Recall Estimation Example

Using Bayes filtering, we can recursively update our belief about the state of the door:

\[ \text{bel}(X_c = 2 = \text{is open}) = 0.983 \]

- Should we take an action?
- Should we probe or explore?
Decision Making Systems

Methods
1. Explicit programming
2. Supervised learning
3. Optimization / optimal control
4. Planning
   • Given a stochastic model, how to algorithmically determine best policy?
5. Reinforcement Learning
   • If model is unknown (or very complex), learn policy through experience
Traffic Alert and Collision Avoidance System (TCAS)

IF (ITF.A LT G.ZTHR) THEN IF(ABS(ITF.VMD) LT G.ZTHR) THEN SET ZHIT; ELSE CLEAR ZHIT; ELSE IF (ITF.ADOT GE P.ZDTHR) THEN CLEAR ZHIT ELSE ITF.TAUV = -ITF.A/ITF.ADOT; IF (ITF.TAUV LT TVTHR AND ((ABS(ITF.VMD) LT G.ZTHR) OR (ITF.TAUV LT ITF.TRTRU))) THEN SET ZHIT ELSE CLEAR ZHIT IF (ZHIT EQ $TRUE AND ABS(ITF.ZDINT) GT P.MAXZDINT THEN CLEAR ZHIT IF (ITF.A LT G.ZTHR) THEN IF(ABS(ITF.VMD) LT G.ZTHR) THEN SET ZHIT; ELSE CLEAR ZHIT; ELSE IF (ITF.ADOT GE P.ZDTHR) THEN CLEAR ZHIT ELSE ITF.TAUV = -ITF.A/ITF.ADOT; IF (ITF.TAUV LT TVTHR AND ((ABS(ITF.VMD) LT G.ZTHR) OR (ITF.TAUV LT ITF.TRTRU))) THEN SET ZHIT ELSE CLEAR ZHIT IF (ZHIT EQ $TRUE AND ABS(ITF.ZDINT) GT P.MAXZDINT THEN CLEAR ZHIT
PROCESS Reverse_modeling:

- Default modeled separation for current RA is 0 if current RA is negative.
- Set own altitude and own rate to own tracked altitude and own tracked rate.
- \textbf{IF} (even does not follow his RA):
  \textbf{THEN} Model separation achieved remaining RA not followed;
  \textbf{ELSE} (current RA is a climb RA):
  \textbf{THEN} CLEAR flag indicating the sense of the RA after a reversal;
  \textbf{ELSE} (model separation achieved by changing current RA greater than 1.2):
  \textbf{IF} (flight path is not toward the target): CLEAR reverse flag as ITF
  \textbf{ELSE} (begin own is intended to follow as RA):
  \textbf{IF} (current RA is greater): model response to current RA:
  \textbf{ELSE} (current RA is a pitch RA): model minimum displayable rate for climb if current rate exceeds minimum displayable rate or minimum displayable rate for descent if current rate is less than minimum displayable rate:
  \textbf{ELSE} (tracked response line models response to an RA direction and time since RA less than a parameter time AND own rate has not changed by more than a parameter):
  \textbf{THEN} set own altitude and own rate to modeled altitude and rate for use in reverse modeling;

- Model separation achieved by changing current RA:
- Set delay time to generate of pilot delay time remaining for last advisory against a new target, and the pilot quick reaction time:
- \textbf{IF} (considering a reversal from a descend RA to a climb RA):
  \textbf{THEN} set own goal rate to generate of own tracked rate (or minimum displayable rate, whichever is less) and minimum climb rate;
  \textbf{ELSE} (even too close to ground to descend):
  \textbf{THEN} set own goal rate to zero;
  \textbf{ELSE} (vertical climb, low VMD geometry was not the reason for considering reversal):
  \textbf{THEN} set own goal rate to be less than own tracked rate (or minimum displayable rate, whichever is greater) and minimum descent rate:

- \textbf{ELSE} (vertical climb, low VMD geometry was not the reason for considering reversal):
  \textbf{THEN} (imagine crossing crossing OR (imagine local AND own crossing):
  \textbf{THEN} use own rate bound to modeled rate:
  \textbf{ELSE} (rate rate bound to model rate):
  \textbf{THEN} (rate rate bound to modeled rate):
  \textbf{ELSE} (rate rate bound to model rate):

- CALL MODEL SEP:
  \textbf{IF} (altitude, goal rate, own altitude, own rate, acceleration response, sensor other reversal, unidirectional, modeled sensor rate, ITF entry)
  \textbf{OUT} (predicted separation for sensor reversal):
  \textbf{ELSE} (predicted separation for sensor reversal is not positive OR (model separation achieved by changing current RA > G.ADM):
  \textbf{THEN} CLEAR reverse flag as ITF
  \textbf{ELSE} (begin own is assumed to follow as RA):

END Reverse_modeling;

RESOLUTION HIGH LEVEL LOGIC

6-92

M.J. Kochenderfer
Why is it hard?

- State Uncertainty
- Dynamic Uncertainty
- Multiple Objectives

Imperfect sensor information leads to uncertainty in position and velocity of aircraft. Variability in pilot behavior makes it difficult to predict future trajectories of aircraft. System must carefully balance both safety and operational considerations.

Slide Credit: Mykel Kochenderfer
## Simplified MDP

<table>
<thead>
<tr>
<th>State space</th>
<th>Action space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative altitude</td>
<td>Clear of conflict</td>
</tr>
<tr>
<td>Own vertical rate</td>
<td>Climb &gt; 1500 ft/min</td>
</tr>
<tr>
<td>Intruder vertical rate</td>
<td>Climb &gt; 2500 ft/min</td>
</tr>
<tr>
<td>Time to lateral NMAC</td>
<td>Descend &gt; 1500 ft/min</td>
</tr>
<tr>
<td>State of advisory</td>
<td>Descend &gt; 2500 ft/min</td>
</tr>
</tbody>
</table>

### Dynamic model

<table>
<thead>
<tr>
<th>Own aircraft</th>
<th>Intruder Aircraft</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACAS X</td>
<td>Slide Credit: Mykel Kochenderfer</td>
</tr>
</tbody>
</table>

### Reward model

<table>
<thead>
<tr>
<th>State of advisory</th>
<th>Dynamic model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear of conflict</td>
<td>Head-on, constant closure</td>
</tr>
<tr>
<td>Alert (-0.01)</td>
<td>Random vertical acceleration</td>
</tr>
<tr>
<td>Reversal (-0.01)</td>
<td>Pilot response delay (5 s)</td>
</tr>
<tr>
<td>Strengthen (-0.009)</td>
<td>Pilot response strength (1/4 g)</td>
</tr>
<tr>
<td>Clear of conflict (0.0001)</td>
<td>State of advisory</td>
</tr>
</tbody>
</table>

- **Slide Credit:** Mykel Kochenderfer
Simplified MDP

State space
- Relative altitude
- Own vertical rate
- Intruder vertical rate
- Time to lateral NMAC
- State of advisory

Action space
- Clear of conflict
- Climb > 1500 ft/min
- Climb > 2500 ft/min
- Descend > 1500 ft/min
- Descend > 2500 ft/min

Dynamic model
- Head-on, constant closure
- Random vertical acceleration
- Pilot response delay (5 s)
- Pilot response strength (1/4 g)
- State of advisory

Reward model
- NMAC (-1)
- Alert (-0.01)
- Reversal (-0.01)
- Strengthen (-0.009)
- Clear of conflict (0.0001)

Slide Credit: Mykel Kochenderfer
Optimized Logic
Both Own and Intruder Level

<table>
<thead>
<tr>
<th>Metric</th>
<th>TCAS</th>
<th>ACAS X</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMACs</td>
<td>169</td>
<td>3</td>
</tr>
<tr>
<td>Alerts</td>
<td>994,317</td>
<td>690,406</td>
</tr>
<tr>
<td>Strengthens</td>
<td>40,470</td>
<td>92,946</td>
</tr>
<tr>
<td>Reversals</td>
<td>197,315</td>
<td>9,569</td>
</tr>
</tbody>
</table>

Slide Credit: Mykel Kochenderfer
Probabilities

1) conditional probability
\[ P(A|B) = \frac{P(A,B)}{P(B)} \]

2) total probability
\[ P(A) = \sum_{B \in \mathbb{B}} P(A|B)P(B) \]

3) Bayes Rule
\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>P(A,B,C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.10</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.14</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.18</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.19</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Bayesian Networks (1)
compact representations of a joint probability distribution

\[ P(X_i | \text{par}_i) \]

every node has a dist:

\[ P(c) \]

\[ P(\text{camera failure} | c) \]

\[ P(\text{perception failure} | \text{camera failure}, c) \]

\[ P(\text{trajectory deviation}) \]

\[ P(\text{safety violation}) \]
Continuous / Mixed Bayes Nets
Temporal Models: Markov Chains

\[ x_t \rightarrow x_{t+1} \rightarrow x_{t+2} \rightarrow \ldots \rightarrow x_n \]

\( x_t \) is the state of our system at time \( t \).

\[ \rightarrow \text{pos} + \text{vel} \]

\[ \begin{array}{c|cc}
    & S & R \\
\hline
S & 0.9 & 0.1 \\
R & 0.2 & 0.8 \\
\end{array} \]
Markov Assumptions and Common Violations

Markov Assumption postulates that past and future data are independent if you know the current state.

What are some common violations?

• Unmodeled dynamics in the environment not included in state
  • E.g., moving people and their effects on sensor measurements in localization

• Inaccuracies in the probabilistic model
  • E.g., error in the map of a localizing agent or incorrect model dynamics

• Approximation errors when using approximate representations
  • E.g., discretization errors from grids, Gaussian assumptions

• Variables in control scheme that influence multiple controls
  • E.g., the goal or target location will influence an entire sequence of control commands
Hidden Markov Models

inference tools:
- filtering $P(X_k|O_{0:k})$
- prediction $P(X_k|O_{0:t}), k < t$
- smoothing $P(X_k|O_{0:t}), k < t$
Sequential Decision Making (1)

Markov Decision Process \((S, A, T, R, \pi)\)

Transition function \(T(s'|s, a)\)

Reward function \(R(s, a)\)
Utility and Reward

MDP rewards are generally additive components in a Utility function:

for some finite horizon \( w \) w/ \( n \) decisions the utility associated w/ the sequence of rewards is

\[
\sum_{t=0}^{n-1} R_t
\]
Policies and Utility

- policy $\pi$ specifies what action to executed from every state:
  $$a_t \leftarrow \pi(s_t) \text{ or } \pi(o_t)$$

- Expected utility of from executing $\pi$ when starting at state $s$ is denoted $U^\pi(s)$

- optimal policy $\pi^*$ maximized expected utility
  $$\pi^*(s) = \arg\max_\pi U^\pi(s)$$
Models of Optimal Behavior

• In the finite-horizon model, the agent should optimize expected reward for the next H steps: \( E \left[ \sum_{t=0}^{H} r_t \right] \)
  • Continuously executing H-step optimal actions is known as receding horizon control

• In the infinite-horizon discounted model, agent should optimize:
  \( E \left[ \sum_{t=0}^{H} \gamma^t r_t \right] \)
  • Discount factor is between 0 and 1, can be thought of as interest rate (reward now is worth more than reward later)
  • Keeps the utility of an infinite sequence finite
Markov Models

<table>
<thead>
<tr>
<th>Markov Decision Processes (MDP)</th>
<th>Partially Observable MDP</th>
<th>Reinforcement Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>![MDP Diagram](MDP Diagram)</td>
<td>![PO-MDP Diagram](PO-MDP Diagram)</td>
<td>![RL Diagram](RL Diagram)</td>
</tr>
</tbody>
</table>

- **Uncertainty in effects of actions**
- **Uncertainty in current state**
- **Uncertainty in model**