- Markov Chains and HMMs
  - Forward, Viterbi...
  - Approx. : Particle Filtering
- Undirected Graphical Models
  - Markov Networks
  - Conditional Random Fields
- Markov Decision Processes and Partially Observable MDP
  - Value Iteration
  - Approx. Inference
- Reinforcement Learning

StarAI (statistical relational AI)
Hybrid: Det + Sto
- Prob CFG
- Prob Relational Models
- Markov Logics

Applications of AI

Query

Planning

Representation

Reasoning

Technique
Simple Decision Net

- Early in the morning. Shall I take my **umbrella** today? (I’ll have to go for a long walk at noon)
- Relevant Random Variables?
Polices for Umbrella Problem

- A **policy** specifies what an agent should do under each circumstance (for each decision, consider the parents of the decision node)

In the **Umbrella** case:

- \( D_1 \) : ? T F
- \( pD_1 \): Rainy, Cloudy, Sunny

**How many policies?**

- \( 1pD_1 \): 3 policies
- \( 1D_1 \): 2 policies

**One possible Policy**

- \( \rightarrow R \): T F T T...
- \( \rightarrow C \): T F T T...
- \( \rightarrow S \): F F F T...
Value of Information

- Early in the morning. I listen to the weather forecast, shall I take my umbrella today? (I’ll have to go for a long walk at noon)
- What would help the agent make a better Umbrella decision?
The value of information of a random variable $X$ for decision $D$ is: $EU(\text{knowing } X) - EU(\text{not knowing } X)$

the utility of the network with an arc from $X$ to $D$ minus

the utility of the network without the arc.

Intuitively:

- The value of information is always $\geq 0$
- It is positive only if the agent changes its policy
Value of Information (cont.)

- The value of information provides a bound on how much you should be prepared to pay for a sensor. How much is a **perfect** weather forecast worth?

- Original maximum expected utility:
  - 77

- Maximum expected utility when we know Weather:
  - 91

- Better forecast is worth at most: 14
Value of Information

- The value of information provides a bound on how much you should be prepared to pay for a sensor. How much is a **perfect** fire sensor worth?

- Original maximum expected utility: $-22.6$
- Maximum expected utility when we know Fire: $-2$
- Perfect fire sensor is worth: $20.6$
Value of Control

- What would help the agent to make an even better Umbrella decision? To maximize its utility.

<table>
<thead>
<tr>
<th>Weather</th>
<th>Umbrella</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain</td>
<td>true</td>
<td>70</td>
</tr>
<tr>
<td>Rain</td>
<td>false</td>
<td>0</td>
</tr>
<tr>
<td>noRain</td>
<td>true</td>
<td>20</td>
</tr>
<tr>
<td>noRain</td>
<td>false</td>
<td>100</td>
</tr>
</tbody>
</table>

- The value of control of a variable $X$ is: the utility of the network when you make $X$ a decision variable minus the utility of the network when $X$ is a random variable.
Value of Control

• What if we could control the weather?

- Original maximum expected utility: 77
- Maximum expected utility when we control the weather: 100
- Value of control of the weather: 23
Value of Control

- What if we control Tampering?

- Original maximum expected utility: \(-22.6\)
- Maximum expected utility when we control the Tampering: \(-20.7\)
- Value of control of Tampering: \(1.9\)
- Let’s take a look at the optimal policy
- Conclusion: do not tamper with fire alarms!
Lecture Overview

Value of Information and Value of Control

Recap Markov Chain

Markov Decision Processes (MDPs)

• Formal Specification and example
Combining ideas for Stochastic planning

- What is a key limitation of decision networks?
  
  \textit{Represent (and optimize) only a fixed number of decisions}

- What is an advantage of Markov models?
  
  \textit{The network can extend indefinitely}

\textbf{Goal: represent (and optimize) an indefinite sequence of decisions}
Decision Processes

Often an agent needs to go beyond a fixed set of decisions — Examples?

- Would like to have an ongoing decision process

Infinite horizon problems: process does not stop

- Robot surviving on planet, Monitoring Nuc. Plant, ...

Indefinite horizon problem: the agent does not know when the process may stop

Finite horizon: the process must end at a give time N in N steps
Lecture Overview (from my 322)

- Recap
- Temporal Probabilistic Models
- Start Markov Models
  - Markov Chain
  - Markov Chains in Natural Language Processing
Simplest Possible DBN

- One random variable for each time slice: let’s assume $S_t$ represents the state at time $t$ with domain $\{v_1, \ldots, v_n\}$

- Each random variable depends only on the previous one

- Thus $P(S_{t+1} | S_0, \ldots, S_t) = P(S_{t+1} | S_t)$

- Intuitively $S_t$ conveys all of the information about the history that can affect the future states.

- “The future is independent of the past given the present.”
• Stationary process assumption: the mechanism that regulates how state variables change overtime is stationary; that is, it can be described by a single transition model.

• \( P(S_t|S_{t-1}) \) is the same for all \( t \)
A stationary Markov Chain: for all \( t > 0 \)
- \( P(S_{t+1} \mid S_0, \ldots, S_t) = P(S_{t+1} \mid S_t) \) and
- \( P(S_{t+1} \mid S_t) \) is the same

We only need to specify
- Simple Model, easy to specify
- Often the natural model
- The network can extend indefinitely
- Variations of SMC are at the core of most Natural Language Processing (NLP) applications!

Variations of SMC are also used in the PageRank algo (used by Google to rank web pages).
Stationary Markov Chain (SMC)

A stationary Markov Chain: for all $t > 0$

- $P(S_{t+1} \mid S_0, \ldots, S_t) = P(S_{t+1} \mid S_t)$ and
- $P(S_{t+1} \mid S_t)$ is the same

So we only need to specify?

A. $P(S_{t+1} \mid S_t)$ and $P(S_0)$

B. $P(S_0)$

C. $P(S_{t+1} \mid S_t)$

D. $P(S_t \mid S_{t+1})$
Stationary Markov–Chain: Example

Domain of variable $S_t$ is \{t, q, p, a, h, e\}

Probability of initial state $P(S_0)$

Stochastic Transition Matrix $P(S_{t+1} | S_t)$

Which of these two is a possible STM?

A. Left one only
B. Right one only
C. Both
D. None

\[ \begin{array}{cccccc} t & q & p & a & h & e \\ \hline t & 0 & .3 & 0 & .3 & .4 & 0 \\ q & .4 & 0 & .6 & 0 & 0 & 0 \\ p & 0 & 0 & 1 & 0 & 0 & 0 \\ a & 0 & 0 & .4 & .6 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 & 1 \\ e & 1 & 0 & 0 & 0 & 0 & 0 \end{array} \]

\[ \begin{array}{cccccc} t & q & p & a & h & e \\ \hline t & 1 & 0 & 0 & 0 & 0 & 0 \\ q & 0 & 1 & 0 & 0 & 0 & 0 \\ p & .3 & 0 & 1 & 0 & 0 & 0 \\ a & 0 & 0 & 0 & 1 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 & 1 \\ e & 0 & 0 & 0 & .2 & 0 & 1 \end{array} \]
### Stationary Markov–Chain: Example

Domain of variable $S_i$ is \{t , q, p, a, h, e\}

We only need to specify \[ P(S_0) \]

Probability of initial state

Stochastic Transition Matrix

\[
P(S_{t+1} | S_t)
\]

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>q</th>
<th>p</th>
<th>a</th>
<th>h</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>0</td>
<td>.3</td>
<td>0</td>
<td>.3</td>
<td>.4</td>
<td>0</td>
</tr>
<tr>
<td>q</td>
<td>.4</td>
<td>0</td>
<td>.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a</td>
<td>0</td>
<td>0</td>
<td>.4</td>
<td>.6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>h</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>e</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Six possible values

Domain of variable $S_i$ is \{t , q, p, a, h, e\}

Stochastic Transition Matrix
Markov-Chain: Inference

Probability of a sequence of states $S_0 \ldots S_T$

$$P(S_0, \ldots, S_T) = P(S_0) \cdot P(S_1 | S_0) \cdot P(S_2 | S_1) \cdot \ldots$$

Example:

$$P(t, q, p) = P(t) \cdot P(q | t) \cdot P(p | q)$$

$$= .6 \times .3 \times .6 = .108$$
Recap: Markov Models

$P(S_0)$ \( \sim \) $n$

$P(S_{t+1} | S_t)$ \( \sim \) $n \times n$

$P(O_t | S_t)$ \( \sim \) $n \times K$

$|\text{dom}(S)| = n$

$|\text{dom}(O)| = K$
Lecture Overview

Value of Information and Value of Control

Recap Markov Chain

Markov Decision Processes (MDPs)
• Formal Specification and example
Markov Models

- Markov Chains
- Hidden Markov Model
- Partially Observable Markov Decision Processes (POMDPs)
- Markov Decision Processes (MDPs)

Noisy Observations
Noisy Actions
Noisy Outcomes
Rewards
How can we deal with indefinite/infinite Decision processes?

We make the same two assumptions we made for\ldots.

The action outcome depends only on the current state

Let $S_t$ be the state at time $t$ \ldots

The process is *stationary* \ldots

We also need a more flexible specification for the utility. How?

- Defined based on a reward/punishment $R(s)$ that the agent receives in each state $s$

\[
\begin{align*}
\sum_{s_0, s_1, \ldots, s_n} & r_0, r_1, \ldots, r_n \\
\end{align*}
\]
MDP: formal specification

For an MDP you specify:

- set $S$ of states and set $A$ of actions
- the process’ dynamics (or \textit{transition model})
  \[ P(S_{t+1} | S_t, A_t) \]
- The \textbf{reward function}
  \[ R(s, a, s') \]
  describing the reward that the agent receives when it performs action $a$ in state $s$ and ends up in state $s'$
- $R(s)$ is used when the reward depends only on the state $s$ and not on how the agent got there
- Absorbing/stopping/terminal state
  \[ P(S_{ab} | a, S_{ab}) = 1 \quad R(S_{ab}, a, S_{ab}) = 0 \]
MDP graphical specification

Basically a MDP is a Markov Chain augmented with actions and rewards/values
When Rewards only depend on the state
Summary Decision Processes: MDPs

To manage an ongoing (indefinite... infinite) decision process, we combine...

Markov Chains & Decision Networks

Markovian
Stationary

Utility not just at the end
But Sequence of rewards

Fully Observable
Example MDP: Scenario and Actions

Agent moves in the above grid via actions Up, Down, Left, Right.

Each action has:
- 0.8 probability to reach its intended effect
- 0.1 probability to move at right angles of the intended direction
- If the agents bumps into a wall, it stays there

How many states?

There are two terminal states (3,4) and (2,4)
Example MDP: Rewards

\[ R(s) = \begin{cases} 
-0.04 & \text{(small penalty) for nonterminal states} \\
\pm 1 & \text{for terminal states} 
\end{cases} \]
Learning Goals for today’s class

You can:

- Define and compute Value of Information and Value of Control in a decision network
- Effectively represent indefinite/infinite decision processes with a Markov Decision Process (MDP)
- Compute the probability distribution on states given a sequence of actions in an MDP
- Define a policy for an MDP
• Read textbook 9.4
• Read textbook 9.5
  • 9.5.1 Value of a Policy
  • 9.5.2 Value of an Optimal Policy
  • 9.5.3 Value Iteration