CSE 473: Artificial Intelligence
Spring 2014

Hidden Markov Models &
Exact Inference

Hanna Hajishirzi

Many slides adapted from Dan Weld, Pieter Abbeel, Dan Klein,
Stuart Russell, Andrew Moore & Luke Zettlemoyer
Outline

- Probabilistic sequence models (and inference)
  - Probability and Uncertainty – Preview
  - Markov Chains
  - Hidden Markov Models
  - Exact Inference
  - Particle Filters
Recap: Reasoning Over Time

- **Stationary Markov models**

  \[ P(X_1) \quad P(X|X_{-1}) \]

- **Hidden Markov models**

  \[ P(E|X) \]

<table>
<thead>
<tr>
<th>X</th>
<th>E</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>umbrella</td>
<td>0.9</td>
</tr>
<tr>
<td>rain</td>
<td>no umbrella</td>
<td>0.1</td>
</tr>
<tr>
<td>sun</td>
<td>umbrella</td>
<td>0.2</td>
</tr>
<tr>
<td>sun</td>
<td>no umbrella</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Hidden Markov Models

- Defines a joint probability distribution:

\[
P(X_1, \ldots, X_n, E_1, \ldots, E_n) = \\
P(X_{1:n}, E_{1:n}) = \\
P(X_1)P(E_1|X_1) \prod_{t=2}^{N} P(X_t|X_{t-1})P(E_t|X_t)
\]
HMM Computations: Inference

- **Given**
  - joint $P(X_{1:n}, E_{1:n})$
  - evidence $E_{1:n} = e_{1:n}$

- **Inference problems include:**
  - **Filtering,** find $P(X_t | e_{1:t})$ for current $t$
  - **Smoothing,** find $P(X_t | e_{1:n})$ for past $t$
  - Most probable explanation, find
    
    $$x^*_{1:n} = \arg\max_{x_{1:n}} P(x_{1:n} | e_{1:n})$$
Inference Recap: Simple Cases

That's my rule!

\[ P(X_1) \quad P(X_t|X_{t-1}) \]
\[ P(E|X) \]

\[ P(X_1|e_1) \]

\[ P(x_1|e_1) = P(x_1, e_1)/P(e_1) \]
\[ \propto_{X_1} P(x_1, e_1) \]
\[ = P(x_1)P(e_1|x_1) \]
Inference Recap: Simple Cases

\[ P(X_1) \]
\[ P(E|X) \]
\[ P(X_1|e_1) \]

\[ P(x_1|e_1) = \frac{P(x_1, e_1)}{P(e_1)} \]
\[ \propto_{X_1} P(x_1, e_1) \]
\[ = P(x_1)P(e_1|x_1) \]

\[ P(X_t|X_{t-1}) \]
\[ P(X_2) \]

\[ P(x_2) = \sum_{x_1} P(x_1, x_2) \]
\[ = \sum_{x_1} P(x_1)P(x_2|x_1) \]
Passage of Time

- We want to know: \( B_t(X) = P(X_t|e_{1:t}) \)
- We can derive the following updates

\[
P(X_{t+1}|e_{1:t}) = \sum_{x_t} P(X_{t+1}, x_t|e_{1:t})
= \sum_{x_t} P(X_{t+1}|x_t, e_{1:t})P(x_t|e_{1:t})
= \sum_{x_t} P(X_{t+1}|x_t)P(x_t|e_{1:t})
\]

- To get \( B_t(X) \) compute each entry and normalize
Passage of Time

- Assume we have current belief $P(X \mid \text{evidence to date})$
  \[ B(X_t) = P(X_t \mid e_{1:t}) \]

- Then, after one time step passes:
  \[ P(X_{t+1} \mid e_{1:t}) = \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid e_{1:t}) \]

- Or, compactly:
  \[ B'(X') = \sum_x P(X' \mid x) B(x) \]

- Basic idea: beliefs get “pushed” through the transitions
  - With the “B” notation, we have to be careful about what time step $t$ the belief is about, and what evidence it includes
Example: Passage of Time

Without observations, uncertainty “accumulates”

\[ B'(X') = \sum_x P(X'|x)B(x) \]

Transition model: ghosts usually go clockwise
Observations

- Assume we have current belief $P(X | \text{previous evidence})$:

$$
P(X_{t+1}|e_{1:t+1}) = \frac{P(X_{t+1}, e_{t+1}|e_{1:t})}{P(e_{t+1}|e_{1:t})}
\propto P(X_{t+1}, e_{t+1}|e_{1:t})
= P(e_{t+1}|e_{1:t}, X_{t+1}) P(X_{t+1}|e_{1:t})
= P(e_{t+1}|X_{t+1}) P(X_{t+1}|e_{1:t})
$$
Observations

- Assume we have current belief $P(X \mid \text{previous evidence})$:

  \[ B'(X_{t+1}) = P(X_{t+1} \mid e_{1:t}) \]

- Then:

  \[ P(X_{t+1} \mid e_{1:t+1}) \propto P(e_{t+1} \mid X_{t+1}) P(X_{t+1} \mid e_{1:t}) \]

- Or:

  \[ B(X_{t+1}) \propto P(e \mid X) B'(X_{t+1}) \]

- Basic idea: beliefs reweighted by likelihood of evidence

- Unlike passage of time, we have to renormalize
Example: Observation

- As we get observations, beliefs get reweighted, uncertainty “decreases”

\[ B(X) \propto P(e|X)B'(X) \]
Online Belief Updates

- Every time step, we start with current $P(X | \text{evidence})$
- We update for time:

$$P(x_t | e_{1:t-1}) = \sum_{x_{t-1}} P(x_{t-1} | e_{1:t-1}) \cdot P(x_t | x_{t-1})$$

- We update for evidence:

$$P(x_t | e_{1:t}) \propto_X P(x_t | e_{1:t-1}) \cdot P(e_t | x_t)$$
The Forward Algorithm

- We want to know: \( B_t(X) = P(X_t|e_{1:t}) \)
- We can derive the following updates

\[
P(x_t|e_{1:t}) \propto_X P(x_t, e_{1:t})
= \sum_{x_{t-1}} P(x_{t-1}, x_t, e_{1:t})
= \sum_{x_{t-1}} P(x_{t-1}, e_{1:t-1})P(x_t|x_{t-1})P(e_t|x_t)
= P(e_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1}, e_{1:t-1})
\]

- To get \( B_t(X) \) compute each entry and normalize
- Problem: space is \(|X|\) and time is \(|X|^2\) per time step
An HMM is defined by:

- **Initial distribution:** $P(X_1)$
- **Transitions:** $P(X_t|X_{t-1})$
- **Emissions:** $P(E|X)$
Forward Algorithm

<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>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>
Example Pac-man
## Summary: Filtering

- Filtering is the inference process of finding a distribution over $X_T$ given $e_1$ through $e_T$: $P(X_T | e_1:t)$
- We first compute $P(X_1 | e_1)$: $P(x_1|e_1) \propto P(x_1) \cdot P(e_1|x_1)$
- For each $t$ from 2 to $T$, we have $P(X_{t-1} | e_1:t-1)$
- Elapse time: compute $P(X_t | e_1:t-1)$

$$P(x_t|e_{1:t-1}) = \sum_{x_{t-1}} P(x_{t-1}|e_{1:t-1}) \cdot P(x_t|x_{t-1})$$

- Observe: compute $P(X_t | e_{1:t-1}, e_t) = P(X_t | e_{1:t})$

$$P(x_t|e_{1:t}) \propto P(x_t|e_{1:t-1}) \cdot P(e_t|x_t)$$