Informatics 2D – Reasoning and Agents

Semester 2, 2011-12

Alex Lascarides
alex@inf.ed.ac.uk

Lecture 26 – Time and Uncertainty I
22nd March 2012
adapted from slides by Michael Rovatsos

Where are we?

Last time . . .

- Completed our account of Bayesian Networks
- Dealt with methods for exact and approximate inference in BNs
- Enumeration, variable elimination, sampling, MCMC

Today . . .

- **Time and uncertainty I**

Time and uncertainty

- So far we have only seen methods for describing uncertainty in static environments
- Every variable had a fixed value, we assumed that nothing changes during evidence collection of diagnosis
- Many practical domains involve uncertainty about processes that can be modelled with probabilistic methods
- Basic idea straightforward: imagine one BN model of the problem for every time step and reason about changes between them

States and observations

- Adopted approach similar to situation calculus: series of snapshots (**time slices**) will be used to describe process of change
- Snapshots consist of observable random variables $\mathbf{E}_t$ and non-observable ones $\mathbf{X}_t$
- For simplicity, we assume sets of (non)observable variables remain constant over time, but this is not necessary
- Observation at t will be $\mathbf{E}_t = \mathbf{e}_t$ for some set of values $\mathbf{e}_t$
- Assume that states start at $t = 0$ and evidence starts arriving at $t = 1$

States and observations

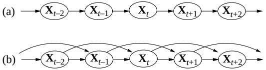
- Example: underground security guard wants to predict whether it is raining but only observes every morning whether director comes in carrying umbrella
- For each day, $\mathbf{E}_t$ contains variable U_t (whether the umbrella appears) and $\mathbf{X}_t$ contains state variable R_t (whether it's raining)
- Evidence $U_1, U_2, \ldots$, state variables $R_0, R_1, \ldots$
- Use notation $a : b$ to denote sequences of integers, e.g. $U_1, U_2, U_3 = U_{1:3}$

Stationary processes and the Markov assumption

- How do we specify dependencies among variables?
- Natural to arrange them in temporal order (causes usually precede effects)
- Problem: set of variables is unbounded (one for each time slice), so we would have to
  - specify unbounded number of conditional probability tables
  - specify an unbounded number of parents for each of these
- Solution to first problem: we assume that changes are caused by a **stationary process** – the laws that govern the process do not change themselves over time (not to be confused with "static")
- For example, $\mathbf{P}(U_t|Parents(U_t))$ does not depend on t

Stationary processes and the Markov assumption

- Solution to second problem: **Markov assumption** – the current state only depends on a finite history of previous states
- Such processes are called Markov processes or Markov chains
- Simplest form: **first-order Markov processes**, every state depends only on predecessor state
- We can write this as $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$
- This conditional distribution is called **transition model**
- Difference between first-order and second-order Markov processes:



Stationary processes and the Markov assumption

- Additionally, we will assume that evidence variables depend only on current state:

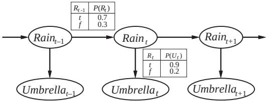
$$\mathbf{P}(\mathbf{E}_t|\mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}) = \mathbf{P}(\mathbf{E}_t|\mathbf{X}_t)$$

- This is called the **sensor model** (**observation model**) of the system
- Notice direction of dependence: state causes evidence (but inference goes in other direction!)
- In umbrella world, rain causes umbrella to appear
- Finally, we need a prior distribution over initial states $\mathbf{P}(\mathbf{X}_0)$
- These three distributions give a specification of the complete JPD:

$$\mathbf{P}(\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_t, \mathbf{E}_1, \ldots, \mathbf{E}_t) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i|\mathbf{X}_{i-1})\mathbf{P}(\mathbf{E}_i|\mathbf{X}_i)$$

Umbrella world example

- Bayesian network structure and conditional distributions
- Transition model $P(Rain_t|Rain_{t-1})$, sensor model $P(Umbrella_t|Rain_t)$



- Rain depends only on rainfall on previous day, whether this is reasonable depends on domain!

Stationary processes and the Markov assumption

- If Markov assumptions seems too simplistic for some domains (and hence, inaccurate), two measures can be taken
  - We can increase the order of the Markov process model
  - We can increase the set of state variables
- For example, add information about season, pressure or humidity
- But this will also increase prediction requirements (problem alleviated if we add new sensors)
- Example: dependency of predicting movement of robot on battery power level
  - add battery level sensor

Inference tasks in temporal models

- Now that we have described general model, we need inference methods for a number of tasks
- **Filtering/monitoring**: compute **belief state** given evidence to date, i.e. $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$
- Interestingly, an almost identical calculation yields the **likelihood** of the evidence sequence $P(\mathbf{e}_{1:t})$
- **Prediction**: computing posterior distribution over a future state given evidence to date: $\mathbf{P}(\mathbf{X}_{t+k}|\mathbf{e}_{1:t})$
- **Smoothing/hindsight**: compute posterior distribution of past state, $\mathbf{P}(\mathbf{X}_k|\mathbf{e}_{1:t})$, $0 \leq k < t$
- **Most likely explanation**: compute $\arg\max_{\mathbf{x}_{1:t}} P(\mathbf{x}_{1:t}|\mathbf{e}_{1:t})$ i.e. the most likely sequence of states given evidence

Filtering and prediction

- Done by **recursive estimation**: compute result for t+1 by doing it for t and then updating with new evidence $\mathbf{e}_{t+1}$. That is, for some function f:

$$\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1}) = f(\mathbf{e}_{t+1}, \mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t}))$$

-

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \\
&= \alpha \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}, \mathbf{e}_{1:t})\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t}) && \text{(Bayes' rule)} \\
&= \alpha \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1})\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t}) && \text{(Markov property)} \\
&= \alpha \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t|\mathbf{e}_{1:t}) && \text{(conditioning)} \\
&= \alpha \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t) P(\mathbf{x}_t|\mathbf{e}_{1:t}) && \text{(Markov assumption)}
\end{aligned}$$

- $\mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1})$ is sensor model; $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t)$ is transition model, $P(\mathbf{x}_t|\mathbf{e}_{1:t})$ is recursive bit (current state distribution).

Filtering and prediction

- We can view estimate $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$ as "message" $\mathbf{f}_{1:t}$ propagated and updated through sequence
- We write this process as $\mathbf{f}_{1:t+1} = \alpha \text{FORWARD}(\mathbf{f}_{1:t}, \mathbf{e}_{t+1})$
- Time and space requirements for this are constant regardless of length of sequence
- This is extremely important for agent design!
- All this is very abstract, let's look at an example

informatics

Example Compute $\mathbf{P}(R_2|u_{1:2})$, $U_1 = true$, $U_2 = true$

- Suppose $\mathbf{P}(R_0) = \langle 0.5, 0.5\rangle$
- Prediction for R_1:

$$\mathbf{P}(R_1) = \sum_{r_0} \mathbf{P}(R_1|r_0)P(r_0) = \langle 0.7, 0.3\rangle \times 0.5 + \langle 0.3, 0.7\rangle \times 0.5 = \langle 0.5, 0.5\rangle$$

- Update with evidence from Day 1 ($U_1 = true$) yields:

$$\mathbf{P}(R_1|u_1) = \alpha \mathbf{P}(u_1|R_1)\mathbf{P}(R_1) = \alpha\langle 0.9, 0.2\rangle\langle 0.5, 0.5\rangle \approx \langle 0.818, 0.182\rangle$$

- Day 2: Observation $U_2 = true$, prediction and update yield

$$\begin{aligned}\mathbf{P}(R_2|u_1) = \textstyle\sum_{r_1} \mathbf{P}(R_2|r_1)P(r_1|u_1) &= \langle 0.7, 0.3\rangle \times 0.818 + \langle 0.3, 0.7\rangle \times 0.182 \\ &= \langle 0.627, 0.373\rangle \\ \mathbf{P}(R_2|u_1, u_2) = \alpha \mathbf{P}(u_2|R_2)\mathbf{P}(R_2|u_1) &= \alpha\langle 0.9, 0.2\rangle\langle 0.627, 0.373\rangle \\ &= \alpha\langle 0.565, 0.075\rangle \\ &\approx \langle 0.883, 0.117\rangle\end{aligned}$$

informatics

Filtering and prediction

- Prediction works like filtering without new evidence
- Computation involves only transition model and not sensor model:

$$\mathbf{P}(\mathbf{X}_{t+k+1}|\mathbf{e}_{1:t}) = \sum_{\mathbf{x}_{t+k}} \mathbf{P}(\mathbf{X}_{t+k+1}|\mathbf{x}_{t+k})P(\mathbf{x}_{t+k}|\mathbf{e}_{1:t})$$

- As we predict further and further into the future, distribution of rain converges to $\langle 0.5, 0.5\rangle$
- This is called the **stationary distribution** of the Markov process (the more uncertainty, the quicker it will converge)

informatics

Filtering and prediction

- We can use the above method to compute **likelihood** of evidence sequence $P(\mathbf{e}_{1:t})$
- Useful to compare different temporal models
- Use a likelihood message $\mathbf{l}_{1:t} = \mathbf{P}(\mathbf{X}_t, \mathbf{e}_{1:t})$ and compute

$$\mathbf{l}_{1:t+1} = \alpha \text{FORWARD}(\mathbf{l}_{1:t}, \mathbf{e}_{t+1})$$

- Once we compute $\mathbf{l}_{1:t}$, summing out yields likelihood

$$L_{1:t} = P(\mathbf{e}_{1:t}) = \sum_{\mathbf{x}_t} \mathbf{l}_{1:t}(\mathbf{x}_t)$$

informatics

Summary

- Time and uncertainty (states and observations)
- Stationarity and Markov assumptions
- Inference in temporal models
- Filtering and prediction
- Next time: **Time and Uncertainty II**