422 big picture: Where are we?

Query Planning

Deterministic

- Logics
  - First Order Logics
- Ontologies
  - Temporal rep.
  - 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

Applications of AI

StarAI (statistical relational AI)

Hybrid: Det + Sto

- Prob CFG
- Prob Relational Models
- Markov Logics

Representation

Reasoning Technique
Lecture Overview (Temporal Inference)

- **Filtering** (posterior distribution over the current state given evidence to date)
  - From intuitive explanation to formal derivation
  - Example
- **Prediction** (posterior distribution over a future state given evidence to date)
- **(start) Smoothing** (posterior distribution over a past state given all evidence to date)
Markov Models

Markov Chains

Hidden Markov Model

Partially Observable Markov Decision Processes (POMDPs)

Markov Decision Processes (MDPs)
Hidden Markov Model

- A Hidden Markov Model (HMM) starts with a Markov chain, and adds a noisy observation/evidence about the state at each time step:

  \[ \mathbb{X}_0 \rightarrow \mathbb{X}_1 \rightarrow \mathbb{X}_2 \rightarrow \mathbb{X}_3 \rightarrow \mathbb{X}_4 \rightarrow \ldots \]

  \[ \mathbb{E}_0 \rightarrow \mathbb{E}_1 \rightarrow \mathbb{E}_2 \rightarrow \mathbb{E}_3 \rightarrow \mathbb{E}_4 \rightarrow \ldots \]

- \( P(X_0) \) specifies initial conditions
- \( P(X_{t+1} | X_t) \) specifies the dynamics
- \( P(E_t | S_t) \) specifies the sensor model

\( |\text{domain}(X)| = k \)
\( |\text{domain}(E)| = h \)
Simple Example

(We’ll use this as a running example)

- Guard stuck in a high-security bunker
- Would like to know if it is raining outside
- Can only tell by looking at whether his boss comes into the bunker with an umbrella every day

<table>
<thead>
<tr>
<th>Transition model</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$R_{t-1}$</th>
<th>$P(R_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>0.7</td>
</tr>
<tr>
<td>$f$</td>
<td>0.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observation model</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$R_t$</th>
<th>$P(U_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>0.9</td>
</tr>
<tr>
<td>$f$</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Transition model:

- $R_t$: Rain at time $t$ (true or false)
- $U_t$: Umbrella used by boss (true or false)

State variables:

- Rain$_{t-1}$
- Rain$_t$
- Rain$_{t+1}$

Observable variables:

- Umbrella$_{t-1}$
- Umbrella$_t$
- Umbrella$_{t+1}$
Useful inference in HMMs

- In general (Filtering): compute the posterior distribution over the current state given all evidence to date

\[ P(X_t \mid e_{0:t}) \]
Intuitive Explanation for filtering recursive formula

\[ P(X_t | e_{0:t}) = \alpha P(e_t | X_t) \times \sum_{X_{t-1}} P(X_t | X_{t-1}) \times P(X_{t-1} | e_0 : e_{t-1}) \]

- \( X_t \) generated evidence \( e_t \)
- \( e_t \) and evidence \( e_0 : e_{t-1} \) must have been generated before getting to \( X_{t-1} \)
- \( X_t \) was reached from whatever \( X_{t-1} \) was, there
Filtering

- Idea: recursive approach
  - Compute filtering up to time $t-1$, and then include the evidence for time $t$ (recursive estimation)

$P(X_t | e_{0:t}) = P(X_t | e_{0:t-1}, e_t)$ dividing up the evidence

$$= \alpha P(e_t | X_t, e_{0:t-1}) P(X_t | e_{0:t-1}) \quad \text{WHY?}$$

$$= \alpha P(e_t | X_t) P(X_t | e_{0:t-1}) \quad \text{WHY?}$$

Inclusion of new evidence: this is available from..

So we only need to compute $P(X_t | e_{0:t-1})$
\[
P(x, y, z) = P(x | y, z) \cdot P(y, z)
\]
\[
P(x, y, z) = P(y | x, z) \cdot P(x, z)
\]

\[
P(y | x, z) = \frac{P(x | y, z) \cdot P(y, z)}{P(x, z)} = P(y | x) \cdot P(z)
\]

\[
\sum_{y} p(x \mid y, z) = \frac{p(x \mid y, z) \cdot p(y, z)}{p(x \mid z)} = \alpha \cdot p(x \mid y, z) \cdot p(y \mid z)
\]
“moving” the conditioning

\[
P(AB|C) = \frac{P(ABC)}{P(C)} \times \frac{P(BC)}{P(BC)} =
\]

\[
= \frac{P(ABC)}{P(BC)} \times \frac{P(BC)}{P(C)} =
\]

\[
= P(A|BC) \times P(B|C)
\]
Filtering

- Compute $P(X_t \mid e_{0:t-1})$

  $P(X_t \mid e_{0:t-1}) = \sum_{x_{t-1}} P(X_t, x_{t-1} \mid e_{0:t-1}) = \sum_{x_{t-1}} P(X_t \mid x_{t-1}, e_{0:t-1}) P(x_{t-1} \mid e_{0:t-1}) = \sum_{x_{t-1}} P(X_t \mid x_{t-1}) P(x_{t-1} \mid e_{0:t-1})$
  because of...

- Transition model!

- Putting it all together, we have the desired recursive formulation

  $P(X_t \mid e_{0:t}) = \alpha P(e_t \mid X_t) \sum_{x_{t-1}} P(X_t \mid x_{t-1}) P(x_{t-1} \mid e_{0:t-1})$

  - Inclusion of new evidence (sensor model)
  - Filtering at time $t-1$

  - Propagation to time $t$

- $P(X_{t-1} \mid e_{0:t-1})$ can be seen as a message $f_{0:t-1}$ that is propagated forward along the sequence, modified by each transition and updated by each observation

- Prove it?
Thus, the recursive definition of filtering at time $t$ in terms of filtering at time $t-1$ can be expressed as a FORWARD procedure:

- $f_{0:t} = \alpha \text{FORWARD}(f_{0:t-1}, e_t)$

which implements the update described in:

$$P(X_t \mid e_{0:t}) = \alpha P(e_t \mid X_t) \sum_{x_{t-1}} P(X_t \mid x_{t-1}) P(x_{t-1} \mid e_{0:t-1})$$

Inclusion of new evidence (sensor model)

Propagation to time $t$

Filtering at time $t-1$
Analysis of Filtering

- Because of the recursive definition in terms for the forward message, when all variables are discrete the time for each update is constant (i.e. independent of $t$)

- The constant depends of course on the size of the state space
Rain Example

- Suppose our security guard came with a prior belief of 0.5 that it rained on day 0, just before the observation sequence started.

- Without loss of generality, this can be modelled with a fictitious state $R_0$ with no associated observation and $P(R_0) = <0.5, 0.5>$

- **Day 1:** umbrella appears ($u_1$). Thus

  $$P(R_1 \mid e_{0:1-1}) = P(R_1) = \sum_{r_0} P(R_1 \mid r_0) P(r_0)$$

  $$= <0.7, 0.3> \times 0.5 + <0.3, 0.7> \times 0.5 = <0.5, 0.5>$$

  **TRUE** 0.5
  **FALSE** 0.5

![Diagram of Rain Example]

<table>
<thead>
<tr>
<th>$R_t$</th>
<th>$P(R_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>0.7</td>
</tr>
<tr>
<td>$f$</td>
<td>0.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$R_t$</th>
<th>$P(U_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>0.9</td>
</tr>
<tr>
<td>$f$</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Rain Example

- Updating this with evidence from for $t=1$ (umbrella appeared) gives

$$P(R_1 | u_1) = \alpha P(u_1 | R_1) P(R_1) =$$

$$\alpha <0.9, 0.2> <0.5, 0.5> = \alpha <0.45, 0.1> \sim <0.818, 0.182>$$

- Day 2: umbrella appears ($u_2$). Thus

$$P(R_2 | e_{0:t-1}) = P(R_2 | u_1) = \sum_{r_1} P(R_2 | r_1) P(r_1 | u_1) =$$

$$= <0.7, 0.3> * 0.818 + <0.3,0.7> * 0.182 \sim <0.627,0.373>$$
Rain Example

- Updating this with evidence from for $t = 2$ (umbrella appeared) gives

$$P(R_2 | u_1, u_2) = \alpha P(u_2 | R_2) P(R_2 | u_1) = \alpha \langle 0.9, 0.2 \rangle \langle 0.627, 0.373 \rangle = \alpha \langle 0.565, 0.075 \rangle \sim \langle 0.883, 0.117 \rangle$$

- Intuitively, the probability of rain increases, because the umbrella appears twice in a row
Practice exercise (home)

Compute filtering at $t_3$ if the 3rd observation/evidence is *no umbrella* (will put solution on inked slides)

$$\langle 0.7, 0.3 \rangle \times 0.883 + \langle 0.3, 0.7 \rangle \times 0.117$$

$$\langle 0.618, 0.264 \rangle + \langle 0.035, 0.081 \rangle = \langle 0.653, 0.345 \rangle$$

$$\lambda \langle 0.653, 0.345 \rangle \times \langle 0.1, 0.8 \rangle$$

$$\lambda \langle 0.065, 0.276 \rangle$$

Normalize/divide by the sum.

$$0.19 \quad 0.81$$
Lecture Overview

• **Filtering** (posterior distribution over the current state given evidence to date)
  • From intuitive explanation to formal derivation
  • Example

• **Prediction** (posterior distribution over a future state given evidence to date)

• (start) **Smoothing** (posterior distribution over a *past* state given all evidence to date)
Prediction $P(X_{t+k+1} \mid e_{0:t})$

- Can be seen as filtering without addition of new evidence
- In fact, filtering already contains a one-step prediction
  
  $$P(X_t \mid e_{0:t}) = a P(e_t \mid X_t) \sum_{x_{t-1}} P(X_t \mid x_{t-1}) \cdot P(x_{t-1} \mid e_{0:t-1})$$

  Inclusion of new evidence (sensor model)

  Filtering at time $t-1$

  Propagation to time $t$

- We need to show how to recursively predict the state at time $t+k+1$ from a prediction for state $t+k$
  
  $$P(X_{t+k+1} \mid e_{0:t}) = \sum_{x_{t+k}} P(X_{t+k+1}, x_{t+k} \mid e_{0:t}) = \sum_{x_{t+k}} P(X_{t+k+1} \mid x_{t+k}, e_{0:t}) \cdot P(x_{t+k} \mid e_{0:t}) =$$
  
  $$= \sum_{x_{t+k}} P(X_{t+k+1} \mid x_{t+k}) \cdot P(x_{t+k} \mid e_{0:t})$$

  Transition model

  Prediction for state $t+k$

- Let’s continue with the rain example and compute the probability of Rain on day four after having seen the umbrella in day one and two: $P(R_4 \mid u_1, u_2)$
Rain Example

- Prediction from day 2 to day 3

\[
P(X_3 | e_{1:2}) = \sum_{x_2} P(X_3 | x_2) P(x_2 | e_{1:2}) = \sum_{r_2} P(R_3 | r_2) P(r_2 | u_1 u_2) =
\]
\[
= \langle 0.7, 0.3 \rangle \times 0.883 + \langle 0.3, 0.7 \rangle \times 0.117 = \langle 0.618, 0.265 \rangle + \langle 0.035, 0.082 \rangle
\]
\[
= \langle 0.653, 0.347 \rangle
\]

- Prediction from day 3 to day 4

\[
P(X_4 | e_{1:2}) = \sum_{x_3} P(X_4 | x_3) P(x_3 | e_{1:2}) = \sum_{r_3} P(R_4 | r_3) P(r_3 | u_1 u_2) =
\]
\[
= \langle 0.7, 0.3 \rangle \times 0.653 + \langle 0.3, 0.7 \rangle \times 0.347 = \langle 0.457, 0.196 \rangle + \langle 0.104, 0.243 \rangle
\]
\[
= \langle 0.561, 0.439 \rangle
\]
Lecture Overview

• **Filtering** (posterior distribution over the current state given evidence to date)
  - From intuitive explanation to formal derivation
  - Example

• **Prediction** (posterior distribution over a future state given evidence to date)

• **(start) Smoothing** (posterior distribution over a *past* state given all evidence to date)
Smoothing

**Smoothing**: Compute the posterior distribution over a past state given all evidence to date

- $P(X_k \mid e_{0:t})$ for $1 \leq k < t$
Smoothing

\[ P(X_k \mid e_{0:t}) = P(X_k \mid e_{0:k}, e_{k+1:t}) \] dividing up the evidence

\[ = \alpha P(X_k \mid e_{0:k}) P(e_{k+1:t} \mid X_k, e_{0:k}) \] using…

\[ = \alpha P(X_k \mid e_{0:k}) P(e_{k+1:t} \mid X_k) \] using…

**forward message from filtering up to state k**, \( f_{0:k} \)

**backward message**, \( b_{k+1:t} \)

computed by a recursive process that runs backwards from \( t \)
Smoothing

\[ P(X_k \mid e_{0:t}) = P(X_k \mid e_{0:k}, e_{k+1:t}) \] dividing up the evidence

\[ = \alpha P(X_k \mid e_{0:k}) P(e_{k+1:t} \mid X_k, e_{0:k}) \text{ using Bayes Rule} \]

\[ = \alpha P(X_k \mid e_{0:k}) P(e_{k+1:t} \mid X_k) \text{ By Markov assumption on evidence} \]

- forward message from filtering up to state k, \( f_{0:k} \)
- \( b_{k+1:t} \) backward message, computed by a recursive process that runs backwards from t
Learning Goals for today’s class

➢ You can:

- Describe Filtering and derive it by manipulating probabilities
- Describe Prediction and derive it by manipulating probabilities
- Describe Smoothing and derive it by manipulating probabilities
• Keep Reading Textbook Chp 8.5
• Keep working on assignment-2 (due on Fri, Oct 20)