PROBABILISTIC REASONING OVER TIME

Outline

- Time and uncertainty
- Inference in temporal models
- Hidden Markov models
- Dynamic Bayesian networks

Based on the textbook by Stuart Russell & Peter Norvig:

Artificial Intelligence, A Modern Approach (2nd Edition)
Chapter 15; excluding Sections 15.4 and 15.6

1. TIME AND UNCERTAINTY

- We have previously developed our techniques for probabilistic reasoning in the context of static worlds.
- E.g., when repairing a car, it is assumed that whatever is broken remains broken during the process of diagnosis.
- However, in certain domains dynamic aspects become essential,

Example. A doctor is treating a diabetic patient,

- Recent insulin doses, food intake, blood sugar measurements, and other physical signs serve as pieces of evidence,
- The doctor decides about food intake and insulin dose.

States and Observations

- The process of change is viewed as a series of snapshots, each of which describes the state of the world at a particular time.
- Each time slice involves a set of random variables indexed by $t$:
  1. the set of unobservable state variables $X_t$ and
  2. the set of observable evidence variables $E_t$.
- The observation at time $t$ is $E_t = e_t$ for some set of values $e_t$.
- The notation $X_{a:b}$ denotes the sets of variables from $X_a$ to $X_b$.
- The interval between time slices depends on the problem!

Stationary Processes and the Markov Assumption

- In a stationary process, the changes in the world state are governed by laws that do not themselves change over time.
- A first-order Markov process satisfies an equation

$$P(X_t | X_{0:t-1}) = P(X_t | X_{t-1})$$

where $P(X_t | X_{t-1})$ forms the transition model of the process,

- In addition, it is typical to assume a sensor model of the form

$$P(E_t | X_{0:t}, E_{0:t-1}) = P(E_t | X_t)$$

so that observations depend only on the current state.
2. INFERENCET IN TEMPORAL MODELS

Having set up the generic temporal model, we may formulate the basic inference tasks that are to be solved.

- **Filtering or monitoring**: the task is to compute the belief state, i.e., the posterior distribution $P(X_t | e_{1:t})$ over the current state.
- **Prediction**: the posterior distribution $P(X_{t+k} | e_{1:t})$ over the future state is of interest for some $k > 0$.
- **Smoothing or hindsight**: the aim is to compute $P(X_k | e_{1:t})$ where $0 \leq k < t$ for some past state.
- **Most likely explanation** is a sequence of states $x_{1:t}$ that maximizes $P(x_{1:t} | e_{1:t})$ for the observations $e_{1:t}$ to date.

Example. When modeling a battery-powered robot wandering in the $xy$-plane, the battery level has to be taken into account.
Example. The security guard has a prior belief \( P(R_0) = \langle 0.5, 0.5 \rangle \) about the state.

1. The prediction from \( t = 0 \) to \( t = 1 \) gives
   \[
P(R_1) = \sum_{r_0} P(R_1 \mid r_0) P(r_0) = \langle 0.7, 0.3 \rangle \times 0.5 + \langle 0.3, 0.7 \rangle \times 0.5 = \langle 0.5, 0.5 \rangle.
   \]
2. Updating this distribution with the evidence \( u_1 \) for \( t = 1 \) gives
   \[
P(R_1 \mid u_1) = \alpha P(u_1 \mid R_1) P(R_1) = \alpha \langle 0.9, 0.2 \rangle \langle 0.5, 0.5 \rangle
   = \alpha \langle 0.45, 0.1 \rangle \approx \langle 0.818, 0.182 \rangle.
   \]
3. Similarly, we obtain
   \[
P(R_2 \mid u_1) \approx \langle 0.627, 0.373 \rangle \text{ and } P(R_2 \mid u_1, u_2) = \alpha \langle 0.565, 0.075 \rangle \approx \langle 0.883, 0.117 \rangle.
   \]
   The probability of rain increases due to repeated evidence.

Example. Let us demonstrate smoothing with the umbrella example:

1. \( P(R_1 \mid u_1, u_2) = \alpha f_{1:2}^{u_2} = \alpha P(R_1 \mid u_1) P(u_2 \mid R_1) \) where we already know the distribution \( f_{1:1} = P(R_1 \mid u_1) = \langle 0.818, 0.182 \rangle \).

2. The distribution \( b_{2:2} = P(u_2 \mid R_1) = \sum_{u_1} P(u_2 \mid u_2) P(u_2 \mid R_1) = 0.9 \times \langle 0.7, 0.3 \rangle + 0.2 \times \langle 0.3, 0.7 \rangle = \langle 0.69, 0.41 \rangle. \)

3. By substituting these distributions and normalizing, we obtain
   \[
P(R_1 \mid u_1, u_2) = \alpha \langle 0.818, 0.182 \rangle \langle 0.69, 0.41 \rangle \approx \langle 0.883, 0.117 \rangle.
   \]
   Thus the smoothed estimate is higher than the filtered estimate.

Smoothing

- The task is to compute \( P(X_k \mid e_{1:t}) \) for \( 0 \leq k < t \) referring to past,
- Using a backward message \( b_{k+1:t} = P(e_{k+1:t} \mid X_k) \), we obtain
  \[
P(X_k \mid e_{1:t}) = \alpha f_{t:k} b_{k+1:t}.
  \]
- The backward message \( b_{k+1:t} \) can be computed using
  \[
b_{k+1:t} = \sum_{X_{k+1}} P(e_{k+1} \mid x_{k+1}) P(e_{k+2:t} \mid x_{k+1}) P(X_{k+1} \mid X_k).
  \]
- Whenever \( k + 1 = t \), the sequence \( e_{k+2:t} \) becomes empty and
  \[
P(e_{k+2:t} \mid x_{k+1}) = P(T \mid x_{k+1}) = 1 \text{ where } T \text{ stands for truth},
  \]
- This leads to a recursive definition, or algorithm
  \[
b_{k+1:t} = \alpha \text{BACKWARD}(b_{k+2:t}, e_{k+1:t}).
  \]

Finding the Most Likely Sequence

Example. Suppose that the security guard makes the following observations during the first five days: \( u_1, u_2, \neg u_3, u_4, u_5 \).

What is the weather sequence most likely to explain this?

- For each pair of states \( x_{i+1} \) and \( x_i \), there is a recursive relationship between the most likely paths to \( x_{i+1} \) and \( x_i \):
  \[
  \max_{x_i} \cdots x_1 P(x_1, \ldots, x_i, X_i \mid e_{1:i+1})
  = \alpha P(e_{i:i} \mid X_{i+1}) \times 
  \max_{x_i} \left( P(X_{i+1} \mid x_i) \max_{x_{i-1}} \cdots x_1 P(x_1, \ldots, x_{i-1}, X_i \mid e_{1:i}) \right).
  \]
- This equation is analogous to the one used in filtering.
4. Dynamic Bayesian Networks

- A dynamic Bayesian network (DBN) represents how the state of the environment evolves over time.
- Each time slice of a DBN may have any number of state variables $X_t$ and evidence variables $E_t$.
- Every HMM can be transformed into a DBN and vice versa.
- By decomposing the state of a complex system into its constituent variables, the DBN is able to take advantage of the sparseness in the temporal probability model.

**Example.** The transition model of a DBN with 20 Boolean state variables, each of which has three parents in the preceding slide, has $20 \times 2^3 = 160$ probabilities while its HMM counterpart has $2^{40}$.

---

3. Hidden Markov Models

- In a hidden Markov Model, or HMM, the world is described by a single discrete random variable $X_t$ taking values $1, \ldots, S$ which correspond to the states of the world.
- The transition model $P(X_t | X_{t-1})$ becomes an $S \times S$ matrix $T$ such that $T_{ij} = P(X_t = j | X_{t-1} = i)$.
- Forward and backward reasoning are simplified as follows:
  $$f_{t+1} = \alpha T f_t$$
  $$b_{t+1} = \alpha T b_{t+1}$$
- Forward-backward type reasoning are of orders $S^2 \times t$ and $S \times t$, respectively.

---

© 2006 TKK / Laboratory for Theoretical Computer Science
Example. A robot is described with state variables $X_t = (X_t, Y_t)$ for position and $X_t = (X_t, Y_t)$ for velocity and $Battery_t$ for the actual battery charge level.

Both position (evidence variables $Z_t$) and the battery charge level (evidence variable $B\text{Meter}_t$) are measured.

**Exact Inference in DBNs**

- The previous algorithms for inference in Bayesian networks can be applied to dynamic Bayesian networks.
- Given a sequence of observations, one can **unroll** a DBN until the network is large enough to accommodate the observations.
- Unrolling can also be done on a slice-by-slice basis.
- In the general case, the complexity of reasoning is exponential.