CSC242: Intro to AI

Lecture 20
Uncertainty Over Time
• Representation of state: $X_t, E_t$

• Transition model: $P(X_t \mid X_{t-1})$

• Markov assumption, stationary process

• Sensor model: $P(E_t \mid X_t)$

• Sensor Markov assumption

• Prior distribution at time 0: $P(X_0)$
Temporal Model

\[ P(X_{0:t}, E_{1:t}) = P(X_0) \prod_{i=1}^{t} P(X_i | X_{i-1})P(E_i | X_i) \]

- Initial State Model
- Transition Model
- Sensor Model
Inference

- Filtering (State Estimation): $P(X_{t+1} \mid e_{1:t+1})$
- Prediction: $P(X_{t+k} \mid e_{1:t})$ for some $k > 0$
- Smoothing: $P(X_k \mid e_{1:t})$ for some $0 \leq k < t$
- Most Likely Explanation: $\arg\max_{x_{1:t}} P(x_{1:t} \mid e_{1:t+1})$
Decision-Making Under Uncertainty
Problem to Solve
Decision to Act!
Perform Action
New State of the World
Decision to Act
Perform Action
New State

Current State

Decision to Act

Current State

Decision to Act

Current State

Decision to Act

Current State
Decision-Making Under Uncertainty

- How to decide what to do given:
  - Uncertainty about the results of actions
  - Uncertainty about the current state
Problem (Domain): \( \langle S, A, \text{Actions}, \text{Result} \rangle \)

\text{Actions} : s \in S \rightarrow 
\{ a \in A : a \text{ can be executed (is applicable) in } s \} 

\text{Result} : s \in S, a \in A \rightarrow 

s' : s' \in S \text{ s.t. } s' \text{ is the result of performing } a \text{ in } s

Problem (Instance): \( \langle \mathcal{I} \in S, \mathcal{G} \subseteq S \rangle \)

Solution: \( \langle a_0, a_1, \ldots, a_n \rangle \in A \text{ s.t.} \)

\text{Result}(\cdots \text{Result}(\text{Result}(\mathcal{I}, a_0), a_1) \cdots , a_n) \in \mathcal{G}
Deterministic Result Function

\[ \text{Result}(a, s) = s' \]

where \( s' \) is the result of performing action \( a \) in state \( s \)
Nondeterministic Result Function

\[ \text{Result}(a, s) = \{ s' \} \]
Nondeterministic Result Function

\[ \text{Result}(a, s) \]

Random variable whose values are the states \( s' \) that are possible as the result of performing action \( a \) in state \( s \)
Distribution on Results

\[ P(\text{Result}(a) \mid a, e) \]

\[ P(\text{Result}(a) = s' \mid a, e) \]
Distribution on Results

\[
P(\text{Result}(a) \mid a, e)
\]

\[
P(\text{Result}(a) = s' \mid a, e)
\]

\[
= \sum_s P(\text{Result}(a, s) = s' \mid a) P(s_0 = s \mid e)
\]
Distribution on Results

\[ P(\text{Result}(a) \mid a, e) \]

\[ P(\text{Result}(a) = s' \mid a, e) \]
The diagram illustrates the process of decision-making and state transitions. It shows a sequence from the current state to a new state after performing an action. The equation $P(Result(a) = s' | a, e)$ represents the probability of the result of an action given the action and the environment.
Goal State/Test

- The state(s) we want to be in
- A state either is or isn’t a goal state
- => A state is or isn’t desirable
Utility
Utility

utility  |yooˈtilət̬|  
noun (pl. utilities)  
1 the state of being useful, profitable, or beneficial: he had a poor opinion of the utility of book learning.
utility  |ˈyooˌtɪlət̬i|  
noun ( pl. utilities )  
1 the state of being useful, profitable, or beneficial:  
he had a poor opinion of the utility of book learning.  
• (in game theory or economics) a measure of that  
which is sought to be maximized in any situation  
involving a choice.
Utility

$U(s)$

Assigns a single number to the desirability of a state
Expected Utility

\[ EU(a \mid e) = \sum_{s'} P(\text{Result}(a) = s' \mid a, e) U(s') \]
$$EU(a \mid e) = \sum_{s'} P(\text{Result}(a) = s' \mid a, e) U(s')$$
Maximum Expected Utility

\[\text{EU}(a \mid e) = \sum_{s'} P(\text{Result}(a) = s' \mid a, e) \text{U}(s')\]

\[a_{\text{MEU}} = \arg\max_a \text{EU}(a \mid e)\]
Utility Theory

- Why is MEU the only rational way to make decisions?
- Why should maximizing average utility be so special?
- What’s the relationship between utility functions and preferences?
Utility Functions

• How can we elicit/design/construct a utility function that accurately represents a real-world agent’s (person’s) preferences?

• Is monetary value a good form of utility? Spoiler: no.

• Are humans always rational? Spoiler: no.

• c.f. Kahneman & Tversky
Maximum Expected Utility

\[ EU(a \mid e) = \sum_{s'} P(\text{Result}(a) = s' \mid a, e) U(s') \]

\[ a_{MEU} = \arg\max_a EU(a \mid e) \]
Sequential Decision Problem

- Utility depends on a sequence of decisions
- Incorporate uncertainty, utility, sensing
- Include search and planning as special cases
$$Actions(s) = \{ \text{Up, Down, Left, Right} \}$$
[ Up, Up, Right, Right, Right ]
Nondeterministic Actions
[ $Up, Up, Right, Right, Right$ ]

$0.8^5 = 0.32768$
[ Up, Up, Right, Right, Right ]

\[ 0.1^4 \times 0.8 = 0.00008 \]
Stochastic Transition Model

\[ P(s' \mid s, a) \]

\[ P(\text{Result}(a) = s' \mid a, e) \]
Markov Assumption

\[ P(s' \mid s, a) \]

\[ P(s' \mid s, s_{-1}, s_{-2}, \ldots, a) = P(s' \mid s, a) \]
**Reward Function**

- In each state $s$ the agent receives reward $R(s)$:
  - $R(s) = +1$ or $-1$ in terminal state
  - $R(s) = -0.04$ otherwise
Markov Decision Process

- Sequential decision problem
- Fully observable, stochastic environment
  - Set of states $S$ with initial state $s_0$
  - Markovian transition model: $P(s' | s, a)$
  - Reward function: $R(s)$
[ Up, Up, Right, Right, Right ]
[ Up, Up, Right, Right, Right ]
Policy

- A policy $\pi$ specifies what the agent should do for any state the agent might reach: $\pi(s)$
Policy Quality

- A policy \( \pi \) specifies what the agent should do for any state the agent might reach: \( \pi(s) \)
- Each time a policy is executed, it leads to a different history
Policies Generate State Sequences

\[ \text{Exec}(\pi, s_0) = \{s_0, s_1, \ldots, s_n\} \]

\[ P(s_1 \mid s_0, \pi(s_0))P(s_2 \mid s_1, \pi(s_1)) \cdots P(s_n \mid s_{n-1}, \pi(s_{n-1})) \]
Policies Generate State Sequences

\[ S^{\pi,s} \]

Random variable over state sequences generated by executing policy \( \pi \) starting in state \( s \)
Policy Quality

- A policy $\pi$ specifies what the agent should do for any state the agent might reach: $\pi(s)$
- Each time a policy is executed, it leads to a different history
Policy Quality

- A policy $\pi$ specifies what the agent should do for any state the agent might reach: $\pi(s)$
- Each time a policy is executed, it leads to a different history
- Quality of a policy is its expected utility
Expected Utility of a Policy

\[ EU(\pi, s) = \sum_{S_{\pi,s}^{\pi,s} \in S_{\pi,s}} P(S_{\pi,s}^{\pi,s}) U(S_{\pi,s}^{\pi,s}) \]
Expected Utility of a Policy

\[
EU(\pi, s) = \sum_{S_i^{\pi,s} \in S^{\pi,s}} P(S_i^{\pi,s}) U(S_i^{\pi,s})
\]

\[
\prod_{i=1}^{n} P(s_i | s_{i-1}, \pi(s_{i-1}))
\]
Expected Utility of a Policy

\[ EU(\pi, s) = \sum_{S_i^{\pi, s} \in S^{\pi, s}} P(S_i^{\pi, s}) U(S_i^{\pi, s}) \]

\[ \prod_{i=1}^{n} P(s_i | s_{i-1}, \pi(s_{i-1})) \]
Additive Rewards

$$EU(\pi, s) = \sum_{S_i^{\pi,s} \in S^{\pi,s}} P(S_i^{\pi,s}) U(S_i^{\pi,s})$$

$$\prod_{i=1}^{n} P(s_i | s_{i-1}, \pi(s_{i-1}))$$

$$\sum_{i=0}^{n} R(s_i)$$
Discounted Rewards

\[ EU(\pi, s) = \sum_{S_i^{\pi,s} \in S^{\pi,s}} P(S_i^{\pi,s}) U(S_i^{\pi,s}) \]

\[ \prod_{i=1}^n P(s_i|s_{i-1}, \pi(s_{i-1})) \]

\[ \sum_{i=0}^n \gamma^i R(s_i) \]
Expected Utility of a Policy

\[ EU(\pi, s) = \sum_{S_i^{\pi, s} \in S^{\pi, s}} P(S_i^{\pi, s}) U(S_i^{\pi, s}) \]

\[ U^\pi(s) = E \left[ \sum_{t=0}^{\infty} \gamma^t R(S_t) \right] \]
Policy Quality

• A policy $\pi$ specifies what the agent should do for any state the agent might reach: $\pi(s)$

• Each time a policy is executed, it leads to a different history

• Quality of a policy is its expected utility
Maximum Expected Utility (MEU) Policy

\[ EU(\pi, s) = \sum_{S_i^\pi, s \in S^\pi, s} P(S_i^\pi, s) U(S_i^\pi, s) \]

\[ U^\pi(s) = E \left[ \sum_{t=0}^{\infty} \gamma^t R(S_t) \right] \]

\[ \pi_s^* = \arg\max_{\pi} U^\pi(s) \]
Policy

• A policy \( \pi \) specifies what the agent should do for any state the agent might reach: \( \pi(s) \)

• Each time a policy is executed, it leads to a different history

• Quality of a policy is its expected utility

• Optimal policy \( \pi^* \) maximizes expected utility
Optimal policies balance risk and reward.
$R(s) < -1.6284$
-0.4278 < R(s) < -0.0850
-0.0221 < R(s) < 0
\[ R(s) > 0 \]
\[ R(s) < -1.6284 \]

\[ -0.4278 < R(s) < -0.0850 \]

\[ -0.0221 < R(s) < 0 \]

\[ R(s) > 0 \]
Markov Decision Process

- Sequential decision problem
- Fully observable, stochastic environment
- Markovian transition model: $P(s' \mid s, a)$
- Additive or discounted rewards: $R(s)$
- Policy: $\pi(s) = \text{action to perform in state } s$
- Quality of a policy is its expected utility
- Optimal policy $\pi^*$ maximizes expected utility
Computing Optimal Policies

• Given MDP (transition & reward models)
• Compute optimal policy
• Inference!
Further Reading

- Partially-observable environments
- POMDPs
- Multi-agent systems
- AIMA 17.5: Game Theory
- AIMA 17.6: Design of MAS
For Next Time:
Quiz
AIMA 18.0-18.4